# Table Of Contents

# Evaluation of the effectiveness of using an electromechanical shock absorber in a subway car

**Indexed by:**
Web of Science Group

Borys Liubarskyi [a], Valeriy Kuznetsov [b], Ewa Kardas-Cinal [c,*], Natalia Lukashova [d], Oleksandr Petrenko [d], Oleh Nikonov [e], Dmytro Nikonov [e]

[a] National Technical University «Kharkiv Polytechnic Institute», Department of Electrical Transport and Diesel Locomotive, Kyrpychova str., 2, Kharkiv, Ukraine, 61002
[b] Warsaw Management University, Department of Technical Science, ul.Kawęczyńska 36, 03-772 Warsaw, Poland
[c] Warsaw University of Technology, Faculty of Transport, ul. Koszykowa 75, 00-662 Warsaw, Poland
[d] O. M. Beketov National University of Urban Economy in Kharkiv, Department of Electrical Transport, Marshala Bazhanova str., 17, Kharkiv, Ukraine, 61002
[e] Kharkiv National Automobile and Highway University, Yaroslava Mudrogo str., 25, Kharkiv, Ukraine, 61002

## Highlights

- Study of vertical oscillations of a subway car due to random track irregularities.

- Indicators of dynamic quality of the metro car with electromechanical shock-absorbers.

- Advantage of electromechanical shock absorbers over central spring suspension.

- Recommendation of using electric shock absorbers in the design of a subway car.

## Abstract

The paper evaluates the effectiveness of the electromechanical shock absorber for a subway car based on dynamic quality indicators. To determine them, a method of synthesis of random forced vertical oscillations of a dynamic model of a subway car was developed, which is based on the method of sliding summation taking into account the spectral density obtained when processing the test results of a subway car. Also in the work the technique on definition of indicators of dynamic quality of the metrocar with electromechanical shock-absorbers which is based on modeling of processes of movement of the subway car on a site of a way with casual roughness and definition, by results of modeling, indicators of dynamic quality is created. The study of forced random oscillations of the dynamic model of a subway car showed that the spring suspension based on electromechanical shock absorbers has a significant advantage over the central spring suspension for all quality indicators in the entire range of speeds. The dependence of the average power generated by the shock absorber is established.

## Keywords

electromechanical shock absorber, subway car, indicators of dynamic quality, spring suspension, running gear.

## 1. Introduction

Investigation of rail vehicles dynamics is one of the main elements in developing new vehicle designs, approval process and shaping and monitoring the railway track geometry [2,10]. A key factor in such investigations is random track irregularities which arise due to repetitive load from running vehicles [11]. These irregularities disturb the railway vehicle motion and result in vehicle vibrations which affect the ride comfort of travelling passengers [6] and the safety of the freight transportation [4,24].

The comfort of subway passengers is directly related to the smooth movement of subway cars. To ensure it, it is necessary to have an effective system for damping vehicle vibrations [15]. The system of damping oscillations on most subway cars usually has two, different parameters of the degree of spring suspension. Conversion of energy of oscillations of a body and elements of car into other types of energy is provided by existence in a running gear of dampers of oscillations - shock-absorbers.

Characteristics of vibration dampers determine the nature of oscillations of the running gear of the rolling stock and directly affect traffic safety and its speed.

Shock absorber performance is determined by the type and design of the shock absorber. As noted in [20] on the smoothness of the best performance inherent in pneumatic shock absorbers, their use requires an additional system of pneumatic power - compressors that reduce the overall energy efficiency of electric rolling stock [14]. To increase the energy efficiency of electric rolling stock, the authors [13] proposed the use of electromechanical shock absorbers. Shock absorbers

of this type do not consume additional energy, instead they are able to partially recover the energy of the oscillations. The generating capabilities of electromechanical shock absorbers make it possible to increase the energy efficiency of rolling stock in general and the chassis in particular [13].

## 2. Analysis of literature data and problem statement

A comprehensive analysis conducted in [19] showed that in world practice the most common is the spring suspension of trolleys. The suspension consists of an axle box and a central part. A similar system is used on subway cars in Ukraine. The purpose of the spring suspension is to reduce the impact of track irregularities and curved road sections that occur during the movement of the metro car. The most common vibration damping systems that occur in the central suspension of trolleys are hydraulic shock absorbers [19]. In [16] it is noted that a common technical solution is friction vibration dampers, supplemented by hydraulic dampers. Such combined systems are installed in the second stage of the spring suspension. To increase the comfort of passengers and increase the smoothness of the metro car series E-KM [9] at PJSC "Kryukiv Carriage Plant" (Kremenchuk, Ukraine) have new carts model 68-7054 with central pneumatic spring suspension, which is supplemented by the shock absorbers. However, the use of pneumatic springs requires additional energy losses and power of the pneumatic system [14].

An alternative way to increase the smoothness of movement is the use of electromechanical energy absorber in the second stage of the spring suspension of metro cars [12, 20].

Electromechanical shock absorbers are most common in road transport. Structurally, such shock absorbers are integrated into the suspension type McPherson [21] together with the spring. The process of recovery of vibration energy from the electromechanical shock absorber and its connection to the onboard network of the vehicle is described in [3]. Estimation of energy efficiency of systems of recovery of energy of mechanical fluctuations for cars KRAZ-65055 (Russia), KAMAZ-45143 (Russia), Gazelle-3302 (Russia), VAZ 2101 (Russia), Reno Megane (France), Toyota Camry (Japan) on the basis Previous results of mathematical modeling of such systems are given in [25]. As noted in [25], the use of electromechanical shock absorbers has provided opportunities to increase the efficiency of vehicles.

To analyze the oscillations of the rail transport body, it is necessary to create a generalized mathematical model of the main approaches to its creation proposed in [23]. The authors of [23] also analyzed the complex processes of oscillations in the mechanical system of the chassis of vehicles.

For an electromechanical shock absorber, the design of which is proposed in [16] and shown in Fig. 1 the authors of [12] developed a mathematical model of a subway car on two two-axle carts with an axial characteristic of $2_0$-$2_0$, in the central stage of the spring suspension which used springs and electromechanical dampers.

The model is developed on the basis of the spatial kinematic scheme of the model of the subway car shown in Fig. 2 and 3.

The peculiarity of the model is that it contains the following components: 17 differential equations of the second order, which describe the operation of the mechanical part of the car-rail track; 8 first-order differential equations that describe the operation of 4 electromechanical shock absorbers. The model is supple-



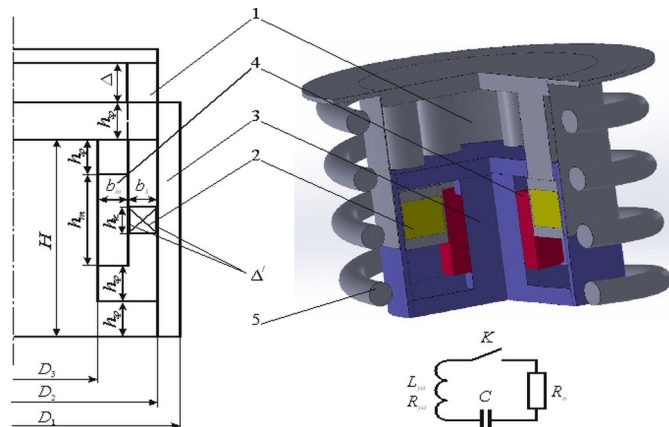*Fig. 1. Electromechanical shock-absorber of a direct current: 1 - an anchor; 2 – armature winding; 3 - bed; 4 - permanent magnet; 5 - spring*
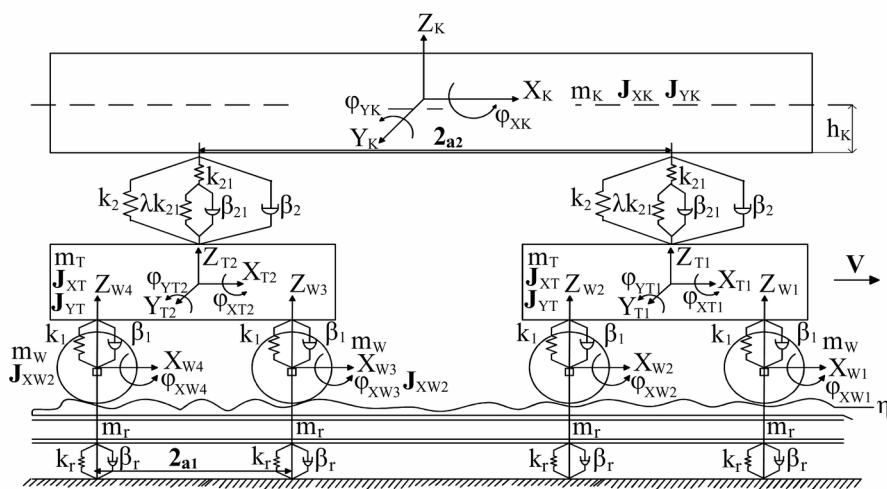


*Fig. 2. Spatial kinematic scheme of the subway car model (side view)*
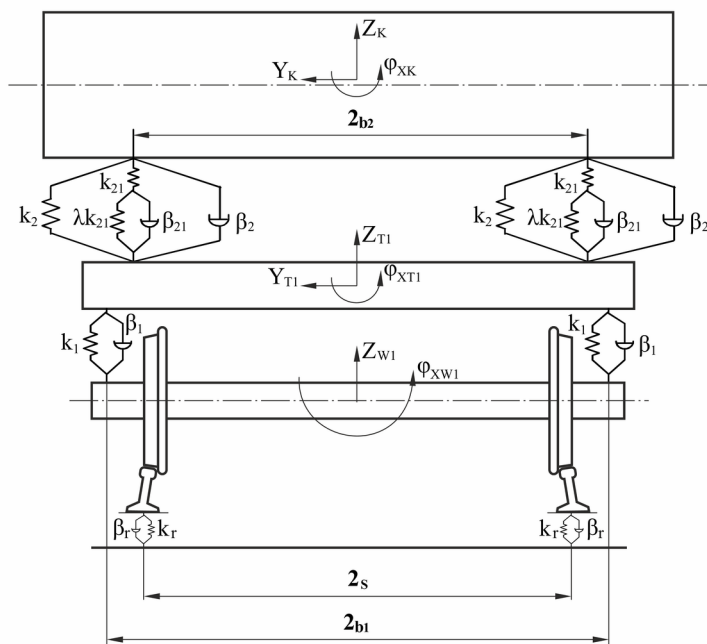


*Fig. 3. Spatial kinematic scheme of the subway car model (front view)*

mented by three polynomials of 32 and 63 orders, which describe the state of the magnetic field of electromechanical shock absorbers and

their electromagnetic force, as well as 4 algebraic coupling equations obtained on the basis of the provisions given in [14].

To assess the effectiveness of the spring suspension, as described in [20], it is necessary to determine for the studied model of the subway car the dependence of the values of their dynamic qualities on the speed of movement. The obtained values of quality indicators must be compared with the permissible values, which are regulated by current regulations [22]. The allowable speed $[v]$ is determined and checked taking into account the unevenness of the track $[v]> v_k$.

Permissible values of indicators of dynamic qualities for subway cars intended for transportation of passengers are accepted according to [22] and are included in table 1.

Table 1. Permissible values of indicators of dynamic qualities of subway cars

| Type of dynamic quality indicator | Valid value |
|---|---|
| Smoothness coefficient, C | 3,25 |
| Maximum vertical acceleration, $\ddot{z}_k$ | 0,35g |
| Coefficient of dynamics in the central suspension, $\kappa_{dc}$ | 0,2 |
| Coefficient of dynamics in axle box suspension, $\kappa_{db}$ | 0,35 |

An important indicator of the energy efficiency of the electromechanical shock absorber is the average power generated by random excitation of track irregularities, which is determined by the expression:

$$ P = \sum_{k=1}^{4} \int_{0}^{t_{end}} R_n i_k^2 \, dt \qquad (1) $$

where $t_{end}$ – time of movement on a track with random irregularities,

$i_k$ – electric current generated in k-th shock absorber (k=1,2,3,4),

$R_n$ – load resistance.

Given the above, it is also possible to note the quality indicator as the maximum movement of the places of installation of auto-coupling on the body, which is important for subway cars moving in the tunnel. Traffic in the tunnel imposes additional restrictions on the amplitude of oscillations of the subway car body.

Determination of dynamic quality indicators is carried out by determining the coefficients.

The coefficient of smoothness of motion C for the implementation of a random process of acceleration of the body duration $t_r$ in the work was calculated according to the expression recommended in [18]:

$$ C = \alpha \cdot 6.67 \sqrt{\sum_{j=1}^{n} \frac{t_j}{t} \left( 2 \int_{f_n}^{f_v} q_n^2(f) G_{\ddot{z}kj}(f) df \right)} \qquad (2) $$

where $\alpha$ – coefficient of proportionality, $\alpha = 4.346$;

$t_j$ – duration of the j-th implementation, $t_p = \sum_{j=1}^{n} t_j$ ;

$f_n, f_v$ – respectively the lower and upper limits of the frequency range of the measured accelerations;

$G_{\ddot{z}kj}(f)$ – spectral density of the j-th implementation of the process of body acceleration;

$q_n(f)$ – normalized amplitude-frequency characteristic of the correcting filter (approximation of physiological coefficient)

$$ q_n(f) = 1.15 \sqrt{\frac{1+0.1f^2}{(1+4.04f^2)\left[(1-0.0364f^2)+0.045f^2\right]}} . \qquad (3) $$

The coefficient of vertical dynamics of the car body is determined [17]:

$$ k_{dv} = \frac{P_{dv} - P_{cm}}{P_{cm}} \qquad (4) $$

where $P_{dv}$ – dynamic vertical force transmitted from the body to the chassis in the central or axle suspension;

$P_{cm}$ – static load transmitted from the body to the chassis.

In [18] a comparative analysis of spring suspension and suspension based on pneumatic springs was performed. A similar approach is rationally used in assessing the quality of use of electromechanical shock absorbers.

As follows from the analysis [3, 9, 12-14, 16, 18-23, 25] of the current state of the problem, the problem is partially solved, because only the following results were obtained. The design and mathematical model of the electromechanical shock absorber for the metrocar have been developed. The model allows to model the processes of metrocar movement during the passage of the track with different types of track irregularities. The authors of the work [3, 9, 12-14, 16, 18-23, 25] developed approaches to determining the indicators of dynamic qualities of subway cars. However, the developed models do not allow to determine these indicators for subway cars equipped with electromechanical shock absorbers. Also the problem of the comparative analysis of efficiency of application of electromechanical shock-absorbers in comparison with traditional designs is not solved.

## 3. The purpose and objectives of the study

The purpose of the study is to conduct a comparative evaluation of the effectiveness of the electromechanical shock absorber for a subway car on the only indicators of the dynamic qualities of subway cars.

To achieve this goal, the following tasks were set:
– to develop a method for determining random forced vertical oscillations of the dynamic model of a subway car;
– to develop a method for determining the indicators of dynamic quality of a metrocar with electromechanical shock absorbers;
– to conduct a comparative analysis of the results of research on the spring suspension of subway cars with an electromechanical shock absorber.

## 4. Comparative evaluation of the effectiveness of the electromechanical shock absorber for the subway car

### 4.1. Random forced vertical oscillations of the dynamic model of a subway car

The coefficient of smoothness is determined by the results of modeling the movement of the metro car. Mathematical model and modeling technique are given in [18]. However, since it is important to determine the coefficient of smoothness according to [18] it is necessary to conduct modeling when driving on a section of track with random unevenness of the rails.

For the synthesis of track irregularities according to [18], it is necessary to determine the spectral power density. The paper applies the

spectral density obtained by processing the test results of the subway car model 81.717. The power spectral density is approximated by a dependence that corresponds to a random process and is differentiated according to the recommendations [18]:

$$G_\eta(\omega, v) = \frac{S_\eta^2}{2\sqrt{\pi}} \sum_{j=1}^{n} \frac{a_j}{\alpha_j \cdot v} \left\{ \exp\left[ -\frac{(\omega + \omega_j v)^2}{4\alpha_j^2 v^2} \right] + \exp\left[ -\frac{(\omega - \omega_j v)^2}{4\alpha_j^2 v^2} \right] \right\}$$

(5)

where $S_\eta^2$ – variance of equivalent geometric irregularity (according to the test results value $S_\eta^2 = 13{,}01$ m$^2$ [20] was obtained,

$\omega$ – current value of frequency, rad/s,

$\omega_j$ – the frequency of the j-th maximum of the spectral density,

$a_j$ – the proportion of variance at the j-th maximum of the spectral density,

$\alpha_j$ – half the width of the j-th maximum of the spectral density at half its height.

For the generation of the track irregularity the method of sliding summation is used. This method is described in [1, 8, 20] and has previously been used both for generating random track irregularities in studies of the railway vehicle dynamics [5, 7] as well as road profiles [1]. The method is equaivalent to the the spectral representation method, where the generated process is modeled as the sum of harmonic signals of fixed amplitudes and random phases.

When applying the sliding summation method at each step of modeling the equivalent geometric irregularity, new values of the generated random numbers are used. As the initial data for modeling, the mean value of the irregularity $\bar{\eta}$ and its power spectral density $G_\eta(\omega)$ are acquired.

Discrete values of the simulated process are obtained with the expression in the form of the sliding summation (moving average):

$$\eta_i = \bar{\eta} + \sum_{j=-M}^{M} c_j \cdot \varepsilon_{i-j}$$

(6)

where $c_j$ – weights,

$\varepsilon_j$ – independent random variables distributed according to Gauss's law, obtained by [20] from the values of the initial stationary discrete white noise distributed according to the uniform law,

$2M$ – number of weights.

To obtain a process whose power spectral density will fully coincide with the assumed spectral density, it is necessary that $2M \rightarrow \infty$. Thus, to reproduce this spectral density at satisfactory level the number of weigths $2M+1$ must be sufficiently large. According to the recommendations given in [17] it was assumed $2M = 9600$.

The value of the weights was determined with the formula:

$$c_j = \frac{1}{\omega_c} \int_0^{\omega_c} \left[ \omega_c G_\eta(\omega) \right]^{0,5} \cdot \cos\frac{j\pi\omega}{\omega_c} d\omega$$

(7)

where $\omega_c$ – the sampling frequency of the random process $\eta(t)$:

$$\omega_c = \frac{\pi}{\Delta t},$$

$\Delta t$ – sampling step.

When modeling $\eta(t)$ by the method of sliding summation, the equivalent geometric irregularity for the speed of motion is modeled, and then in the process of solving the problem "move" on this irregularity with the required speed. This approach is the main one for studying the oscillations of dynamic models of rolling stock during their movement on the simulated section of the track with variable speed according to the operating schedule.

Thus, the functions of the irregularities $\eta$ of left and right rail for the speeds $v = 40, 50, 60, 70, 80, 90$ and $100$ km/h were synthesized. The speed range was chosen taking into account the excess of the design speed ($v_k = 90$ km/).

When modeling the parameters of the analytical expression of the spectral density of expression (5) were determined by $\alpha_j$, $\omega_j$ and $a_j$ (Table 2).

Table 2. Parameters of the analytical expression of spectral density

| The number of the component of the spectrum | $\alpha_j$, (rad/s)/(m/s) | $\omega_j$, (rad/s)/(m/s) | $a_j$ |
|---|---|---|---|
| 1 | 1,0 | 0 | 0,697 |
| 2 | 0,0018 | 0,20096 | 0,12 |
| 3 | 0,002 | 0,25749 | 0,109 |
| 4 | 0,01 | 0,314 | 0,074 |

The duration of the synthesized process according to the recommendations [18, 20] is $t_p = 32{,}768$ s. This roughly corresponds to the implementation length, which is usually accepted in tests of 30 s. The duration of the process in this case corresponds to the number of points of implementation of the reaction of the dynamic system „car - rail track" is $N = 32768$, which corresponds to $2^{15}$. However it meets the requirements for spectral analysis using fast Fourier transform.

## 4.2. Determination of indicators of dynamic quality of a metrocar with electromechanical shock absorbers

The dynamic quality indicators determined by the results of numerical simulation must correspond to the permissible values regulated by the current norms [22], while determining the permissible speed [v], which corresponds to the conditions of inequality $[v] \geq v_k$.

The paper adopted the following indicators of the dynamic qualities of the subway car [20, 22]:

1. Coefficients of smoothness of the course $C_1$ and $C_2$ over the first and second along the movement of pivot points on the floor of the body.

2. The maximum values of accelerations $\ddot{z}_{k1}$ and $\ddot{z}_{k2}$ respectively at the same points as the coefficients of smoothness.

3. Coefficients of dynamics in the relationship between the body and the frame of the trolley, as well as the frame of the trolley and wheelsets: $k_{dc1}$, $k_{dc2}$, $k_{dc3}$ and $k_{dc4}$ - coefficients of dynamics in the elements of the central degree of the spring suspension of the first ($k_{dc1}$, $k_{dc2}$) and the second ($k_{dc3}$, $k_{dc4}$) along the movement of carts; $k_{db1}$, $k_{db2}$, $k_{db3}$, $k_{db4}$, $k_{db5}$, $k_{db6}$, $k_{db7}$ and $k_{db8}$ - dynamics coefficients in the axle stage of the spring suspension of the first ($k_{db1}$, $k_{db2}$, $k_{db3}$ and $k_{db4}$) and the second ($k_{db5}$, $k_{db6}$, $k_{db7}$ and $k_{db8}$) in the course of movement of carts, respectively for the left and right wheels of one wheel pair.

The coefficients of dynamics were defined as the ratio:

$$k_d = \frac{F_{din}^{\max}}{F_{stat}}$$

(8)

where $F_{din}^{max}$ – maximum dynamic force in the above elements of the spring suspension, $F_{stat}$ – the static value of the force in the above elements of the spring suspension.

Statistical characteristics of the distribution of absolute maxima of H processes are determined by [18, 20]:

$$\bar{H} = \int_0^\infty H f(H) dH \qquad (9)$$

where $f(H)$ is the distribution density of the absolute maxima of the random process $q(t)$.

Using the expression for a stationary Gaussian process we find:

$$f(H) = \frac{d}{dH} \exp\left\{ -f_e t_p \exp\left[ -\frac{(H-\bar{H})^2}{S_q^2} \right] \right\} \qquad (10)$$

where $f_e$ – is the effective frequency of the random process $q(t)$;
$t_p$ – time of implementation of a random process [20].

Taking into account the assumptions considered in [17], the values of the average values of the absolute maxima $H$ for the implementations $q(t)$ are determined by the approximate expression:

$$\bar{H}_q \cong S_q\left( \sqrt{2\ln f_e t_p} + \frac{1}{\sqrt{2\ln f_e t_p}} \right) \qquad (11)$$

where $S_q$ is the standard deviation of the random process $q(t)$.

The value of $f_e$ is determined directly by the implementation of reactions $q(t)$:

$$f_e = \frac{n_0}{2 t_p} \qquad (12)$$

where $n_0$ is the number of zeros of the random process $q(t)$.

The standard deviation $S_q$ was determined by the dependence on $q(t)$ by the formula:

$$S_q = \sqrt{\frac{1}{N} \sum_{i=1}^M (q_i - \bar{q})^2} . \qquad (13)$$

The coefficient of smoothness of stroke C on the realizations of the random process of accelerations of the body duration $t_p$ was calculated according to expression (2) [18].

To facilitate the analysis of mathematical models and the obtained results, the natural frequencies of oscillations of subway cars with spring suspension and electromechanical shock absorber were determined.

Eigenfrequency calculations are performed for the frequency of partial oscillations. The results of the calculation are given in table 3.

As can be seen from table 3 the partial frequencies of oscillations of the body and carts are different in the subway cars under study, because the stiffness of the base spring and the spring of the electromechanical shock absorber differ by 40% ($Zw_2$=2174,2 kN/m, $Zw_{21}$=1311,2 kN/m).

Resonant velocities corresponding to the $j$-th maximum of the spectral density of the equivalent geometric irregularity can be determined by the formula [20]:

$$V_p = \frac{f_{ci}}{\omega_j} 3,6 \quad \text{km/h,} \qquad (14)$$

where $f_{ci}$ - the frequency of natural oscillations of the body of the model of the subway car on $i$-th generalized coordinate; $\omega_j$ is the frequency of the $j$-th maximum of the spectral density.

According to the results of calculations, it was found that the resonant velocities even in the body oscillations are outside the studied speed range v = 40-100 km/h. For spring-loaded central suspension, they are in the range of 115-207 km/h [20], and for suspension with electromechanical shock absorbers in the range of 110-196 km/h.

On the other hand, in the results of calculations of the dependences of dynamic quality indicators on the velocity of motion, the first maximum of spectral density approximation with frequency $\omega_1$=0, which accounts for 69.7% of the variance of the equivalent geometric irregularity, should manifest itself.

### 4.3. Comparative analysis of the results of studies of spring suspension of subway cars with electromechanical shock absorbers

The analysis was performed between the parameters of elastic spring suspension and spring suspension using an electromechanical shock absorber.

The dependences of dynamic quality indicators on the speed of movement for the studied model of the subway car are calculated. The obtained values of quality indicators were compared with the permissible values, which are regulated by current regulations (Table 1) and are limited by the movement of the train in the tunnel, as well as among themselves.

*Table 3. Values of partial oscillation frequencies*

| № | Model element | Type of oscillations | The value of the partial frequency, Hz | |
| --- | --- | --- | --- | --- |
| | | | Model with spring spring suspension [1] | Model with electro-mechanical shock absorber |
| 1 | Body | bouncing | 2,32 | 1,89 |
| | | galloping | 2,49 | 1,98 |
| | | lateral sway | 2,21 | 1,82 |
| 2 | Cart | bouncing | 8,09 | 7,93 |
| | | galloping | 5,79 | 5,78 |
| | | lateral sway | 8,08 | 8,07 |
| 3 | Wheelset | bouncing | 45,36 | 45,35 |
| | | lateral sway | 45,92 | 45,92 |

An additional restriction was introduced due to the existing possibility of self-disengagement of metro cars and is not to exceed the difference of the longitudinal axes of auto-couplings, which is allowed, and has the form:

$$z_k \leq \frac{h_{\text{àñ}}}{2} = [z_k] = 0.055 \text{ m.} \qquad (15)$$

The results of the research in the form of graphs of the dependence of the indicators of dynamic quality on the speed $v$ are shown in Figures 4-8.

a)



b)
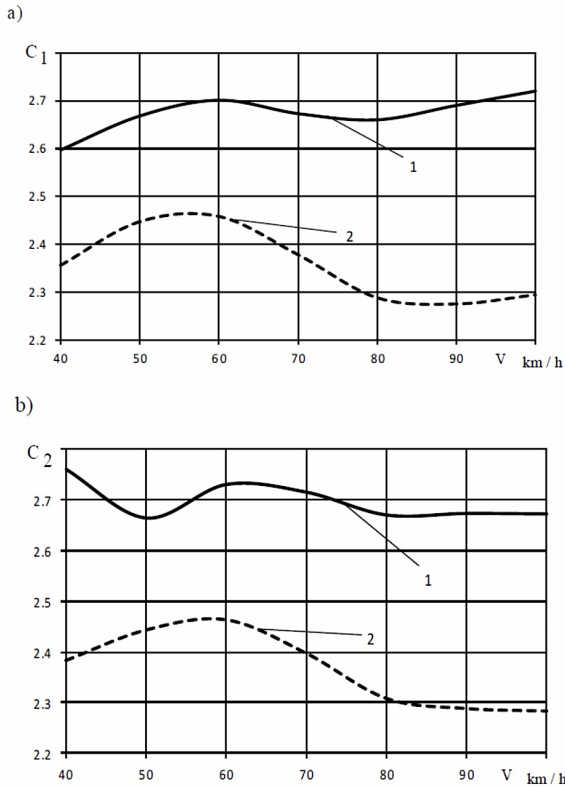


Fig.4. Dependence of the coefficient of smoothness over the first (a) and second (b) pivot points of the body on the speed, 1 - spring central spring suspension; 2 - spring suspension with electromechanical shock absorber

## 5. Discussion of the results of evaluation of indicators of dynamic quality of subway cars equipped with electromechanical shock absorbers

Comparative analysis of graphs shows that the indicators of the dynamic quality of the model of the subway car with both spring and pneumatic spring suspension do not exceed their allowable values up to the speed $v = 100$ km/h.

According to the obtained dependences, it is possible to note that this maximum was especially manifested in the speed range $v = 60$-$70$ km/h on the dependences $C_{1,2}(v)$, $\ddot{z}_{k2}(v)$ and $K_{dc3,4}(v)$.

In Fig. 4-8 we see that the model with spring suspension with electromechanical shock absorber has better performance compared to spring suspension. For example, the plots in Fig. 4 show that this is the case for such an indicator as the coefficient of smoothness $C_1$ for suspension with electromechanical shock absorber (curve 2) in the entire speed range $v = 40$-$100$ km/h, the value of this coefficient is 9.2-15.6% less than for spring suspension (curve 1), and for $C_2$ - at 13.3-14.5%.

According to the value of the maximum acceleration (Fig. 5) on the floor of the body above the first pivot point, the difference between

a)



b)
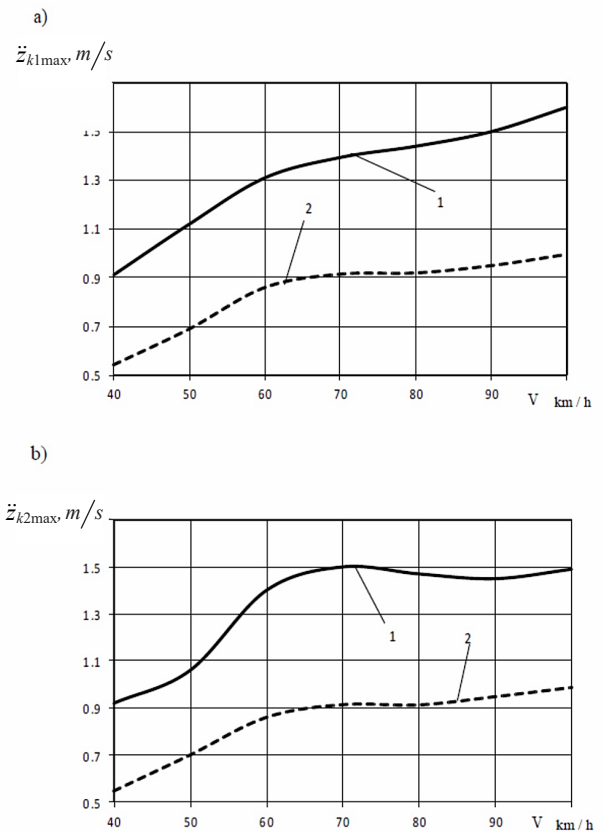


Fig. 5. Dependence of the maximum acceleration over the first (a) and second (b) pivot points of the body on the speed of movement. 1 - spring central suspension; 2 - spring suspension with electromechanical shock absorber

a)



b)
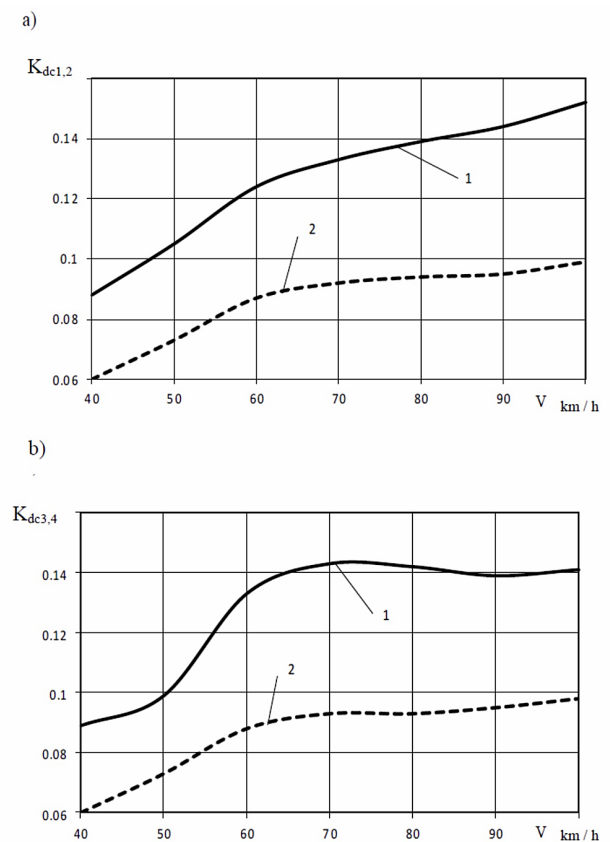


Fig. 6. Dependence of coefficients speakers in the central stage of suspension on the first (a) and the second (b) carts from the speed of movement. 1 - spring central suspension; 2 - spring suspension with electromechanical shock absorber

a)

b)
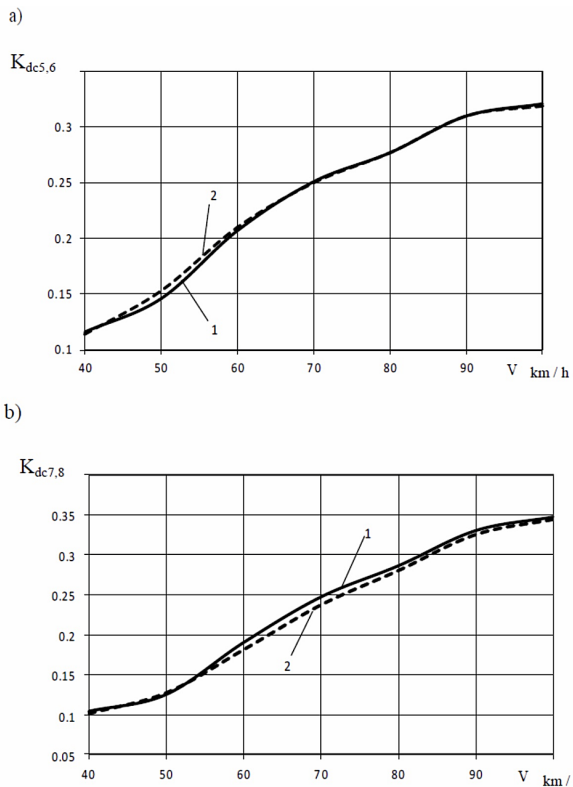
Fig. 7. Dependence of the coefficients of dynamics in the axle stage of suspension of the first wheel pair (a) and the second wheel pair (b) of the first wheelchair from the speed of movement. 1 - spring central suspension; 2 - spring suspension with electromechanical shock absorber

a)

b)
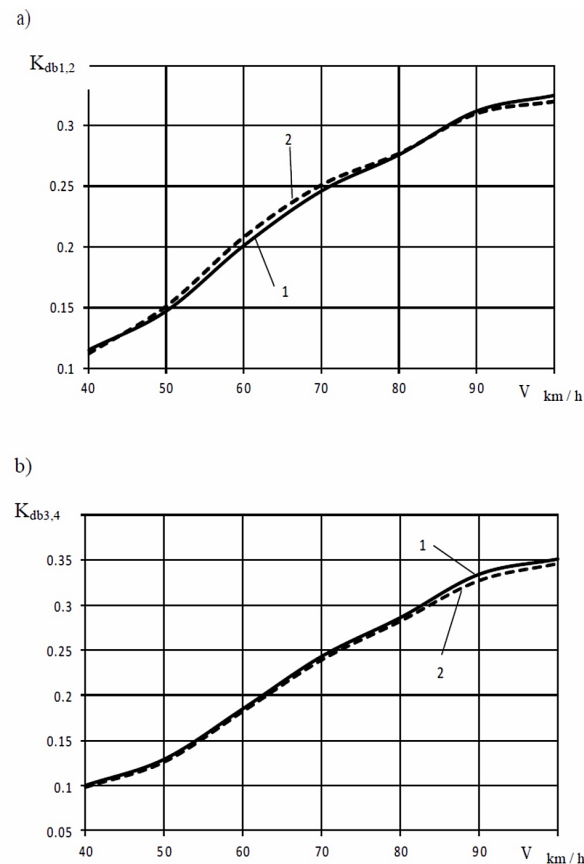
Fig. 8. Dependence of coefficients dynamics in the axle stage of suspension of the first wheel pair (a) and the second wheel pair (b) of the second trolley from the speed of movement. 1 - spring central suspension; 2 - spring suspension with electromechanical shock absorber

the spring and electromechanical suspension shock absorber is in the whole range of speeds 40.4-37.7%, and for - 40.5-33.8%.

The value of the coefficients of dynamics in the central stage with air suspension $K_{dc1,2}$ (Fig. 4) is less by 31.8-34.9% compared to the spring suspension, and for the coefficient of dynamics $K_{dc3,4}$ - by 32.6-30.5% .

The coefficients of dynamics in the axle box stage of spring suspension $K_{db1-8}$ (Fig. 7, Fig. 8) with the same designs of this stage for carts with spring spring suspension and spring suspension with electromechanical shock absorber in this speed range, differ little and, as can be seen from the graphs, do not exceed the permissible value $[K_{db}]$ = 0.35 (except $K_{db3.4}$ for spring central suspension).

According to the results of modeling the operation of the spring suspension with an electromechanical shock absorber, the average power generated by the shock absorber depending on the speed for random irregularity is determined (Fig. 9).

The average power of an electromechanical shock absorber when excited by a random roughness is much greater than with a sinusoidal roughness given in [12], which is due to the higher values of random accelerations and velocities that it dampens. Yes, the average power at speed 100 km/h is 189 watts, and at 60 km/h - 46 watts.

The results were obtained when the value of the variance of the equivalent geometric-irregularity of the rail path of the subway tunnel is 13.01 mm$^2$.

Thus, the results of the calculations should be used for subway cars with spring suspension based on electromechanical shock absorbers.

The method of determining the maximum values of dynamic forces in the elements of the spring suspension, displacements and accelerations of the pivot points of the body on the average value of their absolute maximum is more correct than estimating the value of $3S_q$ . This is due to the fact that the value depends on such important characteristics of the random process $q(t)$ as $t_p$ and $f_e$, and the values of the three standard deviations are clearly not affected by these characteristics.
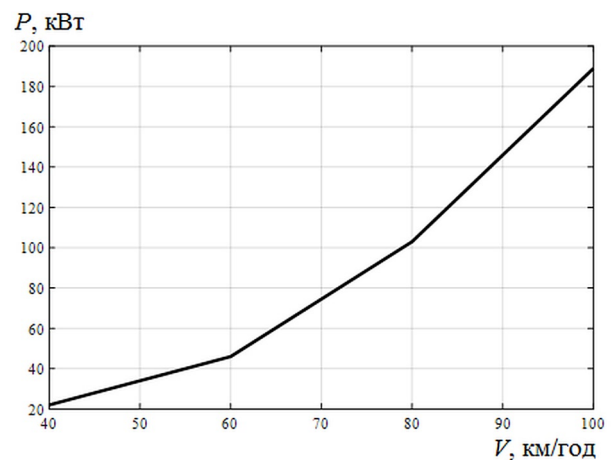


Fig. 9. Graphs of the dependences of the average generated power on the speed of motion at a random excitation irregularity

The presented methods and results of calculation of indicators of dynamic quality of spring suspension of a running gear of the subway car with electromechanical shock-absorbers allow to receive criteria of efficiency of work of shock-absorbers at various modes.

## 6. Conclusions

1. A method for the synthesis of random forced vertical oscillations of a dynamic model of a subway car, based on sliding summation and taking into account the power spectral density obtained by processing the test results of a subway car model 81.717, is adapted and applied.
2. The technique on definition of indicators of dynamic quality of the metro car with electromechanical shock-absorbers which

is based on modeling of processes of movement of the subway car on a section of a way with casual roughness and definition, by results of modeling, indicators of dynamic quality is created.

3. The study of forced random oscillations of the dynamic model of the subway car showed that the spring suspension based on electromechanical shock absorbers has a significant advantage over the central spring suspension for all quality indicators in the entire range of speeds. It is proposed to conduct a comparative analysis in the range of speeds from 40 to 100 km/h.

4. The dependence of the average power generated by the electromechanical shock absorber at different speeds is established. It is noted that, unlike air suspension, the electromechanical shock absorber not only does not require additional losses for the operation of the compressor, but also generates electricity. Therefore, analyzing the results of the calculations, it is recommended to use them for the design of a subway car with spring suspension based on electromechanical shock absorbers.

## References

1. Bogsjö K, Podgorski K, Rychlik I. Models for road surface roughness, Vehicle System Dynamics 2012,(50):725–747, https://doi.org/10.1080/00423114.2011.637566.
2. Gołębiowski P, Kukulski J. Preliminary study of shaping the railway track geometry in terms of their maintenance costs and capacity. Archives of Transport 2020; 53(1): 115–128, https://doi.org/10.5604/01.3001.0014.1787
3. Gysen B L J, Sande van der T P J, Paulides J J H, Lomonova E. Efficiency of a regenerative direct-drive electromagnetic active suspension. IEEE Transactions on Vehicular Technology 2011; 60(4): 1384-1393, https://doi.org/10.1109/TVT.2011.2131160
4. Jacyna M, Semenov I. Models of vehicle service system supply under information uncertainty. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (4): 694–704, http://dx.doi.org/10.17531/ein.2020.4.1.
5. Kardas-Cinal E, Spectral distribution of derailment coefficient in non-linear model of railway vehicle–track system with random track irregularities. ASME Journal of Computational and Nonlinear Dynamics 2013; 8(3):031014, https://doi.org/10.1115/1.4023352.
6. Kardas-Cinal E. Investigation of ride comfort in a railway vehicle in the presence of random track irregularities. Archives of Transport 2006; 18 (1): 5-16.
7. Kardas-Cinal E. Statistical Method for investigating transient enhancements of dynamical responses due to random disturbances: Application to railway vehicle motion. ASME Journal of Vibration and Acoustics-Transactions 2020; 142 (6): 061008, https://doi.org 10.1115/1.4047067.
8. Kisilowski J, Kardas-Cinal E. On a certain method of examining stability of mathematical models of railway vehicles with disturbances occurring in real objects, Vehicle System Dynamics 1994, published online 2008, 23(1):262-273, https://doi.org/10.1080/00423119308969520
9. Kryukov Railway Car Building Plant. Catalogue Passenger car manufacturing. Electronic resource 2019; [ http://www.kvsz.com/images/catalogs/tsn.pdf].
10. Kukulski J, Gołębiowski P, Makowski J, Jacyna-Gołda I, Żak J. Effective Method for Diagnosing Continuous Welded Track Condition Based on Experimental Research. Energies 2021; 14(10):2889, https://doi.org/10.3390/en14102889
11. Kukulski J, Jacyna M, Gołębiowski P.Finite Element Method in Assessing Strength Properties of a Railway Surface and Its Elements. Symmetry-Basel 2019; 8(11): 1-29, http://doi.org/10.3390/sym11081014.
12. Liubarskyi B, Lukashova N, Petrenko O, Iakunin D, Nikonov O, Matsyi O. Building a mathematical model of the oscillations in subway cars equipped with electromechanical shock absorbers. Eastern-European Journal of Enterprise Technologies 2020;6(7(108)):51-59, https://doi.org/10.15587/1729-4061.2020.217183
13. Liubarskyi B, Lukashova N, Petrenko O, PavlenkoT, Iakunin D, Yatsko S,Vashchenko Y. Devising a procedure to choose optimal parameters for the electromechanical shock absorber for a subway car. Eastern-European Journal Of Enterprise Technologies 2019; 4(5 (100)):16-25, https://doi.org/10.15587/1729-4061.2019.176304.
14. Liubarskyi B, Lukashova N, Petrenko O, Yeritsyan B, Kovalchuk Y, Overianova L. Procedure for modeling dynamic processes of the electromechanical shock absorber in a subway car. Eastern-European Journal of Enterprise Technologies 2019; 5(5 (101)): 44-52. https://doi.org/ /10.15587/1729-4061.2019.181117.
15. Lozia Z, Kardas-Cinal E. The use of a linear half - vehicle model for the optimization of damping in the passive suspension system of a railway car. Archives of Transport 2016; 39(3): 31-49, https://doi.org/10.5604/08669546.1225448.
16. Lukashova N, Pavlenko T, Liubarskyi B, Petrenko O. Analysis of constructions of resports lingings of rail city electric mobile composition. Control, Navigation and Communication Systems. Academic Journal 2018; 5(51): 65-68, https://doi.org/https://doi.org/10.26906/SUNZ.2018.5.065.
17. Murzin R V. Vertical oscillations of a motor car of an electric train on four uniaxial trolleys and the choice of parameters of its spring suspension. Doctoral dissertation, Russian University of Transport 2003:110 pages.
18. Ruban V G, Matva A M. Solution of problems of dynamics of railway cars in the Mathcad package: textbook. Rostov State Transport University. Rostov 2009; 99 p.
19. Serdobintsev E V, Zvantsev P N, Han Y. Choice of parameters for a metro coach with pneumatic springs. World of Transport and Transportation 2014, 1: 34-41 (in Russian)
20. Serdobintsev E V, Han Y. Vertical Oscillations of the Metro Wagon with Pneumatic Suspension. World of Transport and Transportation 2013; 2: 78–84 (in Russian).
21. Sergienko A N. Mathematical model of oscillations in the running system of a car with electromagnetic damping. Bulletin of the National Technical University "KhPI". Collection of scientific works. Series: Transport Engineering 2013;  – Kh.: NTU "KhPI". ss.31 (1004).  P. 86–93.
22. Standards for calculation and assessment of strength of load-bearing elements and dynamic qualities and impact on the track of the crew part of the locomotives of the railways of the Ministry of Railways of the Russian Federation gauge 1520 mm. Moscow: MPS of Russia 1998; 145 p.
23. Sulym A O, Fomin O V, Khozia P O, Mastepan A G.Theoretical and practical determination of parameters of on-board capacitive energy storage of the rolling stock.  Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu 2018; 5:79-87 https://doi.org/10.29202/nvngu/2018-5/8.

24. Szaciłło L, Jacyna M, Szczepański E, Izdebski M. Risk assessment for rail freight transport operations. Eksploatacja i Niezawodnosc –Maintenance and Reliability 2021; 23 (3): 476–488, http://doi.org/10.17531/ein.2021.3.

25. Vaskovskyi Yu M, Poda M V. Energy efficiency assessment for energy recovery systems of mechanical vibrations of vehicles. Bulletin of NTU Kharkiv Polytechnic Institute Series Electrical Machines and Electromechanical Energy Conversion 2020; 3 (1357): 52-55, https://doi.org/10.20998/2409-9295.2020.3.09.

# Global non-probabilistic reliability sensitivity analysis based on surrogate model

## Hui Liu[a], Ning-Cong Xiao[b],*

[a] Chengdu University of Traditional Chinese Medicine, School of intelligent medicine, No. 1166, Liutai Avenue, Wenjiang District, Chengdu 611137, China
[b] University of Electronic Science and Technology of China, School of Mechanical and Electrical Engineering, No. 2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu 611731, China

## Highlights

- Global non-probabilistic sensitivity analysis method is proposed.

- The proposed method considers main effect of interval variables.

- The proposed method is easy to use and does not require probability distribution.

## Abstract

Sensitivity analysis is used to find the key variables which have significant effect on system reliability. For a product in early design stage, it is impossible to collect sufficient samples. Thus, the probabilistic-based reliability sensitivity analysis methods are difficult to use due to the requirement of probability distribution. As an alternative, interval can be used because it only requires few samples. In this study, an effective global non-probabilistic sensitivity analysis based on adaptive Kriging model is proposed. The global accuracy Kriging model is constructed to reduce overall computational cost. Subsequently, the global non-probabilistic sensitivity analysis method is developed. Compared to existing non-probabilistic sensitivity analysis methods, the proposed method is a global non-probabilistic reliability sensitivity analysis method. The proposed method is easy to use and does not require probability distribution of the input variables. The applicability of proposed method is demonstrated via two examples.

## Keywords

structural reliability; non-probabilistic reliability sensitivity; adaptive Kriging; surrogate models; interval variables.

## 1. Introduction

In reliability engineering, sensitivity analysis is widely used to find the key variables that have significant effect on system reliability [1, 13]. In general, reliability sensitivity can be classified into two types [11, 17]: local reliability sensitivity and global reliability sensitivity.

Nowadays, probability distribution is often used to represent random parameter, and most of probabilistic-based reliability sensitivity analysis methods have been reported. For example, Proppe [14] presented a local reliability-based sensitivity analysis based on moving particles method. Cadini et al. [1] proposed an adaptive Kriging importance sampling-based method for global sensitivity analysis. Dubourg and Sudret [2] proposed Kriging model-based importance sampling for reliability sensitivity analysis, which can be used in reliability-based design optimization. It should be noted that many samples are required to accurately determine a probability distribution. However, it is impossible to collect enough samples for a product in early design stage. As an alternative, interval or convex model can be used because they only require a few samples, which are useful for a product in early design state. If input variable is represented using interval or convex model, the sensitivity problem is called as non-probabilistic sensitivity in this study. Until now, a few research efforts for non-probabilistic sensitivity analysis have been reported. For example, Li et al. [9] presented the definition of non-probabilistic sensitivity analysis, and optimization-based method is suggested to solve complex non-probabilistic sensitivity problems. Xiao et al. [19] proposed a non-probabilistic sensitivity analysis method under considering correlations among interval variables. Qiao et al. [15] proposed a non-probabilistic reliability sensitivity analysis method based on convex model. Wang et al. [18] used non-probabilistic sensitivity analysis for optimization of aeronautical hydraulic pipelines.

It should be noted that existing non-probabilistic reliability sensitivity methods are, generally, local reliability sensitivity methods. Moreover, performance functions in real applications are typically implicit functions involving time-consuming simulations. Subsequently, non-probabilistic reliability sensitivity analysis involving simulation is extremely computationally expensive. To address these issues, a global non-probabilistic reliability sensitivity analysis based on adaptive Kriging model is proposed in this study. Kriging models have been widely used in reliability engineering in recent years[3, 12, 20]

(*) Corresponding author.
E-mail addresses: N.C. Xiao (ORCID: 0000-0003-0493-1251): ncxiao@uestc.edu.cn

[21, 22]. Therefore, to significantly reduce computational burden for systems with time-consuming simulations, the global accuracy Kriging model is constructed adaptively that can be used to replace simulations. Subsequently, a global non-probabilistic reliability sensitivity analysis method is developed based on Sobol's sensitivity indices [16] for systems with interval variables.

This paper is structured as follows. Section 2 reviews of existing non-probabilistic reliability analysis. Section 3 presents a global non-probabilistic sensitivity analysis method in detail. Two numerical examples are used to demonstrate the applicability of proposed method in section 4. Section 5 is the conclusion.

## 2. Review of existing non-probabilistic reliability analysis

### 2.1. Non-probabilistic reliability analysis

Interval variable can be used to represent random parameter. An interval variable is expressed as:

$$X^I = \left[ X^l, X^u \right] = \left\{ X \mid X^l \le X \le X^u \right\} \tag{1}$$

where $X^l$ and $X^u$ are, respectively, the lower and upper bounds of interval variable. The midpoint and radius of $X^I$ can be, respectively, calculated as follows:

$$\bar{X} = \frac{X^u + X^l}{2}, X^r = \frac{X^u - X^l}{2} \tag{2}$$

Suppose that the performance function of a system is $Z = g\left( X^I \right)$ with interval variables $X^I = \left( X_1^I, X_2^I, \cdots, X_n^I \right)$, then the system response $Z$ is also an interval variable. The non-probabilistic reliability index is defined as follows [8]:

$$\eta = \frac{\bar{Z}}{Z^r} \tag{3}$$

Based on Eq. (3), it shows that if $\eta > 1$, the system is safe; if $\eta < -1$, the system is failure; $-1 \le \eta \le 1$ means that system is in uncertain state. Eq. (3) can be rewritten as:

$$\eta = \frac{\left( Z^u + Z^l \right)}{\left( Z^u - Z^l \right)} \tag{4}$$

where $Z^u, Z^l$ are, respectively, the lower and upper bounds of system response. To accurately calculate $Z^u, Z^l$, the following optimization model can be solved:

$$Z^l / Z^u = \min/\max g\left( X \right)$$
$$\begin{cases} s.t. \\ X^l \le X \le X^u \end{cases} \tag{5}$$

In general, for a system with multiple components, the system non-probabilistic reliability index can be computed as :

$$\eta_{sys} = \begin{cases} \min\left( \eta_1, \eta_2, \cdots, \eta_m \right), \text{for a series system} \\ \max\left( \eta_1, \eta_2, \cdots, \eta_m \right), \text{for a parallel system} \end{cases} \tag{6}$$

where $\eta_i$ is non-probabilistic reliability index of the $i$th component.

### 2.2. Non-probabilistic sensitivity reliability analysis

Reliability sensitivity analysis is useful because it can be used to find the key variable that has significant effect on system reliability. Traditionally, non-probabilistic reliability sensitivity is defined as follows [9]:

$$\left( \frac{\partial \eta}{\partial \bar{X}}, \frac{\partial \eta}{\partial X^r} \right) \tag{7}$$

From Eq. (7), it is easy to know that existing non-probabilistic reliability sensitivity is a local sensitivity measure. Note that there is no direct relationship between non-probability reliability index $\eta$ and interval parameters $\bar{X}$ and $X^r$. Thus, analytical solution for the non-probabilistic reliability sensitivity is, generally, impossible, except for some special cases such as performance function $Z = g\left( X^I \right)$ is a linear function. To approximately estimate non-probabilistic reliability sensitivity in Eq. (7), the finite difference technique can be used as:

$$\begin{cases} \dfrac{\partial \eta}{\partial \bar{X}} = \dfrac{\eta\left( \bar{X} + \Delta \bar{X} \right) - \eta\left( \bar{X} \right)}{\Delta \bar{X}} \\ \dfrac{\partial \eta}{\partial X^r} = \dfrac{\eta\left( X^r + \Delta X^r \right) - \eta\left( X^r \right)}{\Delta X^r} \end{cases} \tag{8}$$

where $\Delta \bar{X}$ and $\Delta X^r$ are very small variations of interval midpoint and radius, respectively. Note that for system non-probabilistic reliability sensitivity problem, Eq. (7) can be revised as follows:

$$\left( \frac{\partial \eta_{sys}}{\partial \bar{X}}, \frac{\partial \eta_{sys}}{\partial X^r} \right) \tag{9}$$

Using the finite difference technique, Eq. (9) can be calculated as follows:

$$\begin{cases} \dfrac{\partial \eta_{sys}}{\partial \bar{X}} = \dfrac{\eta_{sys}\left( \bar{X} + \Delta \bar{X} \right) - \eta_{sys}\left( \bar{X} \right)}{\Delta \bar{X}} \\ \dfrac{\partial \eta_{sys}}{\partial X^r} = \dfrac{\eta_{sys}\left( X^r + \Delta X^r \right) - \eta_{sys}\left( X^r \right)}{\Delta X^r} \end{cases} \tag{10}$$

## 3. Proposed global non-probabilistic sensitivity analysis method

### 3.1. Global non-probabilistic sensitivity analysis

Since non-probabilistic reliability sensitivity in Eq. (7) is a local sensitivity measure, a global non-probabilistic reliability sensitivity analysis method is proposed in this study. Based on Sobol's indices[16], the proposed global non-probabilistic reliability sensitivity measure for a component is defined as:

$$S_{X_j^I} = \frac{\text{Var}_{X_j^I}\left( \eta|_{x_j} \right)}{\sum_{i=1}^{n} \text{Var}_{X_i^I}\left( \eta|_{x_i} \right)} \tag{11}$$

where $\mathrm{Var}_{X_j^I}\left(\eta\big|_{x_j}\right)$ is computed as follows:

$$\mathrm{Var}_{X_j^I}\left(\eta\big|_{x_j}\right)=\frac{1}{N}\sum_{k=1}^{N}\left(\eta_{x_j^k}-\mu\right)^2 \tag{12}$$

where $\eta_{x_j^k}$ is a non-probabilistic index under interval variable $X_j^I=x_j^k$, and $\mu=\frac{1}{N}\sum_{k=1}^{N}\eta_{x_j^k}$. In this study, the range of interval variable $X_j^I$ is evenly divided into $N-1$ sub-intervals. Subsequently, the $N$ samples can be determined, i.e., $x_j^l=x_j^1\leq x_j^2\leq\cdots\leq x_j^{N-1}\leq x_j^N=x_j^u$, where $x_j^k$ is the $k$th sample. To ensure the accuracy, the $N$ is suggested as $N\geq 50$. From Eq. (11), it is easy to know that the value of $S_{X_j^I}$ belongs to interval $[0,1]$. It should be noted that the non-probabilistic reliability index under configuration of interval variable, i.e., $\eta\big|_{x_j^k}$, is not a number in some special cases. Subsequently, $\eta\big|_{x_j^k}$ should be ignored to calculate global component non-probabilistic reliability sensitivity.

Based on Eqs. (6) and (11), global non-probabilistic reliability sensitivity of a system can be calculated as follows:

$$S_{X_j^I}=\frac{\mathrm{Var}_{X_j^I}\left(\eta_{sys}\big|_{x_j}\right)}{\sum_{i=1}^{n}\mathrm{Var}\left(\eta_{sys}\big|_{x_i}\right)} \tag{13}$$

## 3.2. Construct global accuracy surrogate model based on Kriging

In real applications, the system performance functions $\{g_1,g_2,\cdots,g_m\}$ may implicit functions involving time-consuming simulations. Using Eqs. (11) and (13) for calculating global non-probabilistic reliability sensitivity is extremely computationally expensive. For example, using Eq. (11) for calculating component global non-probabilistic reliability sensitivity. Suppose that the average number of simulations for calculating $\eta_{x_j^k}$ is $C$, then the total number of simulations for global non-probabilistic reliability sensitivity analysis of all input variables is $C\times N\times n$. It is easy to know that the computational burden is extremely huge which is impossible in real applications. To reduce computational burden, a global accuracy Kriging model in whole uncertainty space is constructed to replace time-consuming simulation. Kriging is a Gaussian process[10], for an unobserved point $\boldsymbol{x}$, the kriging prediction is a normal random variable with mean value $\mu_{\tilde{g}}(\boldsymbol{x})$ and Kriging variance $\sigma_{\tilde{g}}^2(\boldsymbol{x})$ as follows:

$$\tilde{g}(\boldsymbol{x})\sim\mathbb{N}\left[\mu_{\tilde{g}}(\boldsymbol{x}),\sigma_{\tilde{g}}^2(\boldsymbol{x})\right] \tag{14}$$

For more detailed information of Kriging, please see Ref. [4]. To effectively construct a Kriging model with global accuracy, the best added training sample can be determined as follows[7]:

$$\boldsymbol{x}^*=\underset{\boldsymbol{X}^l\leq\boldsymbol{x}\leq\boldsymbol{X}^u}{\arg\max}\,\sigma_{\tilde{g}}^2(\boldsymbol{x}) \tag{15}$$

From Eq. (15), finding $\boldsymbol{x}^*$ is an optimization problem within the whole uncertainty space. Existing intelligent optimization algorithms such as genetic algorithm can be used to solve the problem, which is complex. In this study, a large number of candidate

samples are randomly generated in whole uncertainty space, i.e., $\{\boldsymbol{x}_c\}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{N_c}\}$. Subsequently, $\sigma_{\tilde{g}}^2(\{\boldsymbol{x}_c\})$ are available based on current Kriging model. The $\boldsymbol{x}^*$ can be approximately determined as follows:

$$\boldsymbol{x}^*=\arg\max\sigma_{\tilde{g}}^2(\{\boldsymbol{x}_c\}) \tag{16}$$

To terminate the process of selecting training samples, the following stopping strategy is adopted:

$$\mathrm{Rmse}(\boldsymbol{x}_{test})\leq\Delta\xi \tag{17}$$

where $\mathrm{Rmse}(\boldsymbol{x}_{test})$ is the root mean square error, $\boldsymbol{x}_{test}$ are randomly generated test samples with the number of $10\times n$, and $\Delta\xi$ is defined as:

$$\Delta\xi=\lambda\times\left|\mathrm{E}\left[g(\boldsymbol{x}_{test})\right]\right| \tag{18}$$

where $|\cdot|$ denotes absolute operator, $\mathrm{E}(\cdot)$ denotes expectation, and $\lambda$ is a small positive number such as $\lambda=0.0001$. $\mathrm{Rmse}(\boldsymbol{x}_{test})$ is calculated as:

$$\mathrm{Rmse}(\boldsymbol{x}_{test})=\sqrt{\sum_{i=1}^{n_{test}}\left[g(\boldsymbol{x}_{test}^i)-\tilde{g}(\boldsymbol{x}_{test}^i)\right]^2\bigg/n_{test}} \tag{19}$$

where $\boldsymbol{x}_{test}^i$ is the $i$th test sample, $n_{test}=10\times n$ is the number of test samples. When the stopping strategy is met, the final surrogate models $\{\tilde{g}_1,\tilde{g}_2,\cdots,\tilde{g}_m\}$ are obtained for system performance functions $\{g_1,g_2,\cdots,g_m\}$.

## 3.3. Summery of proposed method

Compared to existing non-probabilistic sensitivity analysis methods, the proposed method provides a new way for global non-probabilistic sensitivity analysis. The details of proposed method for system global non-probabilistic reliability sensitivity analysis can be summarized in Table 1.

---

Table 1. Algorithm of proposed method

1. Generate a small number of training samples $\{\boldsymbol{x}^s,\boldsymbol{z}^s\}$, where $\boldsymbol{x}^s=\left(x_1^s,x_1^s,\cdots,x_n^s\right)$ and $\boldsymbol{z}^s=\left(z_1^s,z_2^s,\cdots,z_m^s\right)$, and build initial surrogate models.

2. Select training sample using Eq. (16) to refine each surrogate model until stopping strategy in Eq. (17) is met.

3. Global accuracy surrogate models are denoted as $\tilde{\boldsymbol{g}}=\left(\tilde{g}_1,\tilde{g}_2,\cdots,\tilde{g}_m\right)$.

4. **for** $j=1$ to $n$, **do**

    $X_j^I$ is evenly divided into $N-1$ sub-intervals, and $N$ samples can be determined as $x_j^l=x_j^1\leq x_j^2\leq\cdots\leq x_j^{N-1}\leq x_j^N=x_j^u$.

    **for** $k=1$ to $N$, **do**

    Calculate non-probabilistic index $\eta_{sys,x_j^k}$ under interval variable $X_j^I=x_j^k$

    **end for**

    Calculate global non-probabilistic reliability sensitivity $S_{X_j^I}$ using Eq. (13).

    **end for**

5. Sort global non-probabilistic reliability sensitivities $\left\{S_{X_1^I}, S_{X_2^I}, \cdots, S_{X_n^I}\right\}$ in descending order.

## 4. Numerical examples

In this section, two numeral examples are investigated to show the proposed method. The first is a beam with a single failure mode and five interval variables. The second is a parallel system with two highly nonlinear performance functions. To demonstrate the proposed method, all performance functions are viewed as implicit functions to construct surrogate models.

### Example 1–a cantilever with single failure model

A cantilever, as shown in Fig. 1., is considered. The performance function is defined as [9]:

$$g\left(m_{cr}, p_1, p_2, b_1, b_2\right) = m_{cr} - p_1 b_1 - p_2 b_2$$

where $m_{cr}$ is critical limit bending moment, $p_1$ and $p_2$ are two applied loads, $b_1$ and $b_2$ are the length between applied loads and end point. All parameters are interval variables and the detailed information is shown in Table 2.
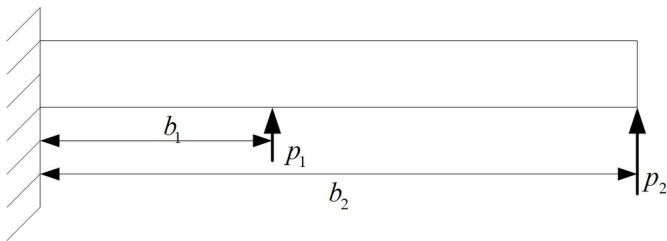


*Fig. 1 A cantilever*

In this example, 10 samples are used to build initial surrogate model, $\lambda = 0.0001$, and the number of test samples is 50. Moreover, 57 training samples are used to construct a global accuracy Kriging model. For global non-probabilistic reliability sensitivity analysis, each interval is evenly divided into 100 subintervals, respectively. The results of global non-probabilistic reliability sensitivity from proposed method are shown in Table 3.

In Table 3, the results with "*" are from the proposed method with true performance function. Based on Table 3, the results are very accurate compared to reference values based on real performance function. The proposed method provides a new way to measure global non-probabilistic reliability sensitivity. It is easy to know that the vari-

able $m_{cr}$ has a significant effect on system reliability, which should pay more attention in design stage. Note that the proposed method is quite different to local non-probabilistic reliability sensitivity method. The local method provides local sensitivity of midpoint and radius of interval variable, whereas the proposed method provides a global non-probabilistic reliability sensitivity of interval variable in whole uncertainty space. In this example, the total number of original function call is 57+50=107.

### Example 2–a parallel system with two failure modes

Suppose that a parallel system with two failure modes, and the corresponding performance functions are defined as follows:

$$\begin{cases} g_1\left(X_1, X_2\right) = \left(X_1 + 2\right)^2 - X_2 - 2 \\ g_2\left(X_1, X_2\right) = \left(X_1 - 3\right)^2 - 2X_1 X_2 + 4 \end{cases}$$

$X_1, X_2$ are two independent interval variables, $X_1 \in [0.5, 1]$, and $X_2 \in [1, 2]$.

In this example, 10 samples are used to build initial surrogate models for $g_1$ and $g_2$, respectively. $\lambda = 0.0001$, and the number of test samples for both is 20. To construct global accuracy Kriging models for $g_1$ and $g_2$, the number of training samples are 14 and 13, respectively. For system global non-probabilistic reliability sensitivity analysis, each interval is evenly divided into 100 subintervals, respectively. The results from the proposed method are shown in Table 4.

In Table 4, the results with "*" are from the proposed method with true performance functions. Based on Table 4, the results are very accurate compared to reference values based on real performance functions. The proposed method provides a new way to measure system global non-probabilistic reliability sensitivity. It is easy to know that the variable $X_1$ has a significant effect on system reliability, which should pay more attention to control it in design stage. In this example, the total number of original function calls are 34 (14+20) and 33 (13+20), respectively.

## 5. Conclusions

Sensitivity analysis is used to find the key variables which have significant effect on system reliability. For a product in early design stage, it is impossible to collect enough samples due to the limitations of time and resources. Thus, the probabilistic-based reliability sensitivity analysis methods are difficult to use because many samples are required to accurately determine a probability distribution. Existing non-probabilistic reliability sensitivity methods are local sensitivity methods. To address the issue, a new global non-probabilistic reliability sensitivity method is proposed in this study. Surrogate models with

*Table 2. Detailed information of interval variables*

| Interval variables | $p_1$ (kN) | $p_2$ (kN) | $b_1$ (m) | $b_2$ (m) | $m_{cr}$ (kN.m) |
|---|---|---|---|---|---|
| Lower bound | 4.4 | 1.7 | 1.8 | 4.5 | 32 |
| Upper bound | 5.6 | 2.3 | 2.2 | 5.5 | 40 |

*Table 3. Global non-probabilistic sensitivities of example 1*

| Sensitivity | $S_{m_{cr}}$ | $S_{p_1}$ | $S_{b_1}$ | $S_{p_2}$ | $S_{b_2}$ |
|---|---|---|---|---|---|
| Proposed method | 0.8220 | 0.0423 | 0.0300 | 0.0732 | 0.0325 |
| | 0.8214* | 0.0426* | 0.0298* | 0.0734* | 0.0328* |

*Table 4. Global non-probabilistic sensitivities of example 2*

| Sensitivity | $S_{X_1}$ | $S_{X_2}$ | $S_{X_1}$ | $S_{X_2}$ |
|---|---|---|---|---|
| Proposed method | 0.9537 | 0.0463 | 0.9541* | 0.0459* |

global accuracy are constructed with adaptive manner to reduce overall computational burden. Subsequently, time-consuming simulations can be replaced by constructed surrogate models, which are much cheaper than simulations. Numerical examples have demonstrated the applicability of proposed method, which provides a new way for global non-probabilistic reliability sensitivity analysis.

Compared to probabilistic-based global reliability sensitivity methods, the major advantage of proposed method is that the distribution type of input variables is not required. Note that the proposed method is different to local non-probabilistic reliability sensitivity methods. The local me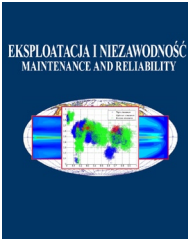thod provides local sensitivity of midpoint and radius of interval variables, whereas the proposed method provides global non-probabilistic reliability sensitivity of interval variables in whole uncertainty space. Moreover, the proposed method only considers main effect of input interval variables.

## References

1. Cadini F, Lombardo S S, Giglio M. Global reliability sensitivity analysis by Sobol-based dynamic adaptive kriging importance sampling. Structural Safety, 2020; 87: 101998. https://doi.org/10.1016/j.strusafe.2020.101998
2. Dubourg V, Sudret B. Meta-model-based importance sampling for reliability sensitivity analysis. Structural Safety, 2014; 49: 27-36. https://doi.org/10.1016/j.strusafe.2013.08.010
3. Echard B, Gayton N, Lemaire M. AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. Structural Safety, 2011; 33: 145-154. https://doi.org/10.1016/j.strusafe.2011.01.002
4. Forrester A I J, Sóbester A, Keane A J. Engineering design via surrogate modelling. Chicheste: John Wiley & Sons, 2008.
5. Guo J, Du X. Reliability sensitivity analysis with random and interval variables. International Journal for Numerical Methods in Engineering, 2009; 78(13): 1585-1617. https://doi.org/10.1002/nme.2543
6. Jiang C, Qiu H, Li X, et al. Iterative reliable design space approach for efficient reliability-based design optimization. Engineering with Computers, 2020; 36(1): 151-169. https://doi.org/10.1007/s00366-018-00691-z
7. Jin S S, Jung H J. Sequential surrogate modeling for efficient finite element model updating. Computers and Structures, 2016; 168: 30-45. https://doi.org/10.1016/j. compstruc. 2016. 02. 005
8. Guo S X, Lu Z Z, Feng Y S. A non-probabilistic model of structural reliability based on interval analysis. Chinese Journal of Computational Mechanics, 2001; 18(1): 56-60. (In Chinese)
9. Li G J, Lu Z Z, Wang P. Sensitivity analysis of non-probabilistic reliability of uncertain structure. Acta Aeronautica et Astronautica Sinica, 2011; 32: 1-7. (In Chinese)
10. Li Y H, Liang X J, Dong S H. Reliability optimization design method based on multi-level surrogate model. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (4): 638–650. http://dx.doi.org/10.17531/ein.2020.4.7
11. Lu Z Z, Li L Y, Song S F, et al. Importance analysis and solution method of uncertain structure system. Beijing: Science Press, 2015. (In Chinese)
12. Meng Z, Zhang Z H, Zhang D Q, et al. An active learning method combining kriging and accelerated chaotic single loop approach (AK-ACSLA) for reliability design optimization. Computer Methods in Applied Mechanics and Engineering, 2019; 357: 112570. https://doi.org/10.1016/j.cma.2019.112570
13. Papaioannou I, Straub D. Variance-based reliability sensitivity analysis and the FORM α-factors. Reliability Engineering and System Safety, 2021; 210: 107496. https://doi.org/10.1016/j.ress.2021.107496
14. Proppe C. Local reliability based sensitivity analysis with the moving particles method. Reliability Engineering and System Safety, 2021; 207: 107269. https://doi.org/10.1016/j.ress.2020.107269
15. Qiao X Z, Su Q W, Li L, et al. Non-probabilistic reliability sensitivity analysis based on convex model. Journal of Mechanical Strength, 2019; 41(4): 895-900. (In Chinese)
16. Sobol I M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 2001; 55(1-3): 271-280. https://doi.org/10.1016/S0378-4754(00)00270-6
17. Torii A J, Novotny A A. A priori error estimates for local reliability-based sensitivity analysis with Monte Carlo simulation. Reliability Engineering and System Safety, 2021; 213: 107749. https://doi.org/10.1016/j.ress.2021.107749
18. Wang W X, Zhou C C, Gao H S, et al. Application of non-probabilistic sensitivity analysis in the optimization of aeronautical hydraulic pipelines. Structural and Multidisciplinary Optimization, 2018; 57: 2177-2191. https://doi.org/10.1007/s00158-017-1848-4
19. Xiao N C, Huang H Z, Li Y F, et al. Non-probabilistic reliability sensitivity analysis of the model of structural systems with interval variables whose state of dependence is determined by constraints. Journal of Risk and Reliability, 2013; 227(5): 491-498. https://doi.org/10.1177/1748006X13480742
20. Xiao N C, Yuan K, Zhou C. Adaptive kriging-based efficient reliability method for structural systems with multiple failure modes and mixed variables. Computer Methods in Applied Mechanics and Engineering, 2020; 359:112649. https://doi.org/10.1016/ j.cma.2019.112649
21. Xiao N C, Yuan K, Zhan H Y. System reliability analysis based on dependent Kriging predictions and parallel learning strategy. Reliability Engineering and System Safety, 2022; 218: 108083. https://doi.org/10.1016/j.ress.2021.108083
22. Yang M D, Zhang D Q, Wang F, et al. Efficient local adaptive Kriging approximation method with single-loop strategy for reliability-based design optimization. Computer Methods in Applied Mechanics Engineering, 2022; 390: 114462. https://doi.org/10.1016/j.cma.2021.114462

# Impact of suspension and route stabilization on dynamic parameters of self-driven mine suspended monorails

Indexed by:
Web of Science Group

Kamil Szewerda[b,*], Jarosław Tokarczyk[b], Jerzy Świder[a], Aneta Grodzicka[c]

[a] Silesian University of Technology, Department of Engineering Processes Automation and Integrated Manufacturing Systems, Faculty of Mechanical Engineering, ul. Konarskiego 18A, 44-100 Gliwice, Poland
[b] KOMAG Institute of Mining Technology, ul. Pszczyńska 37, 44-101 Gliwice, Poland
[c] Silesian University of Technology, Faculty of Mining, Safety Engineering and Industrial Automation, ul. Akademicka 2, 44-100 Gliwice, Poland

## Highlights

- An analysis of potential increase in speed of suspended monorail is presented.

- An analysis of the impact of route stabilization on dynamic loads of the support is included.

- An analysis of the impact of route stabilization on forces affecting the operator is included.

- The impact of dynamic interactions of the monorail on the biomechanical parameter is assessed.

## Abstract

Impact of the method of suspension and route stabilization of suspended monorail on forces loading the roadway roof support system is presented. This is important in the context of possible increasing the speed of monorails during personnel movement. Nature of load and displacement of the route, as well as deceleration of the transport set, with a dynamic excitation - an emergency braking of the transport set, are presented. The results are presented for seven configurations of slings and lashings stabilizing the route. The Head Injury Criterion (HIC), recorded using the Articulated Total Body (HYBRID III) model, during the impact of operator's cabin against an obstacle, is presented in the further part of the article. Analyzes are aimed at developing the guidelines to ensure safety of mining personnel (without exceeding the accepted overloads) and mining infrastructure (without exceeding the maximum accepted load of the roadway support) during operation of the suspended monorail at higher speed. Analyzes are the result of the authors numerical simulations.

## Keywords

coal mining, safety, suspended monorails, transport, numerical simulations

## 1. Introduction

Mine transport is an indispensable link in the mining process of any mining plant. While floor transport has been used since the beginning of mining development, and mine tracks appeared in the 17th century, the use of suspended transport started only in the middle of the 20th century [13]. The suspended monorails quickly became widely used underground mean of transport. The development of self-driven suspended monorails resulted in an increase in the load-bearing capacity and strength of rails and transport sets, as well as an increase in tractive force, compared to suspended cable-driven suspended monorails [12, 13]. Intensive development of suspended monorails is still observed today. This is proved by R&D work in development of electrically powered monorails [9, 13], mechanical modifications, including the possibility of travelling at higher speed or improving the comfort of travelling [21, 24]. The changes also involve introduction of innovative mechatronic solutions in monorails [1, 6, 22]. Along with the development of suspended monorails, the development of monorail routes and the methods of their suspension and stabilization gradu-

ally advanced. In self-driven monorails, their routes were standardized. Currently the most frequently used rails are those with the I155 profile, less often with the I140V profile. One of the "tasks" of the suspended monorail system is to transfer the loads, resulting from the weight of monorail unit and its movement, to the roadway roof support or directly to the rock mass, thus slings and lashings are used. The rails in the roadway are suspended on various types of slings, which, depending on the structure, can be loaded with a maximum force of 40 kN [20]. Depending on the length of rails and the configuration of slings and lashings, this force is transferred in different ways to the arches of yielding roadway roof support, which protects the transport routes against falling rocks. Moreover, to ensure stability, a suspended route must always be stabilized at the turns and the dips. It is recommended to stabilize straight sections, if necessary, at least at every 100 meters. Proper stabilization of the suspended route enables transferring the forces from the moving suspended monorail unit in the longitudinal and transverse directions. This often takes place on inclined routes, when dynamic forces dominate. Emergency braking of a transport set, during which the actuating system of brakes is activated in

(*) Corresponding author.
E-mail addresses: K. Szewerda (ORCID: 0000-0003-2266-1371): kszewerda@komag.eu, J. Tokarczyk (ORCID: 0000-0002-8588-0179): jtokarczyk@komag.eu, J. Świder (ORCID: 0000-0002-4310-0373): jerzy.swider@polsl.pl, A. Grodzicka (ORCID: 0000-0001-5712-8230): aneta.grodzicka@polsl.pl

the shortest possible time is an example. This causes deceleration affecting operator and moved personnel, but also dynamic overloads in connectors, slings and lashings stabilizing the route of the suspended monorail. These situations may lead to damage, such as: deformation of rails, loss of the route continuity (splitting rails at joints) or breaking the slings. In turn, the above-mentioned damage may cause an accident. The probability of hazardous situations and their possible effects increase with the increase in the travel speed of a suspended monorail. According to legal regulations in the Polish hard coal mines [20], the maximum acceptable speed of a suspended monorail during the personnel movement is 2 m/s. Due to the systematically increasing length of the routes to the newly opened mining fields, the time needed for miners to travel to the workplace is also longer. This reduces availability of the mining personnel at workplaces and shortens the effective working time. This happens due to depletion of deposits in close proximity to the shafts where coal is transported to the surface. This affects the financial results of companies in the mining industry. Shortening the travel time of personnel to the longwall panel, e.g. by increasing the acceptable speed of the suspended monorail is the solution. The benefits of shortening the travel time to the workplace are presented in [32]. Reducing the travel time to the longwall face by 15 minutes and by 30 minutes on the way back for each shift at the IM-BAT MIOMNIG Co. Manisa coal mine in Eynez, Turkey, resulted in an increase in mining capacity by 1,606.95 tons per day as mentioned in that article. In the cited example, shortening of the time to reach the workplace was associated with the use of a belt conveyor adapted to personnel movement. Therefore, increasing the maximum acceptable speed of suspended monorails is justified from an economic point of view. However, apart from economic benefits, another extremely important aspect should be ensured – safety of the moved personnel. Introduction of legal changes enabling the increase of the acceptable speed must be preceded by development of the procedures for safe method of the monorail's travel. Analyzing the impact of emergency braking on operator and personnel overloads [3] and assessing the feasibility of introducing a two-stage emergency braking system[23] demonstrated the impact of increasing the speed limit on safety.

Development of guidelines for designing the monorail route on the sections that will be passed at higher speed is one of the preparation stages to increase permissible speed. This aspect is very important, because the wrong way of suspending the monorail route may result in exceeding the permissible load of the roadway roof support's arches, which in extreme cases may lead to roadway damage. Such situations are particularly dangerous during dynamic loads, e.g. during emergency braking from higher speed. The literature on the subject includes the work related to the tests on the load-bearing capacity of the yielding roof support [4], and also related to analysis of static and dynamic loads on the roadway roof support related to transport by suspended monorails [19]. The paper [18] presents the results of research work on impact of the suspended monorail speed on forces in the selected parts of the route. However, the project did not take into account the aspects related to the possibility of using different configurations of slings of the suspended monorail route, which, according to the authors, affects the forces loading the steel arches.

The authors defined the impact of the configuration of slings and lashings on the forces acting on the curves of the roof support's arches as the objective of the work presented in this article. The article presents the difference in the load to the roadway support in relation to seven variants of the configuration of slings and lashings of the suspended monorail route. The load to the route was created by simulating emergency braking from a speed of 5 m/s in a given part of the route. Depending on the configuration of the slings and stabilizing lashings, the suspended monorail route can move within a given range. During emergency braking, especially at higher speeds, the stopped train still moves forward together with the route causing acting the forces on the slings and roadway support, and the overloads affect the operator and the moved personnel. Finding the answer to the question: "How to arrange the slings and lashings of the suspended monorail route to

minimize the load to the roadway roof support arches and the overload to the people in the monorail during emergency braking at higher speed (5 m/s)?" prompted the authors to conduct a series of tests, the results of which are presented. According to the authors, the results of the tests supplement the current state of knowledge on the impact of configuration of slings and lashings of the route on roadway support and at the same time are the basis for the development of guidelines on how to properly construct the route on the sections intended for high-speed suspension of the suspended monorail.

Development of guidelines and correct configuration of slings and lashings of the route is only one of the aspects ensuring the safety of mining personnel in the operator's and passenger cabins. Additional seat belts are the second aspect of the monorail safety improvement. Currently, both the monorail operator and passengers sitting in their seats do not have the possibility of using seat belts or other measures to protect them from being injured in emergency situations, such as emergency braking or a collision with a stationary obstacle on the track. The introduction of seat belts is one of the ideas of protecting the personnel in monorails travelling at higher speeds. This aspect was the subject of further research work of the authors. Its aim was to determine the HIC describing the probability of suffering a severe or fatal head injury. The operator's cabin was tested. The virtual HYBRID III dummy was used during these numerical simulations. HYBRID III dummies are specially designed for crash tests in the automotive industry [8]. These dummies were designed in such a way as to recreate behavior of the human body during traffic accidents [2]. For each part of the dummy (body part), appropriate masses, moments of inertia and the stiffness of connections between them were defined, which corresponded to the stiffness of the joints and the muscle tension. An example of the use of dummies in safety tests in the automotive industry is analysis of the impact of velocity on the risk of injury in relation to the driver and passenger of a passenger car, presented in [33]. Other example of the aforementioned analyzes are the analyzes presented in [17] concerning the impact of position of a person in a car on the risk of injury during a road accident. Another example of the use of dummies is analysis of the impact of car seat vibrations on the comfort of a child traveling in a car, which was presented in [30]. Along with the development of computer techniques and the development of numerical computational methods, a numerical dummy model was developed, which is used for virtual crash tests and simulations aimed at analyzing the safety of a driver and passengers in new automotive solutions.

Simulations with the use of HYBRID III dummies models were also carried out in relation to mining machines used in hard coal mines. The article [28] presents the structure and the method of defining the dummy model, as well as an example of a simulation in which the HYBRID III dummy model was used during the analyzes of the floor railway operator's cabin. As regards the use of suspended monorails, the HYBRID III dummy model was used to identify hazards and to assess their effects in the event of emergency braking from 2 m/s and 4 m/s in relation to passengers traveling in one of the types of passenger cabins [27]. This article presents the results of the simulation of an impact of a monorail traveling at a speed of 5 m/s against a stationary obstacle, with the operator wearing seat belts or without the seat belts. The objective of these research work was to demonstrate the necessity to introduce seat belts as the basic and necessary equipment for the operator's cabin of monorails traveling at a speed of 5 m/s. Due to the justified efforts to increase the permissible speed of the suspended monorail, it is necessary to ensure the safety and comfort of the personnel movement while transporting people. Safety in this aspect relates both to the safety of people as well as to the safety of machines, equipment and mining infrastructure.

Developing the guidelines which determine changes in legal regulations and allow the personnel movement at higher speed in a safe way, even in the event of an emergency is a responsible and difficult task. Results of the research work presented by the authors constitute recommendations on how to configure the slings and lashings of the

route to minimize the roadway support load and overloads affecting people in the monorail in emergency situations. Ensuring these recommendations will contribute to the reliable and safe operation of suspended monorails, even at higher speeds. On the other hand, the introduction of additional components, such as seat belts, in dangerous situations, will allow to minimize injuries sustained by people traveling on the monorail, as indicated by the HIC parameter, calculated in the simulations.

## 2. Computational models of the suspended monorail and its route

Measurement of forces acting on the slings of suspended monorail route requires installation of dedicated sensors in the monitored slings. Both due to the cost and the capabilities of the measuring equipment, the number of sensors that record the forces acting on slings is limited. conditions in which suspended monorails are used are the additional difficulty during this type of research work. The greatest limitations related to tests in in-situ conditions include the fact that the measuring equipment must meet the requirements of the ATEX directive to be used in underground mine workings. Safety requirements are another aspect that limits in-situ testing. It is often difficult to find a sufficient amount of space for the installation of measuring equipment on transport routes in hard coal mines. When testing the monorails under the operating conditions, all legal restrictions related to the movement of suspended monorails, including the speed limits for these monorails, must be met. This means that it is formally impossible to carry out driving tests at a speed 5 m/s in real conditions. In addition, in accordance with the regulations governing the traffic of monorails in mining plants, each time when emergency braking trolley is activated, it is necessary to inspect it, which generates additional costs of tests in real conditions. Testing and measuring the forces in the selected slings on a real object is easier using a special test stand. Such a stand was built under the INESI project [5]. At the stand, while maintaining appropriate safety measures, meeting the regulations required in underground mine are not necessary, and the test equipment does not have to meet the stringent requirements related to the ATEX directive. The stand built for the project was used to test a new type of 4 m elongated and reinforced rails (Fig. 1). The stand also has dedicated sensors for recording the force acting on the slings of the route. On the test stand, emergency braking was tested at a speed of 5 m/s, during which the forces acting on the slings of the route were recorded. However, for economic reasons, it was not possible to modify the method of suspending the monorail route at the stand and recording the forces acting on the slings, using a different method of stabilizing the rails. According to the authors, the method of suspension and stabilization of the rails has an impact on the forces acting on the slings in the case of dynamic excitations, such as emergency braking. Taking into account the limitations, the authors decided to develop a computational model of the suspended monorail and its route, which corresponded to the configuration on the test stand.

The model of suspended monorail consisted of an operator cabin, machinery part, two gear drives, passenger cabin and emergency braking trolley. The model of the monorail route consisted of 23 straight rails, each 4 m long. The rails were placed horizontally, and to suspend them, slings in the configuration from the test stand were used. The created computational model was validated. Emergency braking from a speed of 5 m/s with the boundary conditions consistent with those on the test stand, was simulated. Validation process consisted in comparing the results of the measurements from the test stand and those of numerical simulations, and then by fine-tuning the computational model. The following parameters were analyzed in the validation process:

- acceleration, recorded in the operator cabin and in the passenger cabin (the difference in the maximum acceleration calculated by numerical method and that recorded on the test stand was 6.6% on average),
- effective value of vibrations (RMS), recorded in the operator cabin and in the passenger cabin (the difference in the effective value recorded on the test stand and calculated by numerical simulations was on average 10%),
- forces acting on the selected slings of the monorail route (the difference in maximum force in the selected sling, calculated numerically in relation to values recorded on the test stand, was on average 9%).

The detailed method of validation of the computational model and the results are included in the following sources: acceleration acting on the operator [3]; forces in route suspensions [25]; RMS acceleration [5].

After validation of the suspended monorail model, to assess the impact of configuration of the slings and the route stabilizing lashings on forces transmitted through the slings to the roof support frame, the method of suspending the route in the model was modified by defining seven variants of the computational model. Configuration of the slings on the test stand was marked as variant 6. Each variant of the monorail route suspension differed in the arrangement of the slings located directly above the rail connections and the presence and location of side lashings. Tension of the side chains was adjusted with a turnbuckle. Another, more advanced solution is the use of a yielding lashings of a specific stiffness, their changes in length under the impact of dynamic loads, reduce the peaks of the force loading the sling and the roadway roof support. This solution is definitely more expensive. However, to analyze the impact of using such an element in the lashing in variants 3 and 5, an elastic-damping element with the characteristics corresponding to available industrial solutions was introduced to the side lashings. Each variant of the suspended monorail route model is characterized by the following features:

- Variant 1 – rails are suspended on straight slings, perpendicular to the monorail route. It is the easiest option to install and at the same time the cheapest because it requires the least amount of chain for slings. Difficulties occur with the irregular pitch of the roadway roof support. A section of the route in variant 1 is shown in Fig. 2.

- Variant 2 – this variant complements variant 1 with side lashings, added to the rail No. 2 (in the initial area of the route – braking will take place after the lashing). These lashings are inclined from the horizontal upwards by an angle of 10° and have constant length, Fig. 3. The purpose of this variant is to limit the possibility of the route moving along the monorail axis. At the same time, the forces acting on the lashings during emergency braking were recorded.

- Variant 3 – both the location of the route slings and the lashings were consistent with the



*Fig. 1. View of the reinforced rail and double sling coupler built on a route intended for high-speeds [5]*
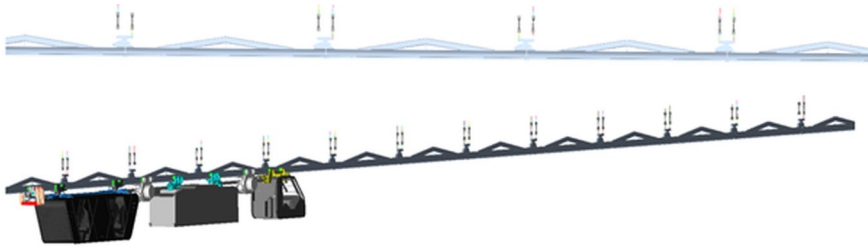
Fig. 2. A section of the suspended monorail route in variant 1, along with the suspended monorail model
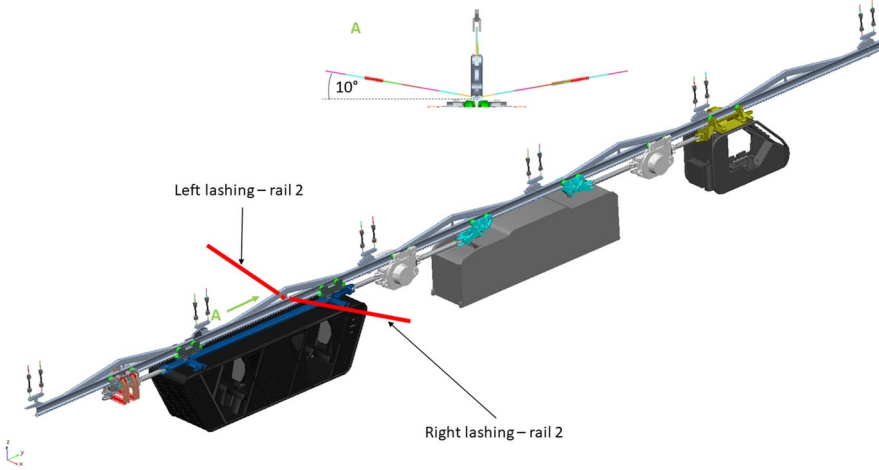


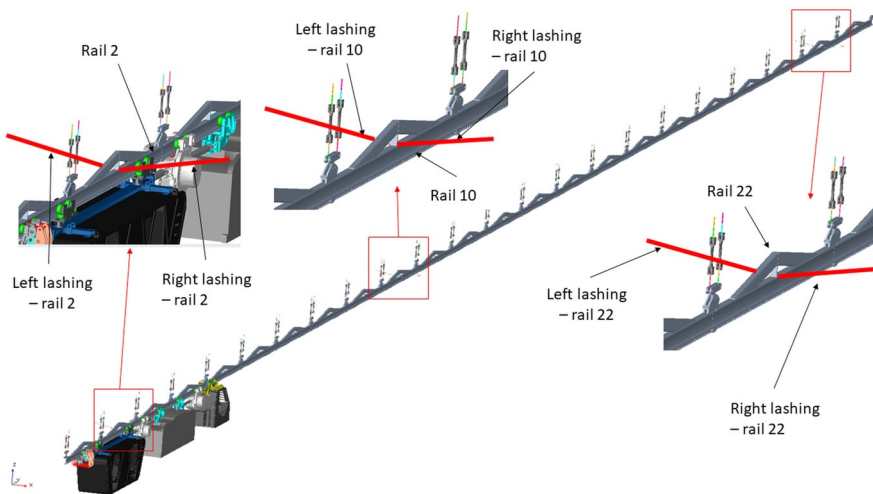Fig. 3. Section of the suspended monorail route variant 2



Fig. 4. Route stabilisation in the computational model for Variant 4

No. 10), and located at the end of the route (rail No. 22) additional slings were added. In addition, the changes in the nature of the route displacement were compared.

• Variant 5 – side lashings were modified (installed as in variant 4) by yielding. Spring elements were used with the same properties as in variant 3. The purpose of this variant was to compare the forces and decelerations during emergency braking in relation to the configuration with fixed-length lashings (variant 4). The results of this simulation may constitute an argument for the use of more expensive technical solutions in selected regions.

• Variant 6 – the computational model includes straight (perpendicular to the rails) and oblique slings placed alternately. Oblique slings were inclined at an angle of 45°; one in the direction of the monorail movement and the other in the opposite direction. This way of constructing the route was used on the test route in the INESI project [5], Fig. 5. This variant was used to validate the computational model of the suspended monorail.

• Variant 7 – all slings of the rails of the computational model are inclined at an angle of 45° in relation to the monorail route. In each pair of slings, one was deflected in the direction of the monorail movement and the other in the opposite direction, Fig. 6. The purpose of this variant is to compare the route displacement and the forces acting on the slings, in relation to the variants with side lashings.

The developed variants of the method of suspending the route of monorail enabled identifying the impact of method of suspending the route on load to the slings, and then further propagation of the load to the arches of the roadway roof support. Such analyzes are important regarding the possibility of increasing the accepted speed of suspended monorails. Situations when it is necessary to use emergency braking from higher speed seem to be dangerous. Then, sudden overloads (load peaks) may take place and that may lead to breaking of the slings and the loss of stability of roadway support. In an extreme case, in poor technical condition, roadway supports may be deformed, which leads to the destruction of the roadway and transport route [4, 18, 19]. Assessment of the impact of installation method of suspended monorail route will increase the safety of mine personnel and will enable specifying the guidelines for the route installation on the sections with increased accepted speed.

## 3. Numerical simulations

As part of the research work, the MultiBody System (MBS) simulation method was used in numerical simulations related to the dynamics of the presented model. In this method, on the basis of defined geometric constraints and defined excitations, the kinematic and dynamic quantities during the analyzed system operation, in the discussed case constituting the suspended monorail assembly with its route, were calculated. In dynamic simulations, the initial conditions in the form of known positions and velocities of all bodies, as well as information about the time processes of forces acting on the bodies, are the input data. Determination of motion of the MBS under the impact

variant 2. The difference was the yielding of the lashings mounted on the rail 2 by introducing an elastic-damping element, enabling the lashings to be extended under the impact of external force. The modulus of elasticity in relation to the side lashings was $6.66 \times 10^6$ N/m. The purpose of this variant was to compare the forces and decelerations during emergency braking in relation to the configuration with fixed-length lashings (variant 2).

• Variant 4 – additional side lashings, installed on rails 10 and 22, were added to the calculation model of the suspended monorail route. The lashings were mounted in the same manner as in the previous variants (10° angle from the horizontal), Fig. 4. To identify differences in the load to the lashing located at the beginning of the route (rail No. 2), in the area of emergency braking (rail
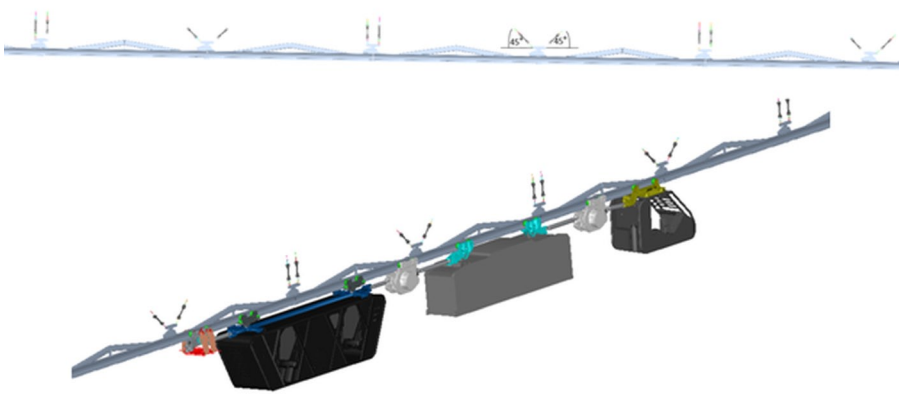
*Fig. 5. View of the suspended route in variant 6*
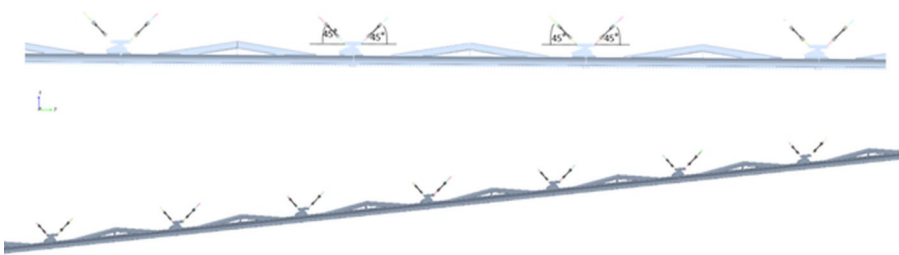


*Fig. 6. Route suspensions in the computational model in Variant 7*

- D – number of guiding constraints.

The solution of the formulated system of equations results in the calculation of positions, velocities, accelerations of each solid, as well as forces and moments acting in the computational model. In the numerical simulations, a computational model was used, which had previously been validated.

The simulation was the same for all variants. Emergency braking was from a speed of 5 m/s. In each variant of the route suspension, the suspended monorail was accelerated to speed of 5 m/s. Then, for approx. 1 s, the speed was constant, then emergency braking started due to the activation of two pairs of jaws in the brake trolley. An example of the monorail speed chart is shown in Fig. 7.

Until the commencement of emergency braking, the speed of the monorail was the same in all simulations. Depending on the suspension method, the route could move along the axis in the direction of the monorail movement. As a result of these movements, the curve in the speed diagram may oscillate depending on the route suspension variant.

of the forces applied to it as well as the reaction forces, in particular geometric constraints, is the result of solving the dynamics problems. From a mathematical point of view, to solve the problem of dynamics, solving the system of differential-algebraic equations is needed. In the software environment (MSC.ADAMS), the system of equations of motion of a mechanical system is formulated based on the Euler - Lagrange's equation (1) [26, 31]:



*Fig. 7. The course of travel speed of suspended monorail unit in the MBS simulation*

## 4. Results of numerical simulations

In a result of the simulations, time processes of various quantities and parameters were obtained.

This article provides the following results:
1. Curve of acceleration of the transport unit during travel and emergency braking.
2. Forces in the slings of the route.
3. Displacement of rails.

Ad. 1) acceleration of the transport unit during travel and emergency braking.

Acceleration and in particular the deceleration, affects other parameters, i.e. the forces in the route suspensions and the route displacements. The greatest changes in the acceleration of the transport set may occur in a situation other than typical operating conditions, e.g. impact loading due to rockfall, hitting an obstacle or (most often) during emergency braking. In the case of dangerous situation emergency braking starts [14]. Although, according to Annex 4 to the Regulation [20], the deceleration cannot exceed 10 m/s$^2$, emergency operation of the braking system results in dynamic overloads, affecting both the suspended route, the frame of roadway support, and most of all everything on the operator and passengers of the suspended

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} + \Phi_q^T \lambda = Q \qquad (1)$$

where:
- L – Lagrange function (2), i.e. the difference between the kinetic energy T and the potential energy V of the system:

$$L = T - V \qquad (2)$$

- $\lambda$ – vector of Lagrange multipliers,
- Q – vector of generalized forces acting on MBS (3):

$$Q = Q(q,\dot{q},t) \qquad (3)$$

- $\Phi$ – vector of left sides of constraints equations (after elimination of redundant constraints) (4):

$$\Phi(q,t) = \begin{bmatrix} \Phi^K(q) \\ \Phi^D(q,t) \end{bmatrix} = 0_{Nx1} \qquad (4)$$

- q – vector of generalized coordinates,
- K – number of kinematic pairs in the system,

| Variant | Acceleration |
|---------|--------------|
| 1. |  |
| 2. |  |
| 3. |  |
| 4. |  |
| 5. |  |

Fig. 8. Acceleration curves for each variant

monorail. Fig. 8 shows the acceleration curves of the transport set in each simulation variant. Maximum and minimum accelerations for each of the analysed variants are shown in Fig. 9.
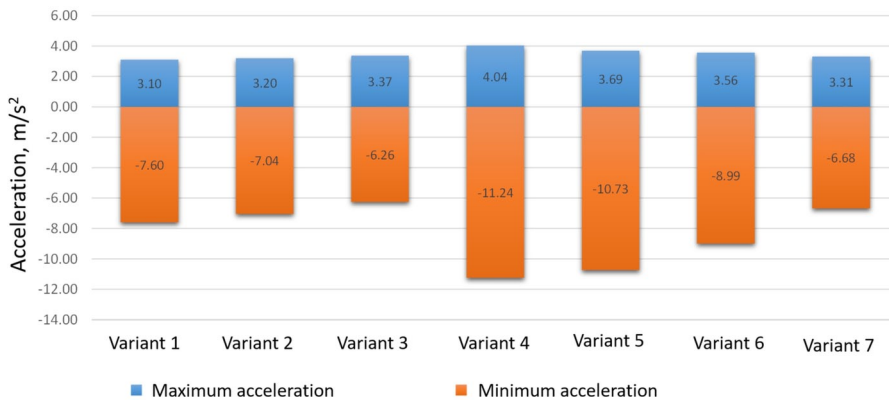


Fig. 9. Maximum and minimum accelerations for each of the analysed variants

The highest deceleration acting on the monorail operator during emergency braking was recorded for variants 4 and 5 – the variants in which 3 side lashings stabilizing the route were defined. On the other hand, the smallest decelerations affecting the operator during emergency braking were recorded for variants 2 and 3, in which one side lashing was used at the beginning of the route. Higher values were recorded in variant 7, in which all the slings were inclined, due to the fact that after braking the route moves further in the direction of monorail movement. The route was stabilized by 3 side lashings In variants 4 and 5. This eliminated the possibility of further movement of the route, what increased deceleration during emergency braking. It is worth mentioning that deceleration in these variants exceed the acceptable values regulated by law in the Polish mines. It is because the way of route installation is not proper on those route sections where a higher speed is allowed. On the other hand, in the case of variants 2 and 3, during emergency braking, the route was stabilized with one lashing, and in variant 7, with oblique slings. Such stabilization limited the effect of lifting the entire transport route upwards, as in variant 1.

As a result, the decelerations acting on the operator were minimized. Comparing variants 2 and 3, yielding of the side lashings resulted in a reduction of deceleration by approx. 11%. On the other hand, in the case of variants with three lashings (variants 4 and 5), the addition of side lashi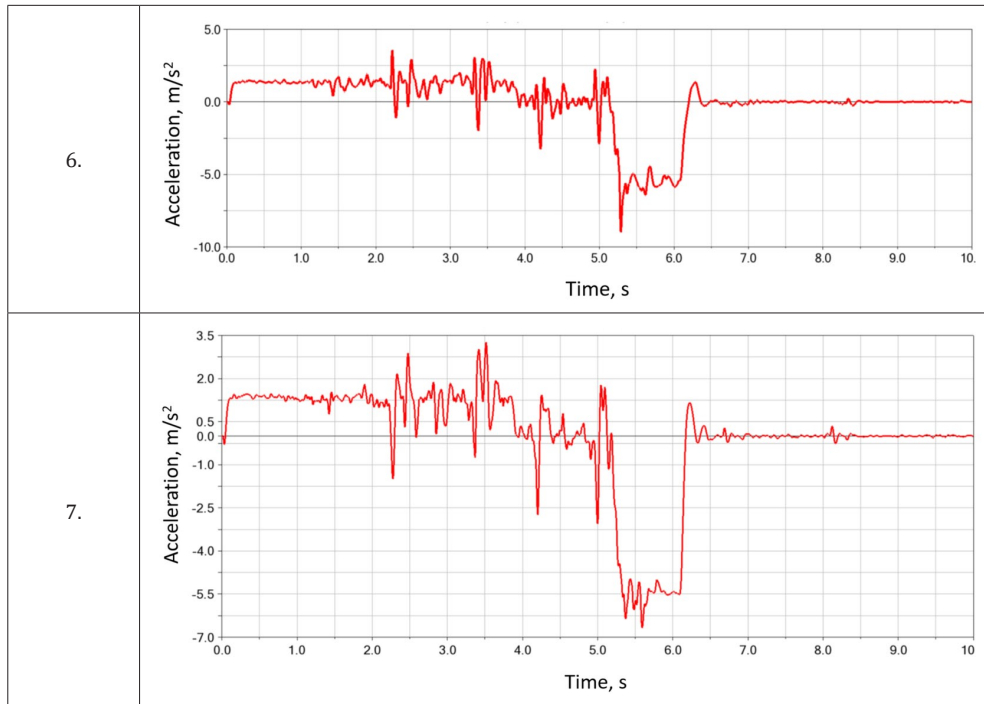ngs resulted in a reduction of the deceleration affecting the operator by about 4.5%. Comparing variant 3 (with one flexible side lashing) and variant 7 (all oblique slings), the deceleration affecting the operator is about 6% higher in variant 7. However, the advantage of variant 7 is the faster stabilization of the route oscillation after emergency braking. In the case of variant 3, after stopping the monorail, transport route forward and backward movement was observed, due to changes in acceleration from positive to negative. In the case of variant 7, oscillation of the route stops with the braking. To sum up, due to minimization of acceleration acting on the suspended monorail operator during emergency braking, the most successful route was the route with straight slings and one side lashing (variant 3) and the route where all slings were installed obliquely (variant 7).

Ad. 2) Forces in route slings

Forces in each sling were another parameter recorded during the simulation. Forces in relation to slings No. 11 – 22 are presented. The slings refer to location of transport set after it stopped (they are located above the transport set). Number of each sling, depending on variant, is shown in Fig. 10 - Fig. 12.

The maximum resultant forces acting on the slings are presented in Table 1. The areas marked in green mean the slings with maximum force at a low level, in most cases not exceeding 10 kN. Yellow and orange colours represent the average range of slings loads. The slings loaded with a force in the range of 10 kN – 30 kN are in this group. It is a load greater than in the group marked in green, however, it is accepted and does not cause any dangerous situations. The areas marked in red indicate slings loaded with a force of more than 30 kN. These are the most loaded slings and special attention should

be paid to them, because in these situations the accepted values may be exceeded.

Due to the stopping point of transport set, a lower load to slings numbers 19-22 was observed. The highest forces acting on the slings were recorded for pairs of slings marked as cz13 and cz14 as well as cz15 and cz16 in all variants. This results from the place, where the monorail stops during emergency braking. The heaviest compo-



*Fig. 10. Numbers of slings in variants 1 – 5*



*Fig. 11. Numbers of slings in variant 6*



*Fig. 12. Numbers of slings in variant 7*

nent of the set (the machinery part) is on the rail located between the slings cz15 and cz18. Direction of the monorail travel determines the direction and sense of inertia force, which loads mainly slings cz15 and cz16. The highest force acting on the sling was recorded for the variant 6 in the sling cz15. Such a large value results from fact that the load from the mass of the machine, i.e. inertia force of this monorail component, and the weight of the travel route, accumulated on this sling. This unfavorable phenomenon takes place when the transport route moves forward or backward at such arrangement of the slings, and this happens during emergency braking. In such a situation, one (front) sling is partially relieved, while the other „takes" part of the first load and the function of route stabilization. In variant 6, accumulation of these loads results in exceeding the maximum accepted forces loading each sling. A similar situation takes place in variant 7, in which 400 N is below the maximum accepted force. When analyzing the force recorded in the most loaded sling, it can be observed that the introduction of one side lashing (variant 2) decreased the force in the sling cz15 by approx. 5% compared to the variant with all straight slings (variant 1). On the other hand, the introduction of 3 side lashings (variant 4) resulted in a reduction of force in the most loaded sling by approx. 16.8%, in relation to the variant without lashings (variant 1). In turn, yielding the side lashings resulted in a reduction of forces in the most loaded sling by approx. 27.5%, comparing variants 1 and 3, and about 12.8% comparing variant 1 and 5. Analyzing the recorded forces in variant 6 and 7, a significant disproportions in force between adjacent slings forming the letter "V" inclined in opposite directions can be observed. This disproportion may reach even approx. 40% as in the variant 7 and slings cz15 and cz16.

In variants 2, 3, 4 and 5 there were side lashings stabilizing the monorail route. Fig. 13 - Fig. 15 show the maximum force vector components recorded in the stabilization lashings during the emergency braking.

The force components are in the following directions:
• OY - in line with the direction of the monorail movement,
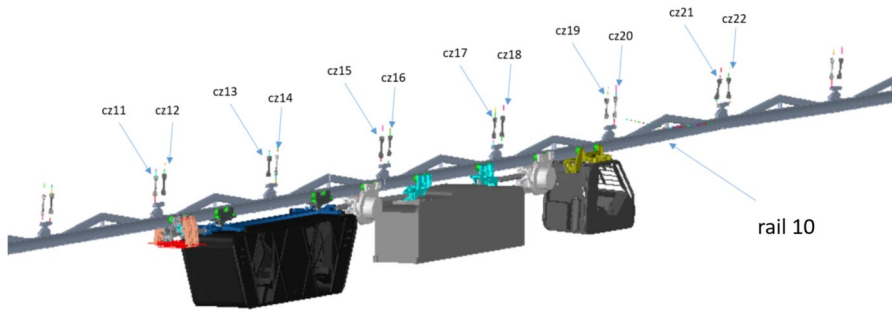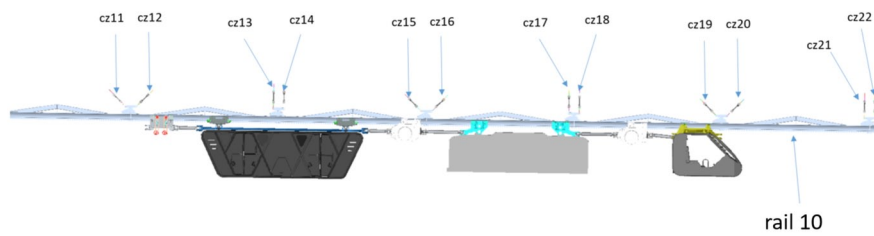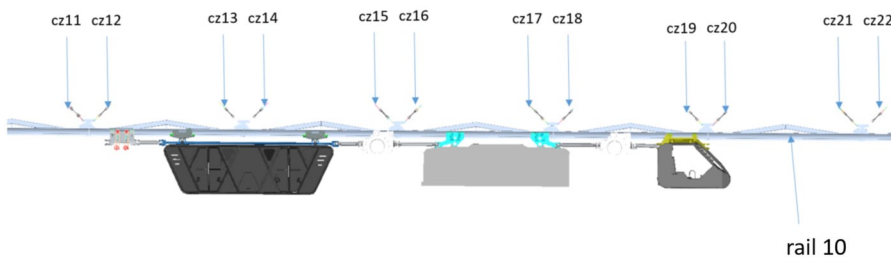
*Table. 1. Maximum resultant forces in slings 11-22*

| Variant number | Maximum resultant force | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | cz11, N | cz12, N | cz13, N | cz14, N | cz15, N | cz16, N | cz17, N | cz18, N | cz19, N | cz20, N | cz21, N | cz22, N |
| 1. | 22052 | 25741 | 25015 | 23675 | 31789 | 24582 | 21596 | 18553 | 7384 | 6079 | 1619 | 1069 |
| 2. | 22227 | 24496 | 25312 | 24801 | 30205 | 23031 | 17607 | 18622 | 7164 | 6816 | 1555 | 1377 |
| 3. | 24519 | 24156 | 23947 | 23188 | 23055 | 19035 | 16939 | 18035 | 6306 | 5980 | 1474 | 1050 |
| 4. | 19222 | 20777 | 19905 | 19830 | 26445 | 21541 | 16879 | 17728 | 6778 | 8171 | 5339 | 2648 |
| 5. | 21733 | 23721 | 25812 | 25048 | 27732 | 27118 | 16996 | 18059 | 5784 | 6122 | 2031 | 4346 |
| 6. | 23998 | 34444 | 23496 | 23471 | 47472 | 29085 | 17363 | 17316 | 11932 | 9786 | 1028 | 1014 |
| 7. | 28042 | 26823 | 29308 | 26103 | 39600 | 24043 | 27935 | 26671 | 9744 | 8837 | 1839 | 1471 |

• OZ - vertical, perpendicular to the axis of the monorail route,
• OX - horizontal, perpendicular to the axis of the monorail route.

The highest values occur in the components of the forces whose direction is consistent with the direction of movement of the transport
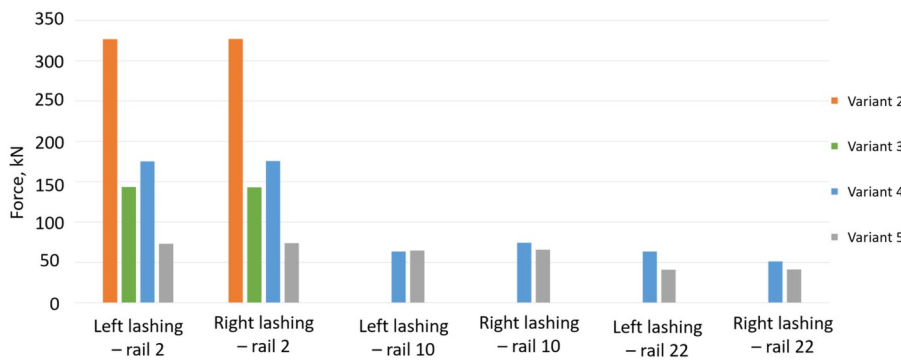


*Fig. 13. Component Y of the force in lashings stabilizing the route*
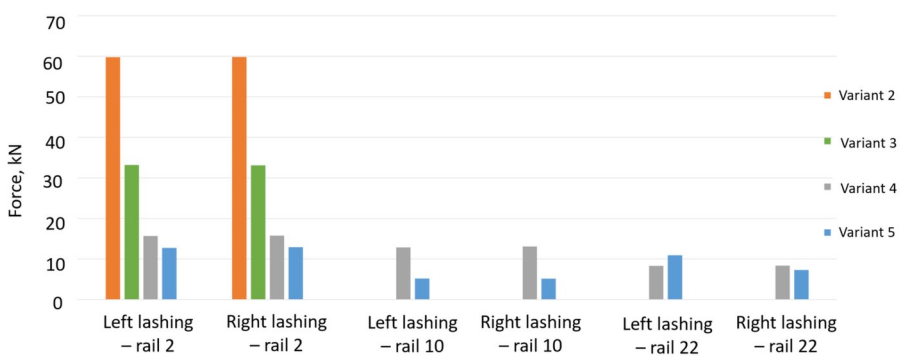


*Fig. 14. Component Z of the force in lashings stabilizing the route*
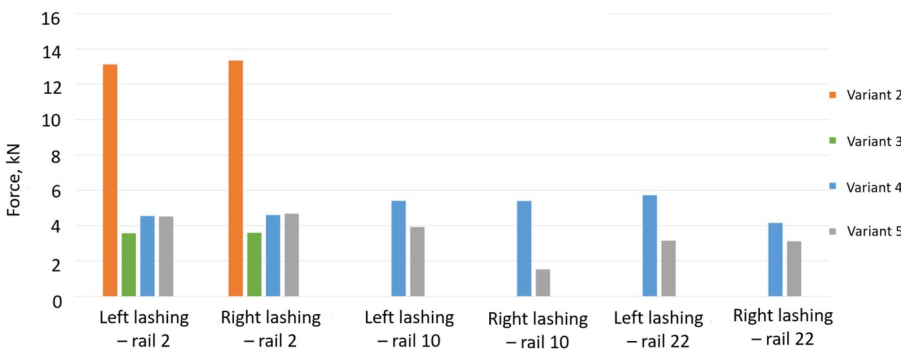


*Fig. 15. Component X of the force in lashings stabilizing the route*

unit. Increasing the number of lashings from 1 (variant 2) to 3 (variant 4) reduces the maximum force by approx. 46.3% in relation to the Y axis, by approx. 73.7% in the vertical axis (Z axis) and by approx. 3 % in the X axis. In addition, the introduction of yielding slings decreased the maximum forces in relation to the configuration with one sling by approx. 56% in the Y axis (in line with the direction of travel), by approx. 45% in the vertical axis (Z axis) and by approx. 73% in the X axis. In relation to the configuration with three side lashings, their yielding resulted in a reduction of the maximum forces in these lashings by approx. 46% in the Y axis, by approx. 20% in the vertical axis (Z axis) and by approx. 60% in the X axis. Therefore, it can be concluded that yielding the side lashings is an effective way to reduce the force in them. This conclusion justifies the purchase of more expensive components used in yielding lashings, especially in the sensitive places, where the roadway roof support is in a worse techni-

cal condition or in the strategic points along the suspended monorail route. The recorded results show that along with introduction of three side lashings, "tearing up" of the rail No. 10 was minimized. This is evidenced by a significant reduction in force in these lashings in the Z axis (vertical axis). The forces acting in the "X" direction have the smallest share in load to the stabilizing lashings, because the computational model covered only the straight section of the route where there were no external forces acting in this direction. The forces recorded in this axis resulted mainly from the deviation of the side lashings by an angle of 10 degrees from the horizon. Minimizing the force in the slings, which are transferred to the roadway roof support during emergency braking from speed of 5 m/s is most effective in variant 3 or variant 5. While analysing the forces in the side lashings, special attention should be paid to the large values of forces recorded in the lashings in axis of the railway movement (Y axis).

The presented forces suggest that during emergency braking from a speed of 5 m/s, the accepted loads to the roadway roof support may be exceeded. An improvement was observed in configurations with added side lashings and with the yielding of these lashings. A comprehensive solution to this problem may be designing the proper installation of the side lashings, the task of which will be to distribute dynamic loads to several adjacent roof support frames. In this way, the lashings will stabilize the monorail route and at the same time the accepted load to each roadway support arch is not exceeded.

Ad. 3) The rail further movement

In Table 2 further movement of the rail 1 and 10, in line with OY, OZ and OX directions is given.

The colours given in the table show that the greatest further movements were in Y direction, i.e. in line with the travel of the transport unit (red and orange).

Further movements in the Z direction, i.e. vertical (orange and green) were smaller and negligible ones were in the X direction (horizontal, perpendicular to the route axis). The largest further movements were in the route arranged according to variant 1 (without stabilization) and the maximum values mean a very large swing of the route, which in real conditions would not be acceptable (fields marked in red in Table 3). The introduction of one, yielding lashing reduces this value by approx. 50%, and the next two by approx. 60%. Further movements of approx. 180 mm during emergency braking do not pose any threats to the crew and passengers and at the same time they reduce the maximum force in the stabilizing lashings.

## 5. Analysis of the results

The cabin hitting a stationary obstacle was simulated as a part of the assessment of operator safetyin the case of an emergency braking when driving at a speed of 5 m/s. During the simulation,the Articulated Total Body (ATB) of Hybrid III dummy model was used, corresponding to a 50 percentile male.

Based on the numerical simulation, the maximum value of the Head Injury Criterion (HIC) during the collision of the cabin with

| Variant | Maximum further movement rail 1 | | | Maximum further movement rail 10 | | |
|---|---|---|---|---|---|---|
| | in Y axis, m | in Z axis, m | In X axis, m | In Y axis, m | In Z axis, m | In X axis, m |
| 1. | 0.365 | 0.132 | 0.005 | 0.365 | 0.132 | 0.001 |
| 2. | 0.133 | 0.084 | 0.001 | 0.129 | 0.017 | 0.001 |
| 3. | 0.188 | 0.078 | 0.001 | 0.184 | 0.03 | 0.001 |
| 4. | 0.111 | 0.079 | 0.001 | 0.104 | 0.107 | 0.0002 |
| 5. | 0.157 | 0.075 | 0.002 | 0.15 | 0.114 | 0.0003 |
| 6. | 0.009 | 0.002 | 0.004 | 0.009 | 0.002 | 0.001 |
| 7. | 0.005 | 0.001 | 0.002 | 0.005 | 0.001 | 0.002 |

an obstacle was determined. This is one of the injury criteria that has been established on the basis of biomechanical responses from experimental tests [7]. The HIC parameter is a function of the time and deceleration during head collision with an obstacle (5). The HIC is expressed by the following formula [11]:

$$HIC = \left[ \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} a(t)dt \right]^{2,5} (t_2 - t_1) \qquad (5)$$

where:
- a – liner acceleration (deceleration) of head centre of gravity, in g,
- $t_1$, $t_2$ – time of starting/ending the contact of head with an obstacle or time interval expressed in sec., at which HIC is maximal (dimensionless parameter).

Exceeding the value of 1000 of HIC means significant increase in the probability of a serious head injury [10, 15, 16]. In relation to the analysed case, two simulations were carried out - the first one, in which the operator sits freely in the operator cabin and has the ability to move around; in the second, the operator is additionally secured with four-point seat belts. Fig. 16 shows the operator initial position, which was the same in both simulations. Then the operator positions are shown after collision with an obstacle comparing the variant with fastened seat belts and without seat belts. The position of the operator is shown in 0.15 s and in 0.21 s of simulation.
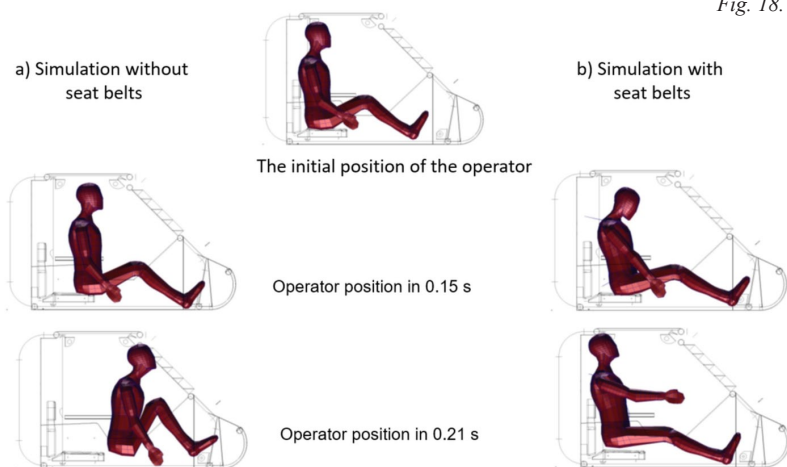


*Fig. 16. Position of Hybrid III dummy during collision with the obstacle at speed equal to 5 m/s: (a) simulation without safety belts, (b) simulation with the safety belts*

Fig. 17 shows the HIC parameter during the simulation of the vehicle hitting an obstacle when the operator does not have seat belts. The value reaches 1200, what means that there is a very high probability of serious or fatal head injury in the result of hitting the head on the cabin front.

Fig. 18 shows the HIC parameters in relation to simulation of the operator cabin collision in a situation with the seat belts fastened. In this case, the maximum HIC parameter is about 350, what means a low probability of suffering severe or fatal head injuries.

For interpretation of the above diagrams, the Abbreviated Injury Scale (AIS) for adults is used. In this way, it is possible to detail the HIC parameter impact on the damage level [16]. For example, when this parameter is equal to 1000,
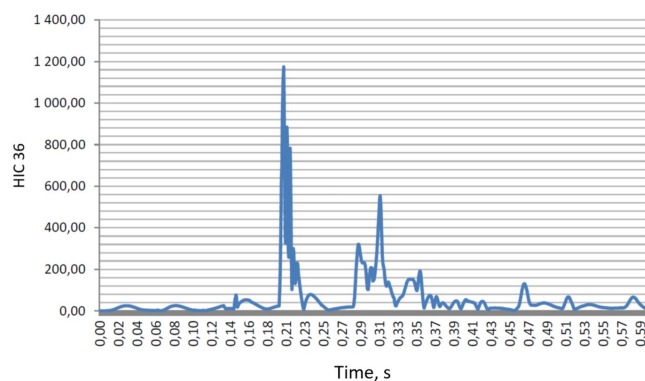


*Fig. 17. The HIC parameter when simulating an impact with an obstacle at a speed of 5 m/s while traveling without seat belts*
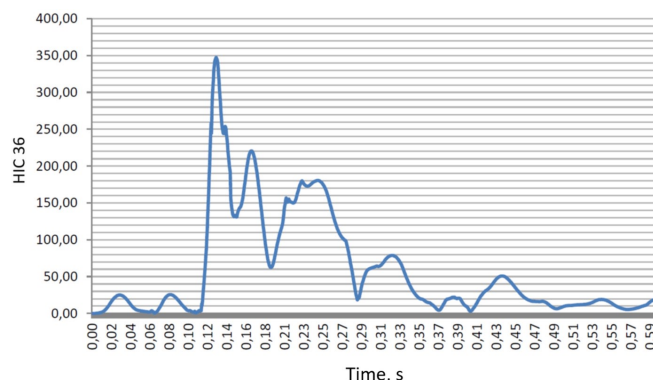


*Fig. 18. The HIC parameter when simulating an impact with an obstacle at a speed of 5 m/s while traveling with the seat belts*

it means the following:
- 18% probability of heavy head injury (AIS4),
- 55% probability of medium head injury (AIS3),
- 90% probability of light head injury (AIS2).

It is assumed that hitting the head on a non-deformable surface with a speed of at least 4 m / s may cause severe brain injuries TBI (Traumatic Brain Injury).

## 6. Conclusions

MBS simulation, along with the analysis of the results, enables the identification and selection of the method of suspending and stabilization of the monorail railway route. The simulations allow for the identification of adverse phenomena that may occur in emergency situations, such as emergency braking from higher speeds. The unfavourable

phenomena include excessive further displacement of the route, observed in variant 1. Excessive stiffening of the route (variants 6 and 7) is also not favourable, due to generation of high forces in the lashings of the route, which may result in exceeding the acceptable load to a single roadway roof support frame. Use of the yielding components in the lashings enables the minimization of the forces loading the roof support frame, which is especially important in the case of emergency. Regarding the suspension options analysed in the article, variants 3 and 7 are the most advantageous regarding minimizing the overloads acting on the operator, and variants 3 and 5 regarding minimizing the forces in the slings. In turn, the smallest forces in the side lashings were recorded in variants 4 and 5. As it results from the analyses presented in the article, there is no one universal and best configuration of the suspensions of the suspended monorail route, which in the situation of dynamic excitations would allow for the maximum minimization of all effects of this extortion. Therefore, numerical simulations should be a normal practice used by designers of transport routes in the production preparation departments of the mining plants, especially in the case of routes intended for high-speed suspension railway, used for the personnel movement. An additional advantage of using this type of simulation at the route designing stage is the possibility of checking and testing many variants of the configuration of slings and lashings proposed by the designers. Another conclusion from the analyses of forces in the side lashings is the need to use components that allow the distribution of forces resulting from dynamic forces into several adjacent roof support arches. This is a valuable information for designers of road transport at higher speed.

Simulations with the ATB, as well as the analyses of the results, indicate that the operator cabin should be equipped with additional passive safety elements, e.g. seat belts or headrests. These elements will protect the operator against severe and even fatal injuries in an emergency. The proposed additional equipment of the operator cabin significantly reduces the HIC coefficient, which should be interpreted as minimizing the likelihood of serious head injuries. Use of special numerical simulations enables both the quantitative and qualitative assessment of the impact of changed driving speed on the operator safety. Inability of this type of tests in real conditions, due to the existing regulations and ensuring the safety of the railway operator is an another argument.

Numerical analyses enable, in a safe and effective way, analysing the level of safety and comparing various scenarios of emergency situations that may happen during operation of the suspended monorail. The analyses of forces in the slings, the decelerations affecting people traveling by the monorail and the route displacements indicate that movement of passengers by the suspended monorail at a speed of 5 m/s, is possible without reducing the level of safety under condition of ensuring the proper design and stabilization of the suspended route as well as introduction of additional equipment (passive safety) in the transport unit. The designer should select the configuration of the slings, so that the route does not move excessively during emergencies. At the same time, care should be taken not to over stiffen the route, which will result in a significant increase in loads transferred to the frame of the roadway roof support.

# References

1. Becker F, Zell M. The state of the art in positively guided rail transport systems for underground mining. Mining Report 2014; 105(1/2): 34 – 46, https://doi.org/10.1002/mire.201400002.
2. Danelson K A, Golman A, Kemper A, Gayzik F. Finite element comparison of human and Hybrid III responses in a frontal impact. Accident Analysis and Prevention 2015, 85: 125-156, https://doi.org/10.1016/j.aap.2015.09.010.
3. Herbuś K, Szewerda K, Świder J. Virtual prototyping of the suspended monorail in the aspect of increasing the permissible travel speed in hard coal mines. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 4, 610-619, http://dx.doi.org/10.17531/ein.2020.4.4.
4. Horyl P, Šňupárek R, Maršálek P, Poruba Z, Pacześniowski K. Parametric Studies of Total Load-Bearing Capacity of Steel Arch Supports. Acta Montanistica Slovaca 2019; 24(3): 213-222.
5. INESI (RFCS) – Increase Efficiency and Safety Improvement in Underground Mining Transportation Routes (Zwiększenie efektywności i poprawa bezpieczeństwa w podziemnych, górniczych drogach transportowych). Contract No. 754169. Okres realizacji: 2017 – 2020.
6. Jagoda J, Hetmańczyk M, Stankiewicz K. Dispersed, self-organizing sensory networks supporting the technological processes. Mining Machines 2021, 2: 13-23, https://doi.org/10.32056/KOMAG2021.2.2.
7. Krzystała E, Kciuk S, Mężyk A. Identyfikacja zagrożeń załogi pojazdów specjalnych podczas wybuchu. Wydawnictwo Naukowe Instytutu Technologii Eksploatacji – Państwowy Instytut Badawczy. Radom 2012, ISBN 978-83-7789-161-2.
8. Lai X, Wang Y, Zhou Q, Lin Z, Culiere P. Development of a finite element pam-crash model of hybrid iii anthropomorphic test device with high fidelity. International Technical Conference on the Enhanced Safety of Vehicles (ESV). Washington:2011: 11-0031.
9. Lutyński A. KOMAG activities in the domestic and international research areas. Mining Machines 2021; 4: 47-60, https://doi.org/10.32056/KOMAG2021.4.5.
10. Mackay M. The increasing importance of the biomechanics of impact trauma. Sadhana 2007; 32(4): 397-408.
11. National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation (DOT). Occupant Crash Protection – Head Injury Criterion. S6.2 of MVSS 571.208, Docket 69-7, Notice 17. NHTSA, Washington, DC, 1972.
12. Pieczora E, Suffner H. Rozwój napędów dołowych kolejek podwieszonych. Maszyny Górnicze 2017; 3: 44-57.
13. Pieczora E, Tokarczyk J. Development of mine underground transportation with use of suspended monorails. Mining-informatics automation and electrical Engineering 2017; 4(532): 96-117, http://dx.doi.org/10.7494/miag.2017.4.532.96.
14. PN-EN ISO 13850:2016-03 Bezpieczeństwo maszyn – Funkcja zatrzymania awaryjnego – Zasady projektowania.
15. Post A, Hoshizaki T B, Glichrist M D, Brien S, Cusimano M, Marshall S. The dynamic response characteristics of traumatic brain injury. Accident Analysis and Prevention 2015; 79: 33-40, https://doi.org/10.1016/j.aap.2015.03.017.
16. Prasad P, Mertz H J. The Position of the United States Delegation to the ISO working group 6 on the use of HIC in the Automotive Environment. Conference: SAE Government Industry Meeting and Exposition SAE Government Industry Meeting and Exposition 1985; https://doi.org/10.4271/851246.
17. Prochowski L, Żuchowski A. Analysis of the influence of passenger position in a car on a risk of injuries during a car accident. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 360–366.
18. Pytkik A, Rotkegel M, Szot Ł. Badanie wpływu prędkości kolejek podwieszonych na siły w wybranych elementach trasy. Przegląd górniczy 2016; 11: 30-37.
19. Pytlik A. Tests of steel arch and rock bolt support resistance to static and dynamic loading induced by suspended monorail transportation.

Studia Geotechnica et Mechanica 2019; 41(2): 81–92, https://doi.org/10.2478/sgem-2019-0009.

20. Rozporządzenie Ministra Energii z dnia 23 listopada 2016 r. w sprawie szczegółowych wymagań dotyczących prowadzenia ruchu podziemnych zakładów górniczych (Dz.U. z 2017r. poz. 1118 z późn. zm).

21. Song ZA, Jiang F. Hydraulic system elaboration and simulation for single-drive light-load monorail locomotive in fully mechanized coal mining applications. IOP Conf. Series: Materials Science and Engineering 2019; 474, http://dx.doi.org/10.1088/1757-899X/474/1/012006.

22. Stankiewicz K. Mechatronic systems developed at the KOMAG. Mining Machines 2020(162); 2: 58 -68, https://doi.org/10.32056/KOMAG2020.2.6.

23. Świder J, Szewerda K, Herbuś K, Jura J. Testing the Impact of Braking Algorithm Parameters on Acceleration and Braking Distance for a Suspended Monorail with regard to Acceptable Travel Speed in Hard Coal Mines. Energies 2021; 14, 7275  https://doi.org/10.3390/en14217275

24. Szewerda K, Tokarczyk J, Bożek P, Michalak D, Drwięga A. Vibrations diagnostics and analysis in operator's and passenger cabins of a suspended monorail. Acta Montanistica Slovaca 2020; 2: 150-158, https://doi.org/10.46544/AMS.v25i2.2.

25. Szewerda K, Tokarczyk J, Wieczorek A. Impact of Increased Travel Speed of a Transportation Set on the Dynamic Parameters of a Mine Suspended Monorail. Energies 2021; 14(6), https://doi.org/10.3390/en14061528.

26. Teixeira R R, Moreira S R D S, Tavares S M O. Multibody dynamics simulation of an electric bus. Procedia Engineering 2015; 114: 470 – 477, https://doi.org/10.1016/j.proeng.2015.08.094.

27. Tokarczyk J. Method for identification of results of dynamic overloads in assessment of safety use of the mine auxiliary transportation system. Archives of Mining Sciences 2016; 61(4): 765-777.

28. Tokarczyk J. Method for virtual prototyping of cabins of mining machines operators. Archives of Mining Sciences 2015; 60 (1): 329-340.

29. Tokarczyk J. Metodyka identyfikacji wybranych zagrożeń mechanicznych w pomocniczym transporcie podziemnych zakładów górniczych. Prace Naukowe – Monografie KOMAG, Monografia nr 52 2017; Instytut Techniki Górniczej KOMAG Gliwice.

30. Wicher J, Więckowski D. Influence of vibrations of the child seat on the comfort of child's ride in a car. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2010; 4 (48): 102–110.

31. Wojtyra M, Frączek J. Metoda układów wieloczłonowych w dynamice mechanizmów. Oficyna Wydawnicza Politechniki Warszawskiej 2007; Warszawa.

32. Yýlmaz AI, Büyükyýldýz G, Ekici A, Çalýk M, Önder Ö, Aksoy CO. Staff transportation two way on the belt conveyor. Acta Montanistica Slovaca 2013; 18: 141-150.

33. Żuchowski A. Analysis of the influence of the impact speed on the risk of injury of the driver and front passenger of a passenger car. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (3): 436–444, http://dx.doi.org/10.17531/ein.2016.3.16.

# Influence of aqueous sodium chloride solutions on operational properties of epoxy coatings

Danuta Kotnarowska [a], Aleksandra Żabińska [a]

[a] University of Technology and Humanities in Radom, Faculty of Mechanical Engineering, ul. Malczewskiego 29, 26-600 Radom, Poland

## Highlights

- A permanent decrease in the operational properties of the brine aged coatings was shown.

- FTIR and DSC studies have documented the degradation of the brine aged epoxy material.

- A significant growth of destruction and roughness of the brine aged coatings was found.

- Silver crackings were revealed on the surface of the coatings aged for 1008 h - 1680 h.

- An original, universal method of multi-criteria studies of coating properties was used.

## Abstract

The main goal of the research presented in the article is to assess the impact of the destruction (degradation) of epoxy coatings, caused by the influence of a 25% aqueous solution of sodium chloride (brine), on the operational properties that determine the coatings reliability (durability). Samples of two-layer epoxy coatings (obtained from powder paints) were aged by immersing them in brine for a maximum period of 1680 h. The following criteria were used to assess the operational properties of the coatings: thickness, roughness and surface morphology, hardness, water absorption (mass increase), surface free energy. The FTIR and DSC tests performed showed a progressive development of the oxidation processes of the epoxy with the ageing period. The long-term impact of the brine resulted in the destruction of the coatings in the form of: craters, grooves, microcracks, including silver cracks. The contact of the coatings with brine also increased their: surface roughness (Ra, Rz, Rmax), thickness, mass, surface free energy, and decreased hardness and heat resistance.

## Keywords

epoxy coating, operational properties, ageing with brine, destruction.

## 1. Introduction

The polymer coatings of technical objects during their use in natural conditions are influenced by various types of harmful operational factors, both climatic and environmental. Among the climatic factors, the strongest destruction of polymer coatings is caused by solar UV radiation, which primarily contributes to the degradation of the coating material and to the initiation of surface cracking of silver coatings [27]. On the other hand, the dominant environmental factors include hard erosive particles (stones, gravel), causing: erosion, abrasion and scratching of the coatings [12], as well as aggressive media.

The aggressive media contributing to the strong destruction of polymer coatings include salt spray (occurring in the coastal area), as well as an aqueous solution of sodium chloride (the so-called brine), which is commonly used in winter to prevent the formation of a layer of ice on the road surface[40].

Operational factors synergistically influencing polymer coatings (climatic and environmental) cause their destruction, both chemical (degradation) and physical, manifested in the form of: craters, etchings, cracks, blisters, interlayer delamination and surface roughness increase [17]. Moreover, the impact of operational factors leads to the loss of durability of the adhesive bonds between the primer coat and the substrate [41], as a result of the development of undercoating corrosion [29].

Operational factors in the initial period of use of polymer coatings cause a progressive loss of their gloss and colour change, thus deteriorating the decorative properties of the coatings [30]. In subsequent stages, due to the increasing physical destruction of the coatings with the passage of their service life, as well as the degradation of the chemical structure of the polymer coating material, deterioration of the barrier properties of the coatings, determining their reliability, is observed [17].

The reliability of polymer coatings is primarily determined by their tightness, conditioned by the optimal proportion of fillers and pigments, limiting the porosity of the coating-forming material [11]. The reason for the formation of pores is the development of ageing processes in the structure of the coatings, mainly under the influence of climatic factors, and above all, solar UV radiation [13]. Pores can also be generated by the influence of environmental factors in the form of: mechanical loads [35], microorganisms [42] and aggressive media, including aqueous solutions of sulphuric acid [15] and sodium chloride [18].

The reliability of polymer coatings significantly depends on the type of coating resin used for their production, which was confirmed

(*) Corresponding author.
E-mail addresses: D. Kotnarowska (ORCID: 0000-0001-8790-3381): d.kotnarowska@uthrad.pl, A. Żabińska (ORCID: 0000-0003-2105-6065): ozabinska@gmail.com

by the test results presented in the paper [2]. The coatings are made of laminates produced on the basis of epoxy or polyester resin. The research proved that epoxy resin coatings showed greater resistance to abrasion and scratching [2].

The mechanical wear resistance of polymeric materials (protective coatings) is also determined by their hardness, the increase of which generally increases their resistance to abrasion and scratching. The authors of the work [30] investigated the impact of thermal fatigue and ageing on the hardness of polymer-ceramic composites, confirming the high hardness stability of new silorane composites, which translates into their high resistance to abrasive wear, compared to traditional composites.

Reliability tests of multi-layer polymer coatings protecting the floor of military cars against destruction caused by an explosion (e.g. grenade) are presented in the paper [36]. Effective protection against this type of rapidly increasing mechanical loads was provided by a coating consisting of three layers, the first of which was a glass composite (6 mm thick), the second was an aluminium sheet (12.5 mm thick), and the third was an aramid composite (2,5 mm thick).

Increasing the tightness of polymer coatings, which determines their reliability, is achieved by modifying the physical structure of the coating material by supplementing with various types of fillers, with optimal dimensions and mass fraction [9]. Microfillers ensuring good barrier properties of coatings include, among others: glass microspheres [12], silicon dioxide (silica) [23], titanium dioxide [38], zinc [24].

In this article, an epoxy primer coating modified with a micro-filler in the form of zinc dust and zinc monoxide is considered.

The reliability of polymer protective coatings is particularly effectively improved by addition of various types of nanofillers, among which the most commonly used is silicon dioxide (the so-called nanosilica) [7]. The composition of polymer coatings is also enriched with nanoparticles of: zinc [6], copper [14], titanium dioxide [31], aluminium trioxide [37], as well as silver nanoparticles, additionally acting as inhibitors of biological corrosion [28].

Titanium dioxide nanoparticles were used to modify the top layer of the epoxy coating system, the examination results of which are presented in this article.

The mass fraction of nanoparticles of fillers is generally ten times lower than in the case of using traditional microfillers. High tightness and good mechanical properties of polymer coatings (high resistance to: erosion, cracking, bending, impact, abrasion) are obtained when each of the filler grains is surrounded by the coating resin [11].

The introduction of nanofillers to the paints reduces the porosity of the coatings, thus improving their barrier properties, which translates into a reduction in the absorption of aggressive media. This was proved by the results of the research obtained in the work [10], showing a decrease in water absorption of epoxy coatings as a result of their modification with aluminium trioxide nanoparticles, with a simultaneous increase in their erosive resistance. It was also found that this type of modification of oxysilane coatings increased their resistance to the effects of salt spray [ 20].

The paper [37] showed that polymer coatings containing aluminium oxide nanoparticles, thanks to their high hardness and even distribution in the coating material, were characterized by increased resistance to abrasion and scratching.

Modification of epoxy coatings with copper nanoparticles [14], with a grain diameter below 66 nm (with a mass fraction of 3.5%), resulted in a reduction of over 30% of the pore volume determined with the use of mercury porosimetry. In addition, an increase in the erosion resistance of modified epoxy coatings was observed, which also showed high thermal stability and resistance to UV radiation with a wavelength below 350 nm.

Research has shown that modifying polymer coatings with small-sized nanoparticles does not reduce their transparency, which allows them to be used for varnish topcoats. For example, the introduction of zinc monoxide nanoparticles into the structure of methacrylate coat-ings (grain size 3÷5 nm and mass fraction 1÷15%) did not reduce their transparency [24].

Polymer coatings modified with nanofillers and nanopigments (called nanocoatings), apart from increasing the tightness [8] (due to the reduction of their porosity), are characterized by: increased resistance to UV radiation [38] and mechanical factors [25]. Moreover, nanocoatings (thanks to their high tightness) show a high durability of adhesive connections with the steel substrate, as a result of limiting the penetration of aggressive media into the surface of the protected element, which may contribute to the development of undercoating corrosion [26].

The content of nanofillers in the structure of polymer coatings also improves their thermal stability [22] and resistance to biological corrosion [42]. Moreover, the work [29] showed a higher resistance to chemical degradation due to UV ageing of polymer coatings containing UV absorbers in the form of nanoparticles of titanium dioxide and/or zinc monoxide, compared to coatings modified with organic absorbers.

In order to increase the service life of polyurethane-epoxy coatings, nanoparticles of aluminium trioxide (grain size d=20 nm) or silica (grain size d=16 nm), the mass fraction of which was 3.5%, were introduced into the polyurethane top layer. Among the coatings aged at the climatic station in the period of 3 years, the lowest chemical and physical destruction, as well as the highest hardness and erosive resistance were distinguished by coating systems with a polyurethane topcoat modified with aluminium trioxide [16].

The epoxy coatings assessed in this article have found a wide application in anti-corrosion protection of elements of technical facilities, mainly exposed to the influence of aggressive media [34]. The recipe of these coatings is constantly being improved, inter alia, thanks to modifications with the use of various types of nanofillers, primarily in order to improve their barrier properties, mechanical and thermal [35].

Multilayer epoxy coatings reinforced with carbon fibres have been used to increase the reliability of coating system protecting the inner surface of railway wagons against damages caused by mechanical and chemical factors, which was demonstrated in the work [4]. The tested coating systems consisted of five to ten layers, made of composite epoxy coatings reinforced with carbon fibres. The best mechanical properties were obtained for the coating consisting of seven layers.

Increased resistance to soaking by aggressive media of the three-layer epoxy coating was obtained as a result of improving its tightness, thanks to the introduction of nanofillers to the structure of each layer [36]. The base layer was modified with aluminium oxide nanoparticles, the interlayer with graphene nanoparticles, and the top layer with silicon dioxide nanoparticles [36]. An increase in the resistance of this coating to UV radiation was also found, as a result of its reflection by silicon dioxide nanoparticles contained in the topcoat.

A significant improvement in the anti-corrosive properties of epoxy coatings was achieved after the addition of 2% (mass fraction) of cerium and graphene monoxide nanoparticles to their composition. The nanofillers sealed their structure, thus limiting the migration of aggressive media to the coating substrate [5].

The effective improvement of the resistance of epoxy coatings (containing zinc microfiller) to the influence of a 3.5% aqueous solution of sodium chloride was achieved by introducing 0.3% (mass fraction) of graphene nanoparticles into their composition [1].

Nanofillers were also used in the production of self-healing epoxy coatings based on: bio-epoxy resin, halloysite nanotubes and silicon dioxide nanoparticles [33]. Halloysite nanotubes (HNTs) were filled with bio-epoxy resin (by vacuum infiltration), and silicon dioxide nanoparticles were mixed with a hardener activated by UV radiation. Then, the filled halloysite nanotubes were mixed with the bio-epoxy resin (together with the hardener). When attempting to scratch the coatings halloysite nanotubes break, and the bio-epoxy resin they contain fills the crackings and comes into contact with the UV activated hardener overlaying the silicon dioxide nanoparticles, initiating the

curing reaction. Complete self-healing of cracks under the influence of sunlight was observed when the mass fraction of halloysite nanotubes filled with bio-epoxy resin was 40%. Moreover, the composite showed increased shielding ability against UV radiation, as well as increased thermal stability and good transparency. Therefore, the developed bio-epoxy composite can constitute an effective, self-repairing protective coating of technical objects exposed to scratching and the influence of solar UV radiation.

The anticorrosive properties of epoxy coatings can also be improved by adding to their composition a filler in the form of flake basalt (chemically inactive), the surface modified with aniline trimer. The addition of 10% (by mass) of this filler to the composition of epoxy coatings increases their corrosion resistance more than twice [39].

Epoxy coatings also effectively protect concrete elements exposed to UV radiation and an aqueous solution of sodium chloride [32].

The resistance of polymer coatings, protecting technical objects against the negative effects of aggressive media, is largely determined by the hydrophilicity of their surfaces, conditioned by the value of the contact angle $\Theta_C$, characteristic for each type of coating, depending on the type of coating material, as well as the condition of the coating surface. On the surface of hydrophilic coatings for which the contact angle $\Theta_C=0°$, complete spreading of the water drops takes place. It should be noted that the most advantageous in terms of protective properties are coatings with low wettability (called hydrophobic coatings), characterized by the contact angle $\Theta_C=90°÷180°$. However, when the value of the contact angle $\Theta_C=180°$, the polymer coating is completely water repellent [21].

The research showed that the contact angle $\Theta_C$ decreases with the increasing degree of ageing of the coatings surface, especially their surface layers, and the lower its value, the larger the area of the surface wetted with drops of aggressive media, which results in increased absorbability of aged coatings. On the other hand, the aggressive media absorbed into the coatings contribute to their swelling and the degradation of the chemical structure of coating material [19].

If the contact angle characteristic for a given coating is $\Theta_C=45°÷90°$, the penetration of aggressive media (water) into its surface layers to a large extent depends on the chemical structure of the coating polymer and the geometry of the wetted surface roughness profile. The development of destruction processes in polymer coatings under the influence of climatic factors (mainly as a result of the influence of solar UV radiation) contributes to the reduction of the contact angle $\Theta_C$ that characterizes them, thus increasing the absorption of the coatings with aggressive media characteristic for the operational environment (e.g. acid rain, fog salt, brine) [19].

It should be emphasized that the reliability of polymer coatings is determined primarily by the durability of adhesive bonds between the coating and the substrate. It is reduced as a result of the development of undercoating corrosion as a result of the penetration of aggressive media into the usually steel substrate [26, 41].

Epoxy coatings considered in this paper, due to their high adhesion to the steel substrate, have found wide application as primer coatings [35].

The ageing of epoxy coatings (obtained from solvent-based paints) with an aqueous solution of sodium chloride results in the degradation of the coating material and the destruction of the coatings in the form of: crackings, etchings, pores and craters (reaching the substrate). Moreover, a significant increase in the surface roughness of the coatings was observed, assessed using the Ra and Rz parameters. The destruction of the coatings progressively increased with the increase of the concentration of the aggressive medium, which, in researches presented in the work [18], was 3%, 10%, 20 %.

Due to the insufficient collection of information contained in the world literature on the development of the processes of destruction of polymer coatings as a result of the influence of climatic factors and/or environmental factors, no effective method of forecasting their reliability (service life) has been developed so far. This is due to the fact that the elements of the "Operational Environment - Coating - Coating Substrate" system are subject to constant changes, inter alia as a result of the continuous development of technology (nanotechnology) in the field of coating materials (polymer materials, fillers and pigments) and their application techniques, and also as a result of progressively increasing environmental pollution and increasing intensity of solar ultraviolet (UV) and thermal (IR) radiation [32].

Due to the development of technology in the area of paints and varnishes, pro-ecological powder paints are produced more and more frequently [35], and environmentally friendly curing of coatings with UV radiation is also used [3,].

This article presents the results of tests of epoxy coating systems made of innovative powder paints.

Due to the fact that the paint formula changes on average every five years, conducting long-term operational tests in order to determine the reliability (durability) of polymer coatings is substantially unjustified. Hence the necessity to carry out accelerated tests [29] of the operational properties of coatings, in conditions modelling the dominant climatic and/or environmental factors.

This article presents the results of research on the physicochemical properties of epoxy coating systems (used in the renovation painting of car chassis elements), which were subjected to accelerated ageing with a 25% sodium chloride aqueous solution, modelling brine used in winter to prevent road slippery. Brine is one of the dominant factors that destroy the polymer protective coatings of the car undercarriage components during its winter operation. Acceleration of ageing was obtained thanks to complete immersion of coating samples in this medium.

## 2. Research methodology

### 2.1. Materials and preparation of samples for testing

Epoxy coating systems, which are used to protect the components of the car chassis, were subjected to investigation.

Samples of epoxy coatings electrostatically deposited on the surface of steel plates with dimensions (160 x 80 x 2) mm were examined. The plates were cut out of sheets, made of calmed structural steel, S 235 JRG 2 (according to PN-EN 10025-1: 2007). The surface of the coatings (before applying the primer coat) was cleaned with an angle grinder (DWE4157), using sandpaper of P120 gradation. Prior to applying the primer, the surface of the steel plates was subjected to a degreasing process with an extraction solvent.

The samples of the coating systems were obtained by applying powder paint to the surface of the steel plates using the electrostatic spray method. Two layers of epoxy coatings were produced successively: a primer epoxy coating (grey) and an epoxy topcoat (blue).

The base coats were obtained from epoxy powder paint, modified with microfiller in the form of stabilized zinc dust (grain size 3÷4 μm and mass fraction 71%) and zinc monoxide (the so-called zinc white, with a mass fraction of 5%).

The topcoats were also obtained from epoxy powder paint. However, for its modification, nanoparticles of titanium dioxide were used, the mass share of which was 2%. The blue colour of the topcoats was obtained by using a mineral pigment with a grain size below 20 μm, consisting of sodium aluminosilicates combined with sulphur ($Na_6Al_6Si_6O_{24}S_4$).

### 2.2. Methodology of accelerated ageing of epoxy coatings with brine

The effects of ageing of epoxy coatings as a result of the action of a 25% aqueous solution of sodium chloride (called brine) were investigated. In order to accelerate the examination, the coatings were fully immersed in a container filled with brine. The surface condition of the coatings was assessed after the appropriate test period, which was: 672 h (28 days), 1008 h (42 days), 1344 h (56 days) and 1680 h (70 days).

Table 1. Standard test methods for the physicochemical properties of epoxy coatings and the test equipment used

| Type of examination method | Apparatus |
| --- | --- |
| Standard number | |
| Determination of the coating thickness<br>PN-EN ISO 2808:2020-01 | Elcometer 456C |
| Method of geometric structure examination of coatings' surface<br>PN-M-04251:1987;  PN-EN ISO 8501-1:2008 | Hommel Tester T500<br>(by Hommelwerke) |
| Buchholz hardness test method<br>PN-EN ISO 2815:2004 | Erichsen hardness tester |

Table 2. Non-conventional examination methods of epoxy coatings and test equipment

| Type of examination method | Apparatus |
| --- | --- |
| Study of changes in the chemical structure of an epoxy coating material | Spectrophotometer FT/IR-6200 (by Jasco) |
| Examination of the coatings thermal resistance | Thermal analyser LABSystem TG /DSC (by SETARAM) |
| Examination of the geometrical structure of the surface | Interferometric microscope Talysurf CCI (by Taylor Hobson) |
| Examination of the surface morphology of epoxy coatings | Scanning electron microscope with field emission SU-70 (by Hitachi) |
| Determination of the elemental composition of the surface layer of epoxy coatings | X-ray microanalyser with EDS energy dispersion by Thermo Scientific (Hitachi SU-70 microscope) |
| Examination of the epoxy coatings surface wettability | Goniometer G10 (by Krüss) |

## 2.3. Methodology of epoxy coatings testing

The physicochemical properties of unaged and aged with brine epoxy coatings were assessed on the basis of results of performed standard tests of: thickness, surface roughness parameters (Ra, Rz, Rmax), and hardness. The applied standard test methods and the test equipment assigned to them, are described in Table 1.

Non-conventional examination methods presented in Table 2 were used to assess the causes and effects of destruction (degradation of the coating material) of epoxy coatings aged with brine.

## 3. Results and their discussion

### 3.1. Influence of ageing with brine on the degradation of the chemical structure of epoxy coating material

Ageing of epoxy coatings with brine increased the intensity of characteristic bands (appearing in the FTIR spectrum) of epoxy material contained in the surface layer of epoxy coatings (Fig. 1). This testifies changes in the chemical structure of the aged epoxy material [13].
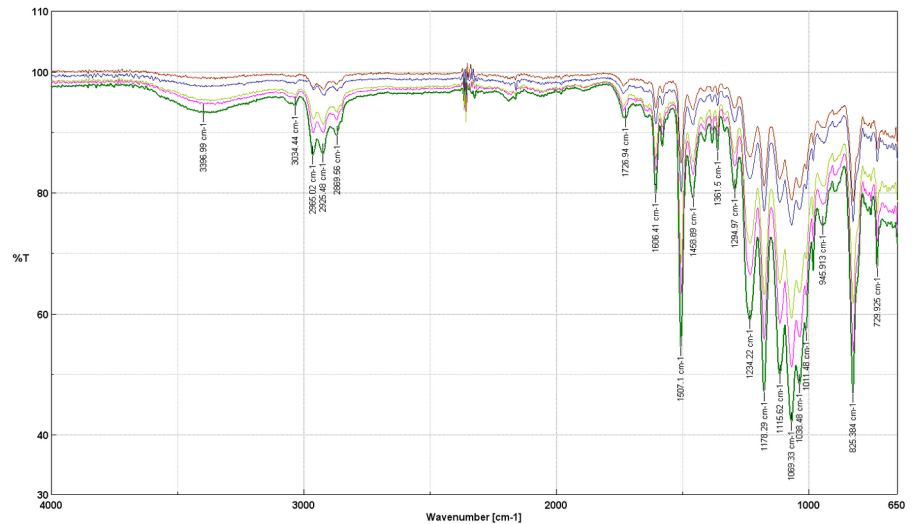


Fig. 1. Characteristics of the FTIR spectrum of an unaged epoxy coating (violet curve) and aged during: 672 h (blue curve), 1008 h (light green curve), 1344 h (pink curve) and 1680 h (green curve)

In the epoxy material, the intensity of the carbonyl group band (C=O), with the peak corresponding to the wave number 1726.94 cm$^{-1}$, increased by more than 3 times in the period of 1680 h. This proves a growth of the oxidation degree of the epoxy material.

Table 3. Temperature values of the energy effects in the epoxy topcoat determined on the basis of the DSC curves course

| Ageing period t [h] | Temperature of energy effects [ºC] | | |
| --- | --- | --- | --- |
| Ageing period [h] | T$_p$<br>temperature of the beginning of the oxidation process | T$_1$<br>temperature of the exothermic peak | T$_k$<br>temperature of the end of the oxidation process |
| 0 | 258 | 320 | 396 |
| 672 | 256 | 320 | 396 |
| 1008 | 255 | 325 | 396 |
| 1344 | 253 | 325 | 396 |
| 1680 | 243 | 325 | 396 |
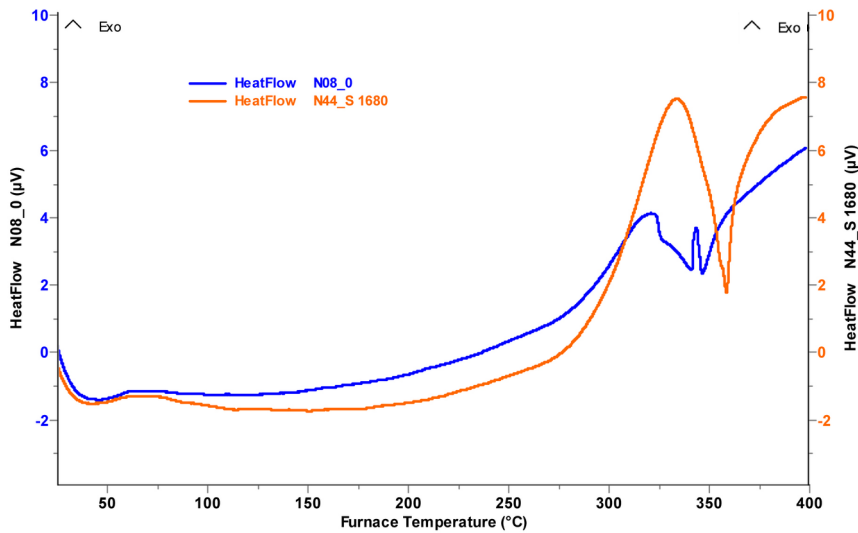
*Fig. 2. Evolution of the DSC curve of an epoxy coating aged with brine in the period of 1680 h (orange curve) compared to the DSC curve of the unaged coating (blue curve)*

The increase (almost 3 times) of the intensity of hydroxyl (OH) groups band (with a peak of $3396.99$ cm$^{-1}$) of also documents the development of epoxy oxidation processes. On the other hand, the increase in the content of CH$_3$ groups (in the band with the peak corresponding to the wave number of $2925.48$ cm$^{-1}$) proves the breaking of chemical bonds in the epoxy material [13].

## 3.2. Influence of ageing with brine on the thermal resistance of coatings

Differential scanning calorimetry (DSC) was used to assess the heat resistance of epoxy coatings. In the examination of the DSC characteristics, samples of epoxy coatings in the form of a powder, obtained as a result of abrasion of the top layer of the epoxy coating, were used. The average mass of the samples used in the examination was $(3.8\pm0.1)$ mg.

The course of the DSC characteristics (Fig. 2) shows that ageing with brine caused a progressive °decrease of the oxidation process onset temper-

*Table 4. Surface topography of epoxy coatings*
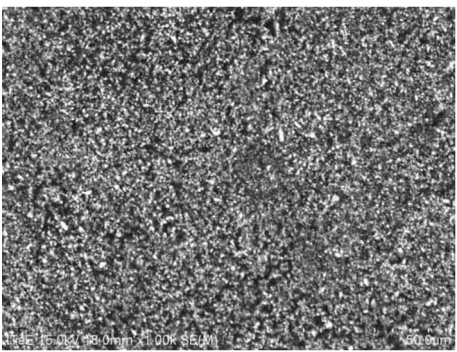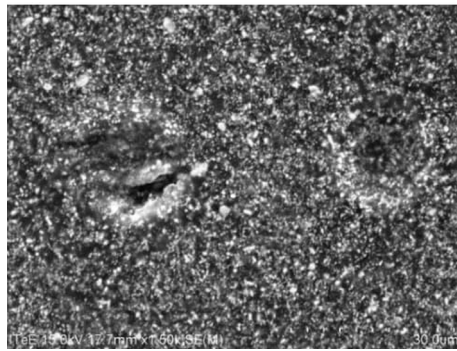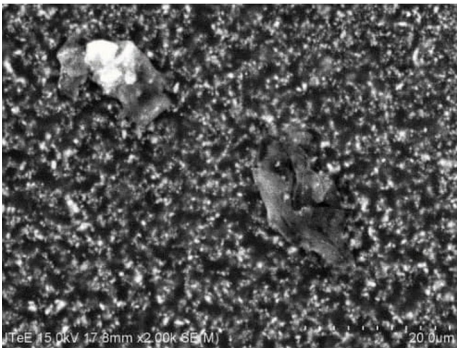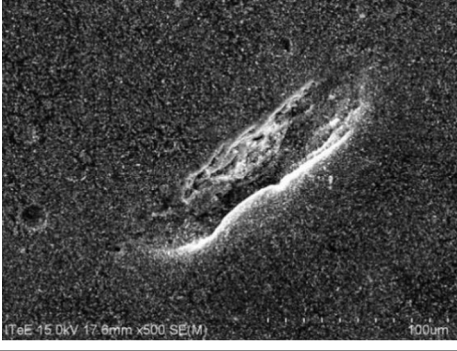
| Type of coating | Epoxy coating surface topography |
|---|---|
| Unaged epoxy coating |  |
| Epoxy coating aged with brine during the 1344 h |  |
| Epoxy coating aged with brine during the 1680 h |  |

ature. For example, after the ageing period of 672 h, the oxidation onset temperature decreased by 2°C, and after ageing for 1680 h, it decreased by 15°C (Tab. 3). Increase of the exothermic peak on the DSC curves also proves the development of oxidation processes in the epoxy material aged with brine (Fig. 2). Thus, the thermal resistance of epoxy coatings decreased.

### 3.3. Influence of ageing with brine on the surface topography of epoxy coatings

Ageing of epoxy coatings with brine caused significant changes in their surface topography (Table 4).

The presented images of the surface topography of epoxy coatings aged with brine show that the surface destruction of the coatings was nonhomogeneous (Table 4).

### 3.4. Influence of ageing with brine on the surface morphology of epoxy coatings

Ageing of epoxy coatings with brine resulted in the destruction of their surface layers (Table 5) in the form of: craters, chippings, rarefactions, crackings. This was due to changes in the chemical structure of aged coatings. Table 5 shows the results of the coating morphology investigation obtained with the use of a scanning electron microscope (SEM).

*Table 5. Morphology of epoxy coatings aged with brine*

| Ageing period t [h] | Epoxy coating images obtained using SEM | Description of the destruction of the epoxy coating aged with brine |
|---|---|---|
| 0 |  | |
| 672 |  | • formation of craters in the topcoat |
| 672 |  | • release of topcoat fragments, • thinning of the top coat |
| 1008 |  | • cracking, • craters |

| | | |
|---|---|---|
| 1008 |  | silver cracking<br>the release of fragments of the topcoat (chipping), rarefactions in the top coat |
| 1344 |  | cracking of the top coat, |
| 1680 |  | cracking of the top coat |

As a result of coatings ageing with brine, development of craters, chipping of coating fragments and rarefactions formation was observed in top coats after the period of 672 h. After ageing in the period of 1008 h new kinds of top coats destruction appeared in the form of crackings, including also silver crackings. Continuing of ageing (for 1344h and 1680 h) contributed to the further development of crackings in top coats.

The chipping of coating components and the formation of craters in coating material was primarily influenced by the increase in the degree of oxidation of epoxy coatings aged with brine. On the other hand, the cause of cracking and the formation of rarefactions in the structure of the coatings was the breaking of chemical bonds (Fig. 1).

### 3.5. Influence of ageing with brine on the elemental composition of epoxy topcoats

X-ray studies (similarly to FTIR and DSC studies) have documented an increase in the degree of oxidation of epoxy coatings aged with brine. A clear increase in the mass fraction of oxygen was found after

*Table 6. X-ray characteristics of epoxy coatings*

| Ageing period t [h] | X-ray characteristics |
|---|---|
| 0 |  |

Full scale counts: 300      N27-S-1008h-an1
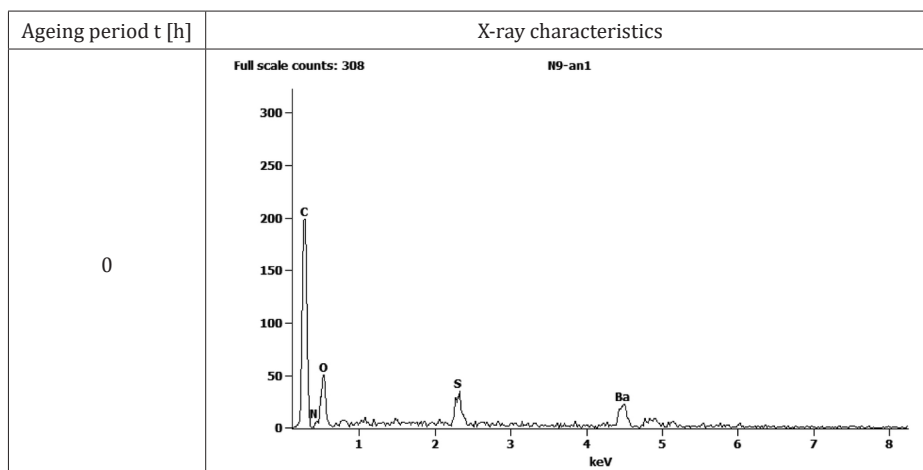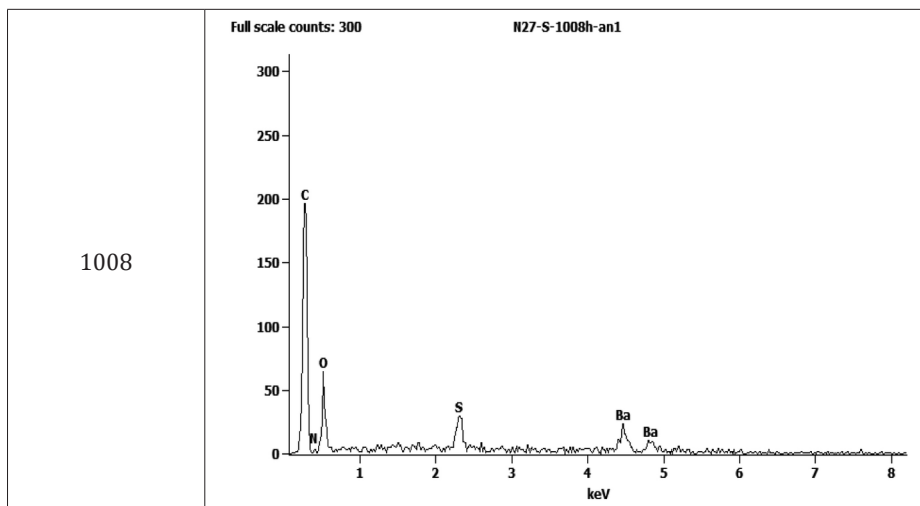
1008

the coatings ageing for 1008 h (Table 6). Due to the dominance of the process of oxidized coating layers chipping, the share of oxygen in the tested coatings, assessed on the basis of X-ray examination, remained unchanged after a longer period of ageing of the coating samples with brine.

### 3.6. Influence of ageing with brine on the wettability of epoxy coatings

The length of the ageing period had a significant effect on the value of the contact angle of the coating while wetting with the measuring liquid (water or diiodomethane). Ageing of epoxy coatings with brine caused (along with the ageing period flow) a decrease in the value of the contact angle $\Theta_C$ (Table 7) with distilled water. After ageing for 1680 h, the value of the contact angle $\Theta_C$ decreased by 10%, thus increasing the contact area of water with the epoxy topcoat.

The obtained results indicate the hydrophilicity growth of aged epoxy coatings, which increases their susceptibility to soaking with aggressive media (water). It should be emphasized that long-term penetration of the coating with aggressive media may lead to the development of undercoating corrosion.

*Table 7. Contact angle $\Theta_C$ values of epoxy coatings surfaces aged with brine*

| Ageing period t [h] | Contact angle $\Theta_C$ [°] | |
|---|---|---|
| | water | diiodomethane |
| 0 | 88.18 | 51.69 |
| 672 | 86.22 | 59.42 |
| 1008 | 80.99 | 54.83 |
| 1344 | 80.10 | 58.62 |
| 1680 | 78.98 | 62.67 |

The contact angle $\Theta_C$ was used to determine the Surface Free Energy (SFE) of epoxy coatings. The Owens-Wendt method, which is commonly used to determine the surface free energy of polymer composite materials, including epoxy coatings, was used to calculate the SFE. In the research carried out with this method, distilled water (polar liquid), characterized by a high value of the polar component, and diiodomethane (non-polar liquid) as a dispersive liquid were used as measuring liquids.

The obtained results of research on the surface free energy of epoxy coatings aged with brine (Tab. 8) showed a significant effect of their ageing period on the value of the SFE components, both polar and dispersive. The free surface energy decreased by 4% for epoxy coatings aged with brine in the period of 1680 h.

On the other hand, ageing with brine contributed to a significant increase in the value of the polar component of the surface free energy of epoxy coatings. After ageing in the period of 1680 h, its over three-fold increase was observed. It should be said with high probability that it was influenced by changes in the chemical structure of the epoxy material contained in the surface layer of epoxy coatings (Fig. 1).

### 3.7. Influence of ageing with brine on the thickness of epoxy coatings

Characteristic of the thickness increase of epoxy coatings aged with brine during the period of 1680 h is presented in Figure 3, while Table 9 shows the trend line equation and the value of the coefficient of determination $R^2$ of this characteristic.



*Fig. 3. Characteristic of the thickness increase of epoxy coatings aged with brine*

*Table 8. Effect of ageing of epoxy coatings with brine on the value of surface free energy and its components (dispersive and polar)*

| Ageing period t [h] | Surface free energy (SFE) $\gamma_s$ [mJ/m$^2$] | Dispersive component of SFE $\gamma_s^d$ [mJ/m$^2$] | Polar component of SFE $\gamma_s^p$ [mJ/m$^2$] |
|---|---|---|---|
| 0 | 35.53 | 33.33 | 2.20 |
| 672 | 32.59 | 28.91 | 3.68 |
| 1008 | 36.49 | 31.54 | 4.94 |
| 1344 | 35.28 | 29.37 | 5.91 |
| 1680 | 34.18 | 27.04 | 7.14 |

The test results proved a progressive increase in the thickness of aged coatings, which after ageing in the period of 1680 h accounted for 5%, which was caused by soaking of the coatings with brine 6.

### 3.8. Influence of ageing with brine on the mass of epoxy coatings

The characteristic of the mass increase of epoxy coatings aged with brine during the period of 1680 h is presented in Figure 4, while Table 10 shows the trend line equation and the value of the coefficient of determination $R^2$ of this characteristic.
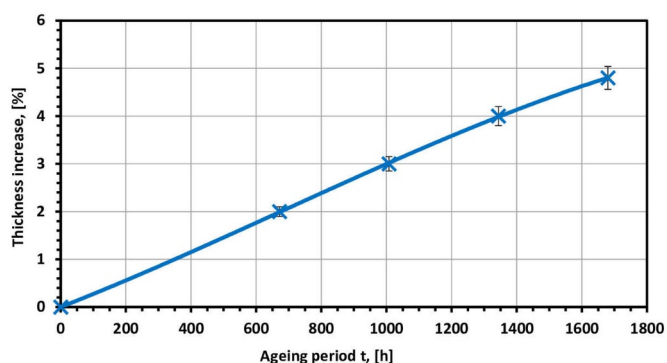
The conducted research shows that ageing with brine caused a progressive mass increase of epoxy coatings. After ageing of the coatings in the period of 1680 h, due to the penetration of brine into its interior, an increase in the mass of the coatings by 8% was observed.

was caused by an increase in the brittleness of the coatings as a result of an increase in their oxidation degree (Fig. 1, Table 3).

### 3.10. Influence of ageing with brine on the surface roughness of epoxy coatings

The impact of brine ageing on the surface roughness of epoxy coatings was assessed on the basis of examination of their parameters (Ra, Rz and Rmax) of surface roughness. Ageing with brine in the period of 1680 h resulted in a significant increase in their value, which was caused by the chipping of components from the top layer of the coatings.

The characteristics of the Ra, Rz and Rmax parameters of the surface roughness of epoxy coatings aged with brine in the period of 1680 h are presented in the figures 6 ÷ 8, while the equations of the trend line and the values of the coefficient of determination $R^2$ of these



*Fig. 4. Characteristic of mass increase of epoxy coatings aged with brine*
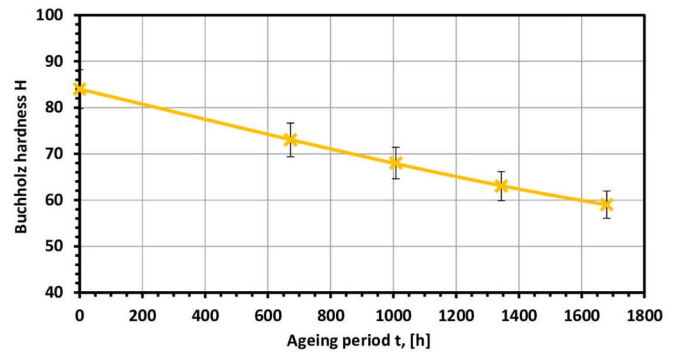


*Fig. 5. Characteristic of Buchholz hardness of epoxy coatings aged with brine*

*Table 9. Equation of the trend line and the value of the coefficient of determination $R^2$ for the thickness increase characteristic of epoxy coatings aged with brine*

| Characteristic name | Mathematical model | Coefficient of determination $R^2$ |
|---|---|---|
| Thickness increase | $y = -3E\text{-}10x^3 + 6E\text{-}07x^2 + 0.0027x + 0.0011$ | 0.9999 |

*Table 10. Equation of the trend line and the value of the coefficient of determination R2 for the mass increase characteristic of epoxy coatings aged with brine*

| Characteristic name | Mathematical model | Coefficient of determination $R^2$ |
|---|---|---|
| Mass increase | $y = -1E\text{-}09x^3 + 5E\text{-}06x^2 + 0.0003x - 0.0027$ | 0.9999 |

*Table 11. Equation of the trend line and the value of the coefficient of determination R2 for the hardness (according to Buchholz) characteristic of epoxy coatings aged with brine*

| Characteristic name | Mathematical model | Coefficient of determination $R^2$ |
|---|---|---|
| Buchholz hardness | $y = 9E\text{-}10x^3 - 7E\text{-}07x^2 - 0.0162x + 83.993$ | 0.9999 |

The reason for this was (among other things) an increase in the water absorption of the coatings, resulting from the reduction of the contact angle of the coating surface with water (Table 8).

### 3.9. Influence of ageing with brine on the hardness of epoxy coatings

The characteristic of Buchholz hardness measurements of epoxy coatings aged with brine in the period of 1680 h is presented in Figure 5, and in Table 11, the equation of the trend line and the value of the coefficient of determination $R^2$ of this characteristic are presented.

Ageing with brine resulted in a progressive reduction of hardness (according to Buchholz) of epoxy coatings. After ageing of the coatings in the period of 1680 h, the hardness decreased by 30%, which



*Fig. 8. Characteristic of the Rmax parameter of the surface roughness of epoxy coatings aged with brine*

Fig. 6. *Characteristic of the Ra parameter of the surface roughness of epoxy coatings aged with brine*
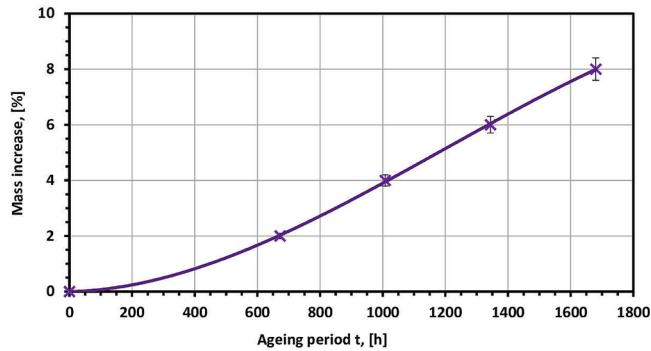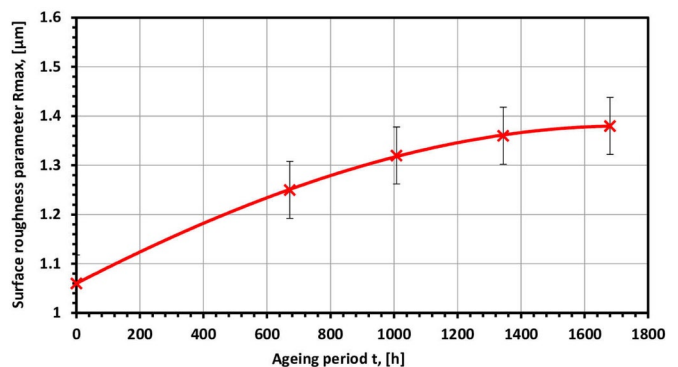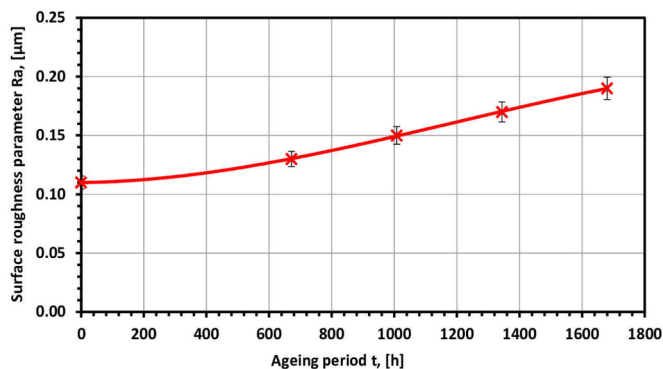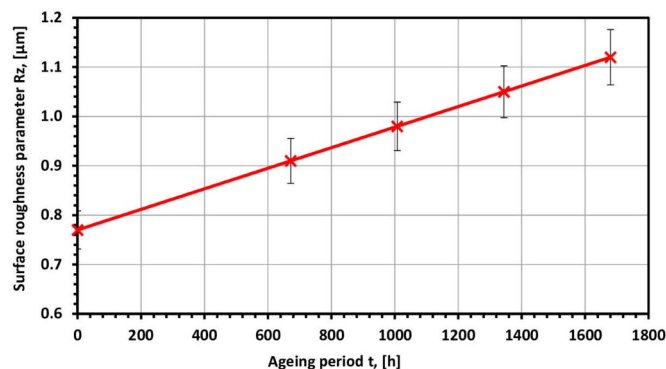


Fig. 7. *Characteristic of the Rz parameter of the surface roughness of epoxy coatings aged with brine*

Table 12. *Equation of the trend line and the value of the coefficient of determination $R^2$ of the Ra parameter of the surface roughness characteristic of epoxy coatings aged with brine*

| Characteristic name | Mathematical model | Coefficient of determination $R^2$ |
|---|---|---|
| Ra parameter | $y = -1E-11x^3 + 5E-08x^2 + 3E-06x + 0.11$ | 0.9999 |

Table 13. *Equation of the trend line and the value of the coefficient of determination $R^2$ for the parameter Rz of the surface roughness characteristic of epoxy coatings aged with brine*

| Characteristic name | Mathematical model | Coefficient of determination $R^2$ |
|---|---|---|
| Rz parameter | $y = 0.0002x + 0.77$ | 1 |

Table 14. *Equation of the trend line and the value of the coefficient of determination $R^2$ for the Rmax parameter of the surface roughness characteristic of epoxy coatings aged with brine*

| Characteristic name | Mathematical model | Coefficient of determination $R^2$ |
|---|---|---|
| Rmax parameter | $y = -1E-11x^3 - 6E-08x^2 + 0.0003x + 1.0599$ | 0.9999 |

characteristics are presented in Tables 12÷14. Ageing with brine resulted in a progressive increase in the Ra Rz and Rmax parameters of epoxy coatings, which after ageing in the period of 1680 h increased by 73%, 45% and 30%, respectively.

## 4. Summary

The conducted literature study shows that the current range of knowledge concerning the operational properties of polymer coatings exposed to brine during use, prevents the effective design of reliable protective and decorative coatings of technical devices.

The main reasons for this are the constant changes in the composition of coatings, caused by the development of nanotechnology in the field of nanofillers and nanopigments used for their modification, as well as the growing pro-ecological requirements in the field of paint and varnish production.

The speed of these changes is in contradiction with the traditional exploitation tests of the protective effectiveness of coatings, the duration of which in the case of innovative coatings can be even several years. The solution to this problem are accelerated tests used in conditions of intensified operational exposures caused by climatic and environmental factors. The degree of intensification of these exposures should be selected in such a way as to be adequate to the exposures affecting the actual period of several years of operation of the device protected with a given protective and decorative coating.

This article presents the results of accelerated (multi-criteria) tests of the variability of the physicochemical properties of epoxy coatings,

such as their thickness, mass (water absorption), hardness, surface roughness, heat resistance, water wettability.

Epoxy coating systems, consisting of a base layer modified with zinc microfiller and a top layer modified with nanoparticles of titanium dioxide, were subjected to multi-criteria tests of coating properties, determining their operational characteristics.

The epoxy coatings were aged by immersing them in a 25% aqueous sodium chloride solution (called brine). The maximum period of ageing with this medium was 1680 h.

Unconventional tests were used to determine the conditions for the destruction of brine-aged epoxy coatings, which include: FTIR infrared spectroscopic tests, DSC differential scanning calorimetry tests, tests of the surface morphology of the coatings performed with the use of scanning electron microscopy, tests of the geometric structure of the surface of the coatings with the use of an interferometric microscope, testing the angle $\Theta_C$ of wetting the surface of the coatings with a model liquid (water or diiodomethane).

The observed increase in brittleness of the top layer of brine-aged epoxy coatings, which progressively increased with the ageing time, was caused by the development of oxidation processes of the epoxy material. This is evidenced by an over 3-fold increase in the intensity of the C=O carbonyl group band (with a peak corresponding to the wave number 1726.94 cm⁻¹), disclosed in FTIR studies. Moreover, a more than 2-fold increase in the intensity of the OH hydroxyl group band (with a peak of 3396.99 cm⁻¹), also documenting the development of oxidation processes of epoxy topcoats. On the other hand, increasing the intensity of the CH₃ group band (with a peak corre-

sponding to the wavenumber of 2925.48cm$^{-1}$) indicates the breaking of chemical bonds in the epoxy material.

The chipping of the components of the top layers of the coatings, as a result of the increase in the degree of their oxidation, resulted in a significant change in the surface roughness profile.

For example, after 1680 h of ageing, the Ra parameter of surface roughness increased by 73%, the Rz parameter by 45%, and the Rmax parameter by 30%.

The effect of brine contributed to a reduction in the heat resistance of aged coatings, as it was found that the temperature of the beginning of the oxidation process (the beginning of thermal decomposition) decreased with the ageing time, as demonstrated in DSC tests. It was found that after 1680 h ageing, the temperature of the beginning of the oxidation process decreased by 15$^{\circ}$C. Moreover, an increase in the intensity of the exothermic peak on the DSC curve was observed. This proves the development of epoxy oxidation processes with the passage of their ageing with brine. X-ray studies have also documented an increase in the susceptibility of epoxy coatings to oxidation due to ageing with brine. They showed a clear increase in the mass fraction of oxygen after ageing of the coatings over the period of 1008 h, determined on the basis of X-ray examinations.

After ageing with brine for 1680 h, a decrease in the hardness (according to Buchholz) of the coatings by 30% was also observed, which contributes to a decrease in their mechanical properties [25].

Ageing of epoxy coatings with brine caused a progressive (with ageing time flow) decrease in the value of the water contact angle $\Theta_C$ on their surface, which after ageing for 1680 h decreased by over 10%. The obtained results indicate an increase in the hydrophilicity of epoxy coatings. Long-term penetration of the coatings with brine and its migration to the steel substrate may lead to the development of undercoating corrosion, as a result of which the coating will lose its adhesion to the steel substrate. Then it ceases to fulfil its protective (barrier) function and needs renovation.

The conducted research also revealed an over 3-fold increase in the value of the polar component of surface free energy in the case of epoxy coating ageing in the period of 1680 h. This was probably caused by significant ageing changes in the form of degradation of the chemical structure of the epoxy topcoat coating material.
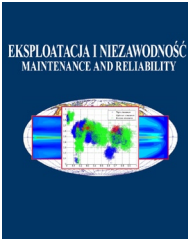
It was also found that after ageing of the coatings in the period of 1680 h, their thickness increased by almost 5%, and their mass by 8%. The likely reason for this was the increase in wettability and porosity of the coatings (due to cracking of chemical bonds).

To sum up, the effect of brine on epoxy coatings resulted in deterioration of their operational properties, determining the coatings reliability, documented by: an increase in thickness and mass (water absorption), a decrease in hardness, a significant increase in the values of parameters (Ra, Rz, Rmax) of surface roughness, a decrease in the angle $\Theta_C$ of surface wetting with water, a decrease in the thermal decomposition onset temperature, as well as increase in the surface free energy and the degree of destruction (in the form of: crackings, silver crackings, rarefactions, craters).

## References

1. Bai W. Ma Y., Meng M., Ying Li Y. The influence of graphene on the cathodic protection performance of zinc-rich epoxy coatings, Progress in Organic Coatings 2021; 161: Article 106456, https://doi.org/10.1016/j.porgcoat.2021.106456.
2. Baier A, Zolkiewski S. Initial research of epoxy and polyester warp laminates testing on abrasive wear used in car sheathing. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15 (1): 37–43 (no paper link available).
3. Bauer F., Decker U., Naumov S., Riedel C. Photoinitiator-free UV curing and matting of acrylatebased nanocomposite coatings: Part 3, Progress in Organic Coatings 2014; 77 (6): 1085-1094, https://doi.org/10.1016/j.porgcoat.2014.03.013.
4. Buchacz A, Baier A, Herbus K, Ociepka P, Grabowski L, Sobek M. Compression studies of multi-layered composite materials for the purpose of verifying composite panels model used in the renovation process of the freight wagon's hull. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (1): 137–146, http://dx.doi.org/10.17531/ein.2018.1.18.
5. Chen L., Yu Z., Yin D., Cao K., Xie C., Zhu L., Jiang Y. Preparation and anticorrosion properties of GO-Ce-MOF nanocomposite coatings, Applied Polymer 2022; 139 (5): Article 51571, DOI: 10.1002/app.51571.
6. Daneshifar M. H., Sajjadi S.A., Zebarjad S., Mohammadtaher M., Abbasi M., Mossaddegh K. The effects of fillers on properties of automotive nanocomposite clear coats: Type, content and surface functionalization, Progress in Organic Coatings 2019; 134: 33–39, https://doi.org/10.1016/j.porgcoat.2019.05.001.
7. Das S., Pandey P., Mohanty S., Nayak S. Effect of nanosilica on the physicochemical, morphological and curing characteristics of transesterified castor oil based polyurethane coatings, Progress in Organic Coatings 2016; 97: 233-243, https://doi.org/10.1016/j.porgcoat. 2016.04.012.
8. Deflorian, F., Fedel, M., Dirè, S., Tagliazucca, V., Bongiovanni, R., Vescovo, L., Minelli, M., & De Angelis, M. G. Study of the effect of organically functionalized silica nanoparticles on the properties of UV curable acrylic coatings. Progress in Organic Coatings 2011, 72(1-2), 44-51, https://doi.org/10.1016/j.porgcoat.2011.01.002
9. Fu S.-Y., Feng X.-Q., Lauke B., Mai Y.-W. Effects of particle size, particle matrix interface adhesion and particle loading on mechanical properties, Composites 2008; Part B, 39: 933-961, https://doi.org/10.1016/j.compositesb.2008.01.002.
10. Golru S. S., et al. Studying the influence of nano-Al$_2$O$_3$ particles on the corrosion performance and hydrolytic degradation resistance of an epoxy/polyamide coating on AA-1050, Progress in Organic Coatings 2014; 77: 1391-1399, https://doi.org/10.1016/j.porgcoat.2014.04.017.
11. Knowles T. The new toolbox. Nanotechnology in paints and coatings, European Coatings Journal 2006; 3: 16-18 (no paper link available).
12. Kotnarowska D. Kinetics of wear of epoxide coating modified with glass microspheres and exposed to the impact of alundum particles, Progress in Organic Coatings 1997; 31: 325-330, https://doi.org/10.1016/S0300-9440(97)00090-8.
13. Kotnarowska D. Influence of ultraviolet radiation and aggressive media on epoxy coating degradation, Progress in Organic Coatings 1999; 37: 149-159, https://doi.org/10.1016/S0300-9440(99)00070-3.
14. Kotnarowska D. Effect of nanofillers on wear resistance of polymer coatings, Solid State Phenomena 2009; 144: 285-290. (Pt. B of Diffusion and Defect Data - Solid State Data), https://doi.org/10.4028/www.scientific.net/SSP.144.285.
15. Kotnarowska D. Epoxy coating destruction as a result of sulphuric acid aqueous solution action, Progress in Organic Coatings 201; 67: 324-328, https://doi.org/10.1016/j.porgcoat. 2009.10.026.
16. Kotnarowska D. Analysis of polyurethane top-coat destruction influence on erosion kinetics of polyurethane-epoxy coating system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (1): 103–114, http://dx.doi.org/10.17531/ein.2019.1.12.
17. Kotnarowska D. Impact of operational environment on the protective and decorative properties of epoxy coatings, Ochrona przed Korozją (Corrosion Protection) 2013; 9: 372-383 (no paper link available).
18. Kotnarowska D. Destruction of Epoxy Coatings under the Influence of Sodium Chloride Water Solutions, Solid State Phenomena 2015; 220: 609-614, https://doi.org/10.4028/ www.scientific.net/SSP.220-221.609.

19. Kotnarowska D. The influence of battery acid on the destruction of acrylic coatings of car bodies, Coatings 2021; 11 (8): 967-987, https://doi.org/10.3390/coatings11080967.

20. Kozhukharov S., Kozhukharov V., Wittmar M., Schem M., Aslan M., Caparrotti H., Veith M. Protective abilities of nanocomposite coatings containing $Al_2O_3$ nano-particles loaded by $CeCl_3$, Progress in Organic Coatings 2011; 71: 198-205, https://doi.org/10.1016/j.porgcoat.2011.02.013.

21. Latthe S. S., at al. Self – cleaning superhydrophobic coatings: Potential industrial applications, Progress in Organic Coatings 2019; 128: 52-58, https://doi.org/10.1016/j.porgcoat.2018.12.008.

22. Le T. T., Nguyen T. V., Nguyen T. A., Thanh T., Nguye H. Thermal, mechanical and antibacterial properties of water-based acrylic Polymer/$SiO_2$–Ag nanocomposite coating, Materials Chemistry and Physics 2019; 232,: 362-366, https://doi.org/10.1016/j.matchemphys.2019.05.001.

23. Leder G., Ladwig T., Valter V., Frahn S., Meyer J. New effects of fumed silica in modern coatings, Progress in Organic Coatings 2002; 45: 139-144, https://doi.org/10.1016/S0300-9440(02)00049-8.

24. Lü N., Lü X., Jin X., Lü C. Preparation and characterization of UV-curable ZnO/polymer nanocomposite films, Polymer International 2006; 56: 138-143, DOI: 10.1002/pi.2126.

25. Maganty S., et al. Enhanced mechanical properties of polyurethane composite coatings through nanosilica addition, Progress in Organic Coatings 2016; 90: 243-251, https://doi.org/ 10.1016/j.porgcoat.2015.10.016.

26. Matin E., Attar M. M., Ramezanzadeh B. Investigation of corrosion protection properties of an epoxy nanocomposite loaded with polysiloxane surface modified nanosilica particles on the steel substrate, Progress in Organic Coatings 2015; 78: 395-403, https://doi.org/10.1016/j.porgcoat.2014.07.004.

27. Narisawa I. The strength of polymeric materials, OHMSHA Publisher, Tokyo, 1982, 397 p. (no paper link available).

28. Nguyen T.N. L., et al. Antimicrobial activity of acrylic polyurethane/$Fe_3O_4$-Ag nanocomposite coating, Progress in Organic Coatings 2019; 132:15-20, https://doi.org/10.1016/j.porgcoat. 2019.02.023.

29. Nguyen T.V. Stability of acrylic polyurethane coatings under accelerated aging tests and natural outdoor exposure: The critical role of the used photo-stabilizers, Progress in Organic Coatings 2018; 124: 137-146, https://doi.org/10.1016/j.porgcoat.2018.08.013.

30. Pieniak D., Niewczas A., Kordos P. Influence of thermal fatigue and ageing on the microhardness of polymer-ceramic composites for biomedical applications, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2012; 14 (2): 181-184 (no paper link available).

31. Pilotek S., Tabellion F. Nanoparticles in coatings. Tailoring properties to application, European Coatings Journal 2005; 4: 170-177 (no paper link available).

32. Prak L., Sumranwanich T., Tangtermsirikul S. Experimental investigation on the degradation of coating on concrete surfaces exposed to accelerated and natural UV in chloride environment, Journal of Adhesion Science 2022; Published online: 22 Jan 2022, DOI:10.1080/01694243.2022.2026707.

33. Pulikkalparambil H., Parameswaranpillai J., Siengchin S., Pionteck J. UV light triggered self-healing of green epoxy coatings, Construction and Building Materials 2021; 305: Article 124725, https://doi.org/10.1016/j.conbuildmat.2021.124725.

34. Report: Epoxy Surface Coatings Market Size, Industry Analysis Report, Regional Outlook, Application Development, Price Trend, Competitive Market Share & Forecast. Global Market Insights 2022 – 2028 (no paper link available).

35. Report: Global Powder Coatings Market Outlook Report 2021-2026: Focus on Thermoset Resin, Epoxy, Epoxy Polyester, Polyester, Polyurethane, Thermoplastic Resin. Research And Markets 2021, https://www.researchandmarkets.com/r/ux993b.

36. Szudrowicz M. Layered composite increasing the resistance of patrol and intervention vehicles to the impact of improvised explosive devices (IED) from below. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (1): 9–15, http://dx.doi.org/10.17531/ein.2018.1.2.

37. Wang Y., Lim S., Luo J.L., Xu Z.H. Tribological and corrosion behaviors of $Al_2O_3$/polymer nanocomposite Coatings Wear 2006; 260: 976-983, https://doi.org/10.1016/j.wear. 2005.06. 013.

38. Wojciechowski K., Zukowska G., Z, Korczagin I, Malanowski P. Effect of $TiO_2$ on UV stability of polymeric binder films used in waterborne facade paints, Progress in Organic Coatings 2015; 85: 123-130, https://doi.org/10.1016/j.porgcoat.2015.04.002.

39. Yue Q., Wu L., Lv J., Wang A., Ding R., Wang Y. Study on anti-corrosion performance and mechanism of epoxy coatings based on basalt flake loaded aniline trimer, Colloid and Interface Science Communications 2021; 45: Article 100505, DOI 10.1016/j.colcom.2021.100505.

40. Zhang W., Li L., Yao S., Zheng G. Corrosion protection properties of lacquer coatings on steel modified by carbon black nanoparticles in NaCl solution, Corrosion Science 2007; 49: 654-661, https://doi.org/10.1016/j.corsci.2006.06.017.

41. Zubielewicz M., Królikowska A. The influence of ageing of epoxy coatings on adhesion of polyurethane topcoats and protective properties of coating systems, Progress in Organic Coatings 2009; 66: 129-136, https://doi.org/10.1016/j.porgcoat.2009.06.014.

42. Zyska B. (red.), Żakowska Z. (red.), Mikrobiologia Materiałów., Wydawnictwo Politechniki Łódzkiej, Łódź 2005, ISBN 83-7283-150-5, http://hdl.handle.net/11652/1807.

# Double Layered Priority based Gray Wolf Algorithm (PrGWO-SK) for safety management in IoT network through anomaly detection

**Indexed by:**
Web of Science Group

Akhileshwar Prasad Agrawal [a,*], Nanhay Singh [a]

[a] Guru Gobind Singh Indraprastha University, Dept of Computer Science and Engg., Ambedkar Institute of Advanced Communication Technologies and Research (now NSUT-E), Geeta Colony, Delhi-110031, India

| Highlights | Abstract |
|---|---|
| • Proposed swarm intelligent based PrGWO resulting in optimal feature set.<br><br>• Contributed by improving classification of individual classes present in datasets.<br><br>• Overall the accuracy improved for two of the datasets, with third very close to best.<br><br>• Proposed new fitness function resulting in inclusive performance measurement. | For mitigating and managing risk failures due to Internet of Things (IoT) attacks, many Machine Learning (ML) and Deep Learning (DL) solutions have been used to detect attacks but mostly suffer from the problem of high dimensionality. The problem is even more acute for resource starved IoT nodes to work with high dimension data. Motivated by this problem, in the present work a priority based Gray Wolf Optimizer is proposed for effectively reducing the input feature vector of the dataset. At each iteration all the wolves leverage the relative importance of their leader wolves' position vector for updating their own positions. Also, a new inclusive fitness function is hereby proposed which incorporates all the important quality metrics along with the accuracy measure. In a first, SVM is used to initialize the proposed PrGWO population and kNN is used as the fitness wrapper technique. The proposed approach is tested on NSL-KDD, DS2OS and BoTIoT datasets and the best accuracies are found to be 99.60%, 99.71% and 99.97% with number of features as 12,6 and 9 respectively which are better than most of the existing algorithms. |

**Keywords**

Gray Wolf Optimizer, anomaly detection, feature selection, predictive maintenance.

## 1. Introduction

Internet of things (IoT) application is ever increasing since its inception due to its widespread use in areas of smart buildings, smart vehicles, smart highways and wireless sensor network which will grow exponentially in coming years [27]. The network layer is particularly more vulnerable to attacks due to attackers' ability to launch it at several locations. Various famous attacks aimed at the network layer are categorized into Denial of service (DoS) attacks, User to Root (U2R) attacks etc [8]. Apart from the network layer, attacks may be directed at the application layer or other layers, too. Over the years research efforts have been aimed at improving the algorithms for classification and thereby improving Intrusion detectin system (IDS), however still there are lot of problems left to be adequately addressed in order to make the system robust.

*Problems and motivation:*

1. Developing a feature selection technique to find the most optimal reduced feature vector. The requirement for optimized feature vector is necessary to a) reduce the training time of the model and b) test the incoming traffic in real time.

2. Developing an IDS that can improve classification accuracy, detection rate, false positive rate and other important performance metrics using this optimal reduce feature vector.

*Objective:*

Motivated by the problems listed above, the authors intend to devise an IDS technique/methodology **to reduce and find optimal feature vector giving maximum fitness in terms of accuracy, Detection rate (DR) and False positive rate (FPR)** for a given traffic/dataset. Thus the present problem can be conceived as a **multiobjective optimization** problem. This optimal feature vector can then be used for classification on the IoT nodes. The IoT nodes based on the trained model can then test for incoming traffic using this optimal reduced feature set (instead of considering the entire feature set)in real time.

*Contribution:*

1. A priority based Gray wolf Optimizer with SVM and kNN (PrGWO-SK) is proposed, in which the support vector Machine (SVM) alongwith first layer fitness function is used to find the reduced feature vector which then acts as feeder vector for the initialization of Priority based Gray wolf optimizer (PrGWO) wolves/particles.

(*) Corresponding author.
E-mail addresses: A.P. Agrawal (ORCID: 0000-0001-6366-0311): kpw.ce08@gmail.com, N. Singh (ORCID: 0000-0002-3303-1386): nanhay.singh@nsut.ac.in

2. PrGWO is proposed wherein the intra-group priority of different leader hunting wolves is used by other wolves to update their own positions during each iteration. The leader wolves uses the same concept in a modified form to update their own positions.

3. In the present work, two fitness functions are proposed to be used in the two different layers. The first layer fitness function focuses solely on the accuracy while the second layer fitness function balances not only the requirement of accuracy but considers other metrics like detection rate, false positive rate and length of feature vector.

4. PrGWO-SK was tested on three datasets-legacy NSL-KDD dataset and newly captured DS2OS, BoTIoT datasets using various metrics- Accuracy,FPR, Recall, Precision and F1-score and the results were found to be encouraging.

## 2. Related work

The IDS system working on resource staved IoT nodes has limited capacity. Antunes Rodrigues et al. [3] emphasized on preventive maintenance of sensor nodes in industrial application using Ishikawa diagram and FMECA. M. Almiania et al. [2] used the recurrent neural network concept to classify the data but it suffered from delays and was not amenable to real time applications. Tian et al. [39] used the deep belief network alongwith non-Mean gaussian distribution to classify the network packets. However, feature selection was not considered here. Baranowski [4] used the Bayesian workflow to predict attacks and failures in IoT particularly considering device variance. Modi et al. [22] developed a framework which used Snort and combined different classifiers, viz Decision tree, Bayesian & Associative. However, this was done for cloud network which is not resource starved as IoT nodes. Sivapalan et al. [32] suggested using lightweight Neural network to detect attacks in the wearable IoT(ECG). Gao et al. [6] in their paper proposed to use memory augmented autoencoder for detecting attacks in time series sensor data. A very important supervised technique is SVM which is used to draw hyper-plane between different classes. In one of the papers, the authors analysed threats using artificial neural networks [12].Vijayanand et al. [40] used the technique of SVM in the field of mesh networks security and found it to be efficient in distinguishing the attack from normal types. Similarly, E.Shams et al. [30] in their paper used the SVM technique to

implement security mechanism in vehicular adhoc network. To reduce the training time, subsampling technique was used to filter out the less useful data thus validating the utility of SVM in different applications. In the literature, extensive research has been done in the field of metaheuristic algorithms for traversing the search space and converging to a solution. Tama et al. [35] in their paper have used ensemble technique to classify the data. In this paper feature selection was accomplished through Ant colony optimization and PSO. However the work suffers from methodological complexity in using a large number of algorithms. Kunhare et al. [19] used the random forest algorithm for feature selection coupled with the PSO algorithm. The accuracy and number of features were optimally generated, however the comparison with other work was not extensive. Wei et al. [41] in their paper used the jaccard fitness function for evaluating the optimality of the feature set. The accuracy was measured to be very good but the number of features increased. J.Gu et al. [9] used the concept of Naive Bayes to enhance the differences in the feature values to enable SVM to clearly distinguish between normal and attack types. Though the accuracy improved, the number of features was neglected. T.Wisanwanichthan et al. [42] in their paper used double layered approach i.e. SVM and Naive Bayes to classify the data. In their work the emphasis was on R2L and U2R but overall the accuracy suffered. Inspired by swarm intelligence concept, Mirjalili S. et al. [21] developed gray wolf optimization algorithm which used swarm technique to combine multiple greedy best solutions to update the subsequent solutions. Here the optimal solution was reached by calculating the fitness function. E. Emary et al. [5] in their paper has proposed the binary version of Gray wolf optimizer for feature selection. The proposed approach was validated by using it over a number of datasets. However, for NSL-KDD dataset no experiments were performed. M. Safaldin et al. [28] in their work used five leader wolves to guide the new positions of all wolves instead of four. The paper discussed about the enhancement ratio concept and compared the results according to population size. Apart from the above mentioned works, some of the important related work is depicted in the Table 1.

### 3.1. Proposed PrGWO-SK

Fig.1 depicts the proposed methodology of PrGWO-SK in graphical form. The proposed methodology starts with acquisition of net-

*Table 1. Related work at a glance*

| References | Year | Main Technique | Feature sel. | Important notes |
|---|---|---|---|---|
| Pajouh et al. [25] | 2017 | Linear discriminant analysis | Yes | Both NB and kNN used for two tier feature selection and classification |
| Shone et al. [31] | 2018 | S-NDAE | Yes | Used S-NDAE and RFA |
| Yao et al. [45] | 2019 | HMLD | Yes | Hybrid feature selection and hybrid classification through SVM and ANN |
| Gu et al. [10] | 2019 | DT-EnSVM2 | No | Used SVM along with feature transformation |
| Kumar et al. [18] | 2021 | Xgboost and kNN | Yes | Used information gain, correlation coefficient to reduce the features based on importance |
| Gao et al. [7] | 2019 | Adapting ensemble | Yes | Used multiple ML algorithms like RF, DNN, kNN, RF and using the voting mechanism for ensembling them |
| Golrang et al. [8] | 2020 | NSGAII-ANN | Yes | Proposed the use of Random Forest and multi-objective fitness criteria for FS |
| Wisanwanichthan et al. [42] | 2021 | DLHA | Yes | Division of features was done to use Naive Bayes with SVM |
| Gu et al. [9] | 2020 | NB-SVM2 | No | Used the Naive Bayes concept to enhance the features of data for classification by SVM |
| Teng et al. [38] | 2018 | DT | Yes | Used the SVM along with DT for adaptive classification |
| Alazzam et al. [1] | 2020 | Sigmoid_PIO | Yes | Pigeon inspired algorithm for feature selection. |
| Tama et al. [35] | 2019 | PSO | Yes | Used two stage classifier ensembles for anomaly detection. |
| Wei et al. [41] | 2020 | NIA+GHSOM-pr | Yes | Used Jaccard's coefficient for measuring fitness |
| Safaldin et al. [28] | 2020 | GWOSVM-IDS | Yes | Modified the GWO algorithm to detect anomalies |
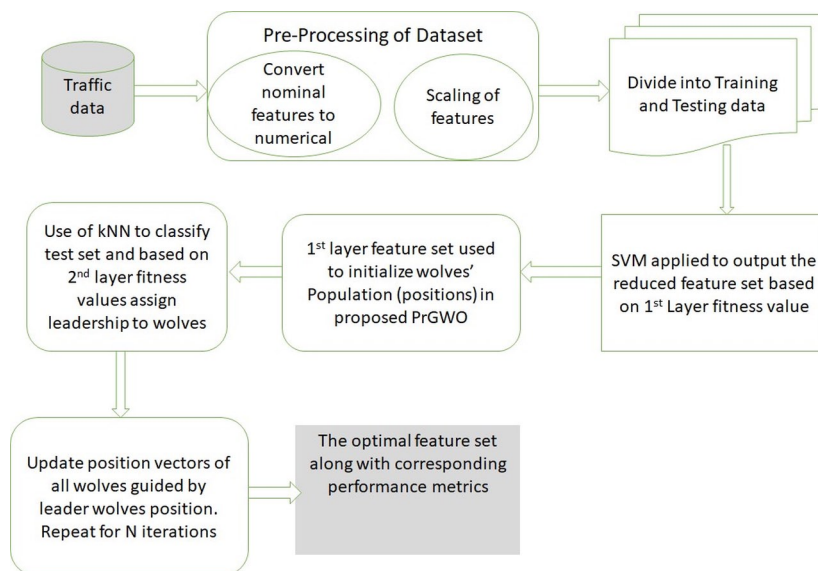
*Fig. 1. Methodology of the PrGWO-SK algorithm*

work traffic data for which NSL-KDD[36][37], BoTIoT[17] and DS2OS[23] datasets are used. NSL-KDD dataset is a legacy and widely used network dataset for intrusion detection work thus making it possible to compare the present work with others work in a comprehensive way. The DS2OS dataset is a collection of traces captured at the application layer in IoT environment while the BoTIoT dataset is a relatively recent dataset captured by designing network environment. Thus the testing of proposed approach on three different datasets-each one having their unique characterstic- helped validate the proposed work. Secondly, the used dataset is pre-processed using section 3.2 to convert into a form acceptable to the algorithm. Thirdly the data set is divided into 80% training and 20% testing subsets through random selection. Fourthly, the data is passed to the first layer (SVM with Radial basis Function(rbf)) which used first layer fitness function for generating reduced feature set. The SVM with rbf kernel has the advantage of seperating non-linearly seperable, closely related classes in dataset. The reduced feature set is passed to the second layer which used this to initialise the population of search agents of proposed algorithm 2 (These initialized search agents represent distinct feature sets). Initialization of search agents is an important step in swarm intelligence algorithms thus this work focusses on this aspect specifically. For achieving this two layered model is used in PrGWO-SK. Here the first layer exclusively considers accuracy as the basis for classification as it is one of the most important metrics in context of security mechanism. Thus the objective of first layer is solely to provide good solution for initialization of swarm population.

In the second layer the focus is not only on accuracy but also on other performance metrics, especially the length of feature vector finally selected. Thus, the second layer seeks to optimize both the length of feature vector alongwith accuracy and other important measures. To reiterate this can then be conceived as the multi objective optimization problem.

The algorithm then used these initialized search agents coupled with kNN as wrapper technique for generating classification results. kNN is used for its utility to evolve with new datapoints i.e new attack and normal access types as it does not involve explicit function generation. Thus it can give flexible non linear decision boundaries in a simple way. The relative performance of these distinct feature sets is evaluated through use of second layer fitness function. Accordingly four top search agents are assigned as leader wolves. In each subsequent iteration the position of each search agent is updated with the guidance of assigned leader wolves (four best search agents of the previous iteration). At the end of last iteration the feature set corresponding to best search agent is assigned as the optimal feature vector and passed to the real world classifier for detection.

Algorithm 1 gives step by step account of PrGWO-SK. Notations used in the algorithm are specified below. $W_i$ (i=1..N) denotes the wolves, FNR stands for False negative rate, TNR for True negative rate; rest of the notations are defined in the algorithm itself. In the first phase/layer data preprocessing is done using techniques of section 3.1. Also the number of wolves participating in the algorithm are specified as parameter. This population of wolves is initialized randomly and algorithm 2 called with following parameters- Initialized population vector, fitness function 1 and wrapper method as SVM. The algorithm 1 stored the returned reduced feature vector. In the second layer/phase the feature vector returned by the first layer is used to initialize the population of wolves. All features of the returned feature vector are mandatorily selected union random selection over rest of the features. Once again the algorithm 2 is called with this set of initialized population, fitness function 2 and wrapper method kNN. The returned feature vector by algorithm 2 is stored as optimal feature vector and given to the real world classifier for classification using only the features present in this optimal feature vector.

### 3.2. Preparing dataset and preprocessing

NSL-KDD: Published in 2009, it is an enhanced version of the CUP99 dataset from KDD. Before the assessment of NSL-KDD, several researchers used the KDD'CUP99 dataset. But KDD'CUP99 has many duplicates and this makes the dataset redundant and biased towards some of the attacks. NSL-KDD has got 41 independent features and one dependent feature. These 41 independent features can broadly be classified as- basic features, content features and traffic features. Regarding dependent feature it is categorised into 39 attack types. Broadly the attack types can be categorised into DoS, U2R, Remote to local (R2L) and Probe types. Additionally one type present is 'normal' to denote normal class. Regarding approximate distribution of various attack and normal types it is as:
Normal=53.45%, Probe=9.25%, DoS=36.45%, U2R=0.04% and R2L=0.78%.

DS2OS: Published in 2018, this open-source data set was obtained via Kaggle. In a virtual IoT environment, the distributed smart space orchestration system (DS2OS) is used to create the dataset. The entire virtual architecture, which is a collection of many micro-services in an IoT context. In the DS2OS dataset there are 12 independent features- majority having nonnumerical values. Regarding attack types these can be categorised into 7 types- DoS, Malicious control (MC), Malicious Operations (MO), Probe, Scan, Spy and Wrongsetup (WS). Regarding approximate distribution they are as:
Normal=97.2%, DoS=1.59%, MC=0.25%, MO=0.23%, Probe=0.09%, Scan=0.43%, Spy=0.14% and WS=0.03%.

BoTIoT: Published in 2019, the UNSW Canberra Cyber Range Center's practical network configuration was built in order to build the BoT-IoT dataset. A workable substitute for IoT solutions, this data collection is produced utilising the message queuing telemetry transport (MQTT) protocol, which connects machine-to-machine interactions. In total there are 43 independent features while the attack categories present in BoTIoT dataset are 4 in number- DoS, Distributed denial of service (DDoS), Reconnaissance and theft. Regarding approximate distribution of attack and normal types it is as:

Normal=0.012%, DDoS=52.5%, DoS=44.9%, Recon=2.4% and theft=0.002%.

As discussed before, all datasets are preprocessed before they can be used in the actual ML algorithm. Below described techniques are used to transform them in a form amenable to classification algorithms:

*Handling Missing values:* Very often the dataset contains few missing values which affect the classification model's performance. In case of NSL-KDD and BoTIoT there are no missing values, while for DS2OS dataset the missing values are replaced with "missing" term.

*Feature mapping:* A few independent features like protocol type, flag and service are of nominal type which needs to be converted into numerical type. In the literature approaches like one hot encoding have been used. However, one hot encoding makes the dataset much more sparse thereby increasing the computation complexity. Therefore, in the present work ordinal encoding is used for converting nominal into numerical values. For example for protocol_type attribute the original values are converted as: TCP=1, UDP=2, ICMP=3. The same technique is used for other features also.

*Feature normalization:* Feature normalization or scaling is required to infuse uniformity in the values of features. In absence of feature normalization the higher values tend to dominate the trained model parameters. In the present work, all the independent features are scaled to a uniform level using min-max scaling technique:

$$F = \frac{f - f_{min}}{f_{max} - f_{min}} \tag{1}$$

Here, $f$ is feature which is normalized, $f_{max}$ is a feature's maximum value, $f_{min}$ is its minimum value present in the dataset.

After scaling is done the next step is to split entire data set into training and test subdataset. Random selection technique is used for dividing the entire dataset into 80:20 ratio. The algorithm was repeated for 10 runs with different train:test subset using randomization. This is required to avoid chance selection bias in results.

---

**Algorithm 1: PrGWO-SK**

**Input:**

D = $D_i \cup D_d$  where \⋆ $D_i$ are the independent features and $D_d$ the label in the dataset ⋆\

**Output:**

$F_{optimal}$ : The set of optimal features
**Steps:**
1. Load .csv file

2. D= ordinal encoding(D)   \⋆ Convert the non-numeric data into distinct numeric types. ⋆\

3. $D_i$ = D.drop(label)   \⋆ Get the matrix of independent features ⋆\

4. $D_d$ = D.drop( $D_i$ )   \⋆ Get the column vector of dependent feature or label ⋆\

5. $D_i$ = Normalize( $D_i$ )   \⋆ Scale the dataset $D_i$ to create uniformity among various features.

6. Initialize N value   \⋆ No. of wolves or search agents used ⋆\

7. Repeat for 10 times:

  7.1. $[D_{train}, D_{test}]$ = Random selection on ( $D_i \cup D_d$ ) \⋆ Random selection to create training:test as 80:20
**Layer 1:**
  7.2. Initialize and call PrGWO with
    $W_i$ (i=1..N) = [ $X_i^d$ ] where $X_i^d \in \{0,1\}$.    \⋆ d is the vector dimension and i is the search agent no. ⋆\,
    Fitness = $fit_{1layer}$ and
     Classifier= SVM (rbf)
  7.3. $F_{SVM}$  = W[alpha]
**Layer 2:**
  7.4. Initialize and call PrGWO with:
$W_i\left(i=1....N\right)=\left[F_{SVM} \cup \left(D_i - F_{SVM}\right)\in \{0,1\}\right]$,

Fitness = $fit_{2layer}$ and

 Wrapper Method = kNN

  7.5. $F_{optimal}$ = W[alpha]

Output $F_{optimal}$ , Acc, DR, FPR, FNR, TNR, Precision, F1-score.

---

### 3.3. Proposed PrGWO

**Background**: The Gray wolf optimizer algorithm [21] is inspired by the metaheuristic approach which searches through the solution space for optimal solutions. Notionally, here a pack of leader wolves is responsible for guiding the followers' pack towards the prey's location and then hunting it down. The leader pack consists of best three solution vectors and are called as alpha, beta, delta wolves. Though all wolves start from a random solution vector but they move towards optimal solution vector through help of their leader wolves.

The update equation for all wolves was given as:

$$W_i\left(iter+1\right) = \frac{\omega 1 + \omega 2 + \omega 3}{3} \tag{2}$$

Here:

$$\omega 1 = \left|W\left[alpha\right] - A_1 . D\left[alpha\right]\right| \tag{3}$$

$$\omega 2 = \left|W\left[beta\right] - A_2 . D\left[beta\right]\right| \tag{4}$$

$$\omega 3 = \left|W\left[delta\right] - A_3 . D\left[delta\right]\right| \tag{5}$$

where $W_i$(iter+1) is the i$^{th}$ wolf position in next iteration, $W\left[alpha\right], W\left[beta\right], W\left[delta\right]$ are the positions of the Alpha, Beta and Delta wolves, $A_i$ are the factor values and D[alpha], D[beta], D[delta] are the distance vectors of i$^{th}$ wolf from alpha, beta and delta wolves' position vector respectively.

For adapting the general algorithm having continuous values to feature selection task with each feature represented by only binary variable $\in 0,1$ the above algorithm was modified [5]. In this algorithm the mapping was done from continuous values to the binary values. In the modified scenario, the update equation was written as:

$$W_i\left(iter+1\right) = crossover\left(\omega 1, \omega 2, \omega 3\right) \tag{6}$$

Here crossover is defined as :

$$crossover(\omega1,\omega2,\omega3)=\begin{cases}\omega1 & if \quad rand<\frac{1}{3}\\[2mm]\omega2 & if \quad \frac{1}{3}\le rand<\frac{2}{3}\\[2mm]\omega3 & if \quad rand\ge\frac{2}{3}\end{cases} \qquad (7)$$

$$\omega1=\begin{cases}1 & if \quad (W[alpha]+step[alpha])\ge1\\0 & otherwise\end{cases} \qquad (8)$$

$$step[alpha]=\begin{cases}1 & if \quad sigmoid\left(10\left(A_1 D[alpha]-.5\right)\right)>rand\\0 & otherwise\end{cases} \qquad (9)$$

Same equations (6)-(9) are used for other leader wolves- beta, delta and omega with relevant substitutions.

**PrGWO**: In PrGWO, firstly instead of using only three leaders four active leader wolves have been taken. To each non-leader wolf this provides more variety; availability of increased number of leader position vectors for updating their own position vector.

Secondly, the criteria of relative importance of individual four leaders in influencing updation of positions of all the other non-leader wolves was incorporated. For instance, alpha wolf is the best leader among all leader wolves hence it needs to be given higher weight influence than the other three. Similarly beta, delta and omega wolves need to be given weightage in order of their importance. Unlike proposed in the present paper, the original GWO and bGWO had given equal weight influence to all the wolves disregarding the fact that the alpha wolf is the most important leader in terms of acquiring best position followed by beta, delta and omega wolves respectively. In this proposed algorithm, different impact factor to different wolves viz alpha, beta, delta and omega wolves is proposed.

The mathematical model used in the proposed algorithm is as:

$$W_i(iter+1)=\begin{cases}\omega1 & if \quad rand<\frac{1}{2}\\[2mm]\omega2 & if \quad \frac{1}{2}\le rand<\frac{3}{4}\\[2mm]\omega3 & if \quad \frac{3}{4}\le rand<\frac{9}{10}\\[2mm]\omega4 & otherwise\end{cases} \qquad (10)$$

$$\omega1=\begin{cases}1 & if \quad \left(sigmoid\left(W[alpha]-A_1D[alpha]\right)\right)\ge rand\\0 & otherwise\end{cases} \qquad (11)$$

$$\omega2=\begin{cases}1 & if \quad \left(sigmoid\left(W[beta]-A_2D[beta]\right)\right)\ge rand\\0 & otherwise\end{cases} \qquad (12)$$

$$\omega3=\begin{cases}1 & if \quad \left(sigmoid\left(W[delta]-A_3D[delta]\right)\right)\ge rand\\0 & otherwise\end{cases} \qquad (13)$$

$$\omega4=\begin{cases}1 & if \quad \left(sigmoid\left(W[omega]-A_4D[omega]\right)\right)\ge rand\\0 & otherwise\end{cases} \qquad (14)$$

$\omega i=1...4$ are binary vectors and $W_i$ (iter+1) are the updated position vectors of individual non-leader wolves. In equation no.10 different weights of influence have been assigned to different leader wolves

based on their relative importance. For example, if the random number (generated in range [0,1]) is less than 0.5 the updation of non-leader wolf is based on alpha wolf. This weight of influence is reduced respectively for the beta, delta and omega wolves as 0.25, 0.15 and 0.10 respectively. This is contrary to the original bGWO where the weight influence for all the wolves was uniformly distributed as 0.33.

Thirdly, for alpha, beta, delta and omega wolves (leaders guiding the other wolves)a different criteria for updating of their own positions is proposed. Each leader wolf updates its position relative to its own, its predessesors and its immediate successor. This was done to ensure that already better position holding wolves do not get deviated under influence of other less efficient successor leader wolves except its immediate successor. Immediate succcessor was required as in absence of it the alpha wolf position would not have updated and exploration of search space would have been restricted. Accordingly, updation of alpha wolf position was done relative to its own and beta wolf's position, updation of beta wolf's position was done with respect to alpha, beta and delta wolves' position, delta wolf position was done with respect to positions of alpha, beta, delta and omega wolves while for omega wolf it was with respect to all leader wolves. For updation of alpha,beta, delta wolves equations 15-17 were used.

$$W[alpha](iter+1)=Crossover(\omega1,\omega2) \qquad (15)$$

$$W[beta](iter+1)=Crossover(\omega1,\omega2,\omega3) \qquad (16)$$

$$W[delta](iter+1)=Crossover(\omega1,\omega2,\omega3,\omega4) \qquad (17)$$

$$Crossover(\omega1,\omega2)=\begin{cases}\omega1 & if \quad rand<\frac{2}{3}\\[2mm]\omega2 & otherwise\end{cases} \qquad (18)$$

$$Crossover(\omega1,\omega2,\omega3)=\begin{cases}\omega1 & if \quad rand<\frac{1}{2}\\[2mm]\omega2 & if \quad \frac{1}{2}\le rand<\frac{4}{5}\\[2mm]\omega3 & otherwise\end{cases} \qquad (19)$$

$$Crossover(\omega1,\omega2,\omega3,\omega4)=\begin{cases}\omega1 & if \quad rand<\frac{1}{2}\\[2mm]\omega2 & if \quad \frac{1}{2}\le rand<\frac{3}{4}\\[2mm]\omega3 & if \quad \frac{3}{4}\le rand<\frac{9}{10}\\[2mm]\omega4 & otherwise\end{cases} \qquad (20)$$

Omega wolf can also be updated similar to delta wolf using equation 17. From eq 18-20 it can be seen that for updation of alpha wolf's position, ratio of 0.66:0.33 in terms of weight of influence is used between alpha and beta wolves. Similarly for beta wolf's updation ratio used is 0.5:0.3:0.2. In case of the delta wolf's updation, the weights of influence used are 0.5:0.25:0.15:0.10.

Fig. 2 depicts the PrGWO in graphical form while algorithmic depiction is presented as Algorithm 2. In this algorithm the initialized N feature vectors corresponding to N wolves are used to create N models. These models are then applied to the test data to classify the data. Based on the classification report and fitness function, fitness value of each wolf's feature vector is calculated. The fitness values are sorted in ascending order and the position/feature vector corresponding to the best fitness value is assigned as alpha position. Similarly the
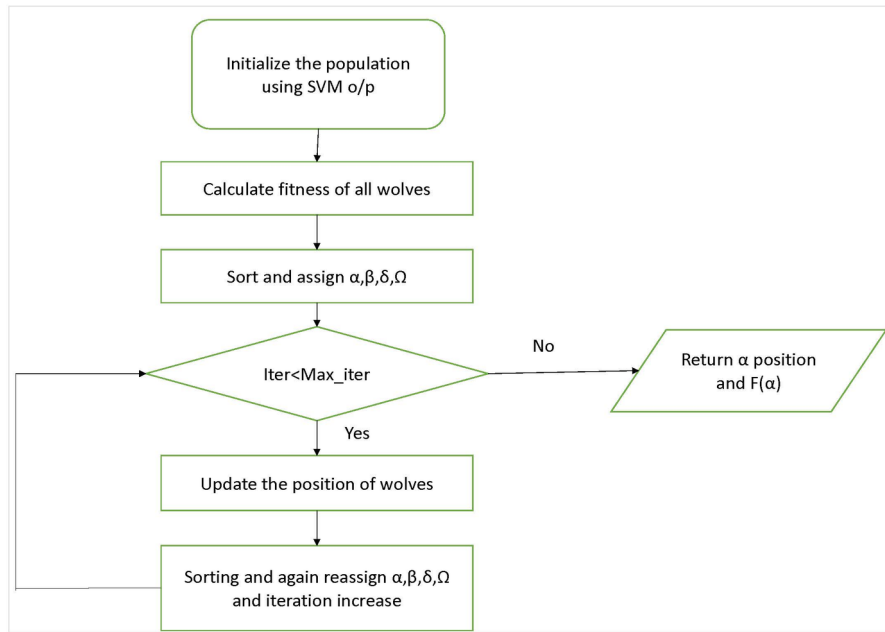
*Fig. 2. Flowchart of PrGWO*

second best is assigned as beta, third best as delta and fourth one as omega. In the subsequent iteration these values(alpha, beta, delta and omega) are used to update the position vectors of other wolves and their own using equations enlisted in the algorithm. Again the fitness value of each wolf's updated feature vector is calculated. Based on these newly sorted fitness values, reassignment of alpha, beta, delta and omega is done. The cycle is repeated till the last iteration. At the end of last iteration the alpha position vector is returned as optimal reduced feature vector.

### 3.4. Proposed fitness function

Fitness function has an important role in determining the optimal number of features for use in classification algorithms like SVM and kNN without compromising on performance metrics. For systematic initialization of the Gray wolves in the second layer, SVM is used as the wrapper method in the first layer. In this first layer, the accuracy measure has been used for getting the reduced feature set. This is done to ensure that the most important measure of efficiency is accounted for exclusively in the first layer; the other measures being accommodated in the second layer later on without sacrificing accuracy.

Thus the fitness function used in first layer is:

$$fit_{1layer} = 1 - \left( \frac{cor}{tot} \right) \qquad (21)$$

where $fit_{1layer}$ is the fitness function of first layer, *cor* is the correctly predicted instances and *tot* are the total number of instances. As for the second layer wherein PrGWO along with kNN is being used there are other important metrics like Detection rate, False positive rate and number of features which needs to be incorporated to make the fitness function truly inclusive. However, recognizing the accuracy as the most important measure for ensuring good security here again, a fitness model is proposed which gives higher weightage to the accuracy.

---

**Algorithm 2: PrGWO**

---

**Input:**
Iteration = max_iteration
Initialized $W_i$ (i=1....N)

*FitnessFn*
Wrapper method {SVM,kNN}

**Output:**
W[alpha] \⋆ Optimal feature set got after the algorithm is complete ⋆\

f(W[alpha]) \⋆ Fitness value corresponding to optimal feature set ⋆\

**Steps:**
1.Initialize A
2. For all $W_i$

  2.1 $Model_i$ = Classifier( $D_{train}$ )
  2.2 $Predict_i$ = Classifier(Model$_i$, $D_{test}$ )
  2.3 Fitness value $fit_i = Fitnessfn(Predict_i)$
3. Sort( $fit_i$ ),corresponding $W_i$ and assign:

 W[alpha]= $W_i$ corresponding to best( $fit_i$ )

 W[beta] = $W_i$ corresponding to $2^{nd}$ best( $fit_i$ )

 W[delta]= $W_i$ corresponding to $3^{rd}$ best( $fit_i$ )

 W[omega]= $W_i$ corresponding to $4^{th}$ best( $fit_i$ )
4. While(iterations < max_iteration)
 4.1. For i=1 to N do

  *if* ($W_i$ ==W[alpha])   Apply eq. 15

   *elseif* ($W_i$ ==W[beta])   Apply eq. 16

   *elseif* ($W_i$ ==W[delta] or W[omega])   Apply eq. 17
  Else Apply eq. 10
  Endif
  endfor
 4.2.Update A
 4.3.Repeat step 2.1-2.3
 4.4.Sort( $fit_i$ ),corresponding $W_i$ and reassign W[alpha],   W[beta], W[delta] and W[omega].
Return W[alpha] and f(W[alpha])

---

The proposed fitness function is as:

$$fit_{2layer} = a.error + \left(\frac{1-a}{2}\right).DR_{error} + \left(\frac{1-a}{2}\right).FPR + (b.Nfeat) \quad (22)$$

$$error = 1 - \frac{cor}{tot} \quad (23)$$

$$DR_{error} = 1 - \frac{Tpos}{Tpos + Fneg} \quad (24)$$

$$FPR = \frac{Fpos}{Tneg + Fpos}. \quad (25)$$

where $fit_{2layer}$ is the fitness function of second layer, $Tpos$ is True-Positive cases, $Fneg$ is false negative, $Fpos$ is false-positive and $Tneg$ is true-negative. The last term $Nfeat$ is the number of features. Multiplying by b factor is necessary to scale this term to the level of other terms in the expression. Lastly, in the model the goal of the algorithm is to minimize the fitness function. The least valued fitness function will be assigned as alpha wolf and its space position will be the best feature set.

### 3.5. Performance metrics used

Present work was tested using various quality parameters. Equation(26)-(34) depicts these parameters:

$$Acc = \frac{True\_pos + True\_neg}{True\_pos + False\_pos + True\_neg + False\_neg} \quad (26)$$

$$DR = \frac{True\_pos}{True\_pos + False\_neg} \quad (27)$$

$$FPR = \frac{False\_pos}{False\_pos + True\_neg} \quad (28)$$

$$TNR = \frac{True\_neg}{False\_pos + True\_neg} \quad (29)$$

$$FNR = \frac{False\_neg}{True\_pos + False\_neg} \quad (30)$$

$$PR = \frac{True\_pos}{False\_pos + True\_pos} \quad (31)$$

$$F1-score = \frac{2 \times DR \times PR}{DR + PR} \quad (32)$$

$$Feat = No.\,of\ features \quad (33)$$

$$Ratio(R) = \frac{Acc}{Feat} \quad (34)$$

True positive signifies that the data point was predicted as malicious and actually it was malicious. False Positive signifies that the data point was predicted as malicious, however actually it was not. True negative signifies that the data point was predicted as not malicious and actually it was not malicious. False negative signifies that the data point was predicted as not malicious, however it was actually malicious.

Detection Rate (DR) is also known as recall and TPR. It identifies the number of times the classifier predicted a 'positive' result over the number of times positive results were to be predicted.

Precision (PR) determines the measure of correctness of results got. Thus it uses true positive over true positive and false positive combined.

F1-score is used when class distribution is not balanced. It calculates the weighted average of detection rate and precision.

## 4. Experimental results and discussion:

Experiments were performed using Matlab 2021 and Python programming language on anaconda distribution over the Intel Core i7-10th generation machine.

*Table 2. Table of parameters*

| Parameter name | Parameter values |
|---|---|
| No. of runs | 10 |
| No. of population-N | 5,7,10 |
| No. of iterations | 25 |
| Technique for feature selection in 1st layer | SVM(RBF) |
| Fitness function for SVM-RBF | $fit_{1layer}$ |
| Technique used FS and classification in second layer | PrGWO with kNN |
| Fitness function for PrGWO-kNN | $fit_{2layer}$ |
| a in fitness function- $fit_{2layer}$ | 0.7 |
| b in fitness function $fit_{2layer}$ for optimal accuracy | 0.0001 |
| b in fitness function for optimal no. of features | 0.001 |

### 4.1. Parameters used:

Various parameters were tested experimentally for finding the best combination of parameters. Table 2 illustrates these parameters. The algorithm was run 10 times, Number of wolves were taken to be 5,7 and 10, No of iterations within each run was 25, SVM was used with RBF, two layers were used with different fitness functions and values of b were taken as 0.001 and 0.0001. The reason for taking different values of b is to give different weightage to number of features in the fitness function. In table 3 variation in results due to size of population is depicted. Notationwise NoF in the table stands for No. of features. Here 3 values of N were taken: 5,7 and 10 wolves. The results were compared not only in regard to DR, FPR, F1-score etc but also accuracy and length of feature set finally selected by the algorithm. The results show that when the size of population was 7, the length of feature set was least and accuracy was maximum. Only two parameters-FPR and TNR were found to be better in case of N=10. Similarly the 'k' value in the kNN was experimented for k=1,3 and 5. Table 4 shows the variation in results with change in k value. For k=1, the experiments yielded best results in terms of accuracy i.e 99.60% alongwith other metrics However, the best length of feature vector i.e 8 was found for k=3 though the accuracy and other metrics were not better than k=1.

Hence based on experimental results it is appropriate to remark that a cost benefit analysis needs to be done in terms of accuracy and number of features while choosing value of k for real world scenarios.

Table 3. Variation of results with value of 'N' for NSL-KDD

| N | Accu-racy | NoF | DR | FPR | TNR | FNR | Preci-sion | F1 |
|---|---|---|---|---|---|---|---|---|
| 5 | 99.46 | 13 | 99.44 | 0.34 | 99.66 | 0.55 | 99.45 | 99.44 |
| 7 | **99.60** | **12** | **99.61** | 0.28 | 99.72 | **0.39** | **99.61** | **99.61** |
| 10 | 99.47 | 14 | 99.45 | **0.27** | **99.73** | 0.54 | 99.46 | 99.46 |

Table 4. Variation of results with value of 'k' for NSL-KDD

| k | NoF | Accuracy | DR | FPR | TNR | FNR | Precision | F1 |
|---|---|---|---|---|---|---|---|---|
| 1 | 12 | **99.60** | **99.61** | **0.28** | **99.72** | **0.39** | **99.62** | **99.61** |
| 3 | **8** | 99.32 | 99.33 | 0.53 | 99.47 | 0.67 | 99.30 | 99.32 |
| 5 | 10 | 99.26 | 99.26 | 0.53 | 99.47 | 0.74 | 99.25 | 99.25 |

Table 5. Classwise performance of PrGWO-SK for NSL-KDD dataset

| Class | DR | FPR | TNR | FNR | Precision | F1 score |
|---|---|---|---|---|---|---|
| Normal | 99.76 | 0.478 | 99.52 | 0.24 | 99.58 | 99.67 |
| DoS | **99.89** | 0.062 | 99.94 | **0.11** | **99.89** | **99.89** |
| Probe | 98.75 | 0.061 | 99.94 | 1.25 | 99.39 | 99.07 |
| R2L | 88.58 | **0.036** | **99.964** | 11.41 | 95.57 | 91.94 |
| U2R | 66.67 | 0.043 | 99.957 | 33.33 | 42.10 | 51.61 |

Table 6. Classwise performance of PrGWO-SK for DS2OS dataset

| Class | DR | FNR | FPR | TNR | Precision | F1-score |
|---|---|---|---|---|---|---|
| DoS attack | 99.45 | 0.55 | 0.0015 | 99.9985 | 99.48 | 99.46 |
| Data Probing | 99.75 | 0.25 | 0.0002 | 99.9998 | 99.76 | 99.75 |
| Malicious control | 99.6 | 0.4 | 0.0013 | 99.9987 | 99.41 | 99.50 |
| Malicious operation | 99.8 | 0.2 | 0.0025 | 99.9975 | 98.91 | 99.35 |
| Scan | 98.73 | 1.27 | 0.0054 | 99.9946 | 98.65 | 98.69 |
| Spying | **99.84** | **0.16** | **0** | **100** | **99.77** | **99.80** |
| Wrongsetup | 99.74 | 0.26 | **0** | **100** | 99.75 | 99.74 |
| Normal | 99.81 | 0.19 | 0.19 | 99.81 | 99.76 | 99.78 |

Table 7. Classwise performance of PrGWO-SK for BoTIoT dataset

| Class | DR | FNR | FPR | TNR | Precision | F1-score |
|---|---|---|---|---|---|---|
| DoS | 99.97 | 0.03 | **0.0001** | **99.9999** | 99.81 | 99.89 |
| DDoS | **99.98** | **0.02** | **0.0001** | **99.9999** | **99.92** | **99.95** |
| Reconnaissance | 99.89 | 0.11 | 0.0008 | 99.9992 | 99.79 | 99.84 |
| Theft | 96.2 | 3.8 | 0.0045 | 99.9955 | 99.15 | 97.65 |
| Normal | 99.2 | 0.8 | 0.0009 | 99.9991 | 98.9 | 99.05 |

## 4.2. Performance of PrGWO-SK for different classes:

The performance of present approach was tested on different classes of dataset. Table 5 shows the measures for the Detection rate, False positive rate, F1-score, Precision etc for all the classes. The experimental results show DR to be highest for DoS attacks being 99.89% while for R2L and U2R the DR was least among all. However, the False positive rate was found to be the best for these two classes viz. 0.036% and 0.043% among all. Considering the precision, the

two classes -Normal and DoS performs to the extent of 99.58% and 99.89%.

As regards DS2OS dataset table 6 depicts the classification results for various classes present in the dataset. Most of the performance metrics shows best results for the spying class. In case of FPR and TNR both Wrongsetup and Spying classes shows the results as 0 and 100% respectively, i.e no false alarm is generated if the classes are neither Wrongsetup or Spying. However, the algorithm performs worst for the Scan class where the Detection rate was measured as 98.73% in contrast to the other classes where the DR was measured more than 99%.

Similar to DS2OS and NSl-KDD the classwise performance of the BoTIoT dataset was also evaluated. From table 7 it can be seen that for DDoS the algorithm performs best in terms of DR at 99.98%, FPR at 0.0001%, precision at 99.92% and F1-score at 99.95%. The class DoS shows results very close to the DDoS class. The least Detection rate was recorded for class 'Theft' at 96.2% which shows the difficuly in correctly predicting this class of attack.

## 4.3. Convergence of algorithm:

Table 8 depicts the various performance measures iteration-wise for b=0.0001. For this value of b the algorithm focusses more on getting optimal accuracy even at the cost of increasing the length of feature vector.

Similarly table 9 depicts the experimental results for b=0.001. Here the focus shifts more towards optimizing the length of feature vector.

Table 8. Iteration wise results for NSL-KDD with b=0.0001

| Itera-tion | Fea-tures | Accu-racy | DR | FPR | Preci-sion | F1-score |
|---|---|---|---|---|---|---|
| 1 | 19 | 99.506 | 99.510 | 0.306 | 99.500 | 99.505 |
| 5 | 15 | 99.569 | 99.572 | 0.268 | 99.581 | 99.576 |
| 10 | 15 | 99.565 | 99.568 | **0.241** | 99.577 | 99.572 |
| 15 | 13 | 99.577 | 99.579 | 0.275 | 99.586 | 99.582 |
| 16 | 13 | 99.577 | 99.579 | 0.275 | 99.586 | 99.582 |
| 17 | 12 | 99.577 | 99.579 | 0.275 | 99.586 | 99.582 |
| 18 | 13 | 99.577 | 99.579 | 0.255 | 99.586 | 99.582 |
| 19 | 13 | 99.577 | 99.579 | 0.255 | 99.586 | 99.582 |
| 20 | 13 | 99.577 | 99.579 | 0.255 | 99.586 | 99.582 |
| 21 | 13 | 99.577 | 99.579 | 0.255 | 99.586 | 99.582 |
| 22 | 13 | 99.585 | 99.587 | 0.257 | 99.597 | 99.592 |
| 23 | **12** | **99.603** | **99.606** | 0.284 | **99.616** | **99.611** |
| 24 | 12 | 99.603 | 99.606 | 0.284 | 99.616 | 99.611 |
| 25 | 12 | 99.603 | 99.606 | 0.284 | 99.616 | 99.611 |

Table 9. Iteration wise results for NSL-KDD with b=0.001

| Iteration | Feature | Accuracy | DR | FPR | Preci-sion | F1-score |
|---|---|---|---|---|---|---|
| 1 | 18 | 98.810 | 98.790 | 0.560 | 98.890 | 98.840 |
| 5 | 14 | 98.790 | 98.800 | 0.501 | 98.870 | 98.835 |
| 10 | 14 | 99.202 | 99.193 | 0.565 | 99.191 | 99.192 |
| 15 | 14 | 99.202 | 99.193 | 0.565 | 99.191 | 99.192 |
| 16 | 13 | 99.212 | 99.193 | 0.565 | 99.207 | 99.200 |
| 17 | 13 | 99.212 | 99.193 | 0.565 | 99.207 | 99.200 |
| 18 | 15 | 99.345 | 99.337 | **0.425** | 99.340 | 99.339 |
| 19 | 12 | 99.345 | 99.337 | 0.441 | 99.338 | 99.338 |
| 20 | 11 | 99.345 | 99.337 | 0.447 | 99.338 | 99.338 |
| 21 | 8 | 99.342 | 99.343 | 0.445 | 99.331 | 99.337 |
| 22 | 8 | 99.342 | 99.343 | 0.445 | 99.331 | 99.337 |
| 23 | 8 | 99.342 | 99.343 | 0.445 | 99.331 | 99.337 |
| 24 | **8** | **99.361** | **99.367** | **0.568** | **99.366** | **99.367** |
| 25 | 8 | 99.361 | 99.367 | 0.568 | 99.366 | 99.367 |
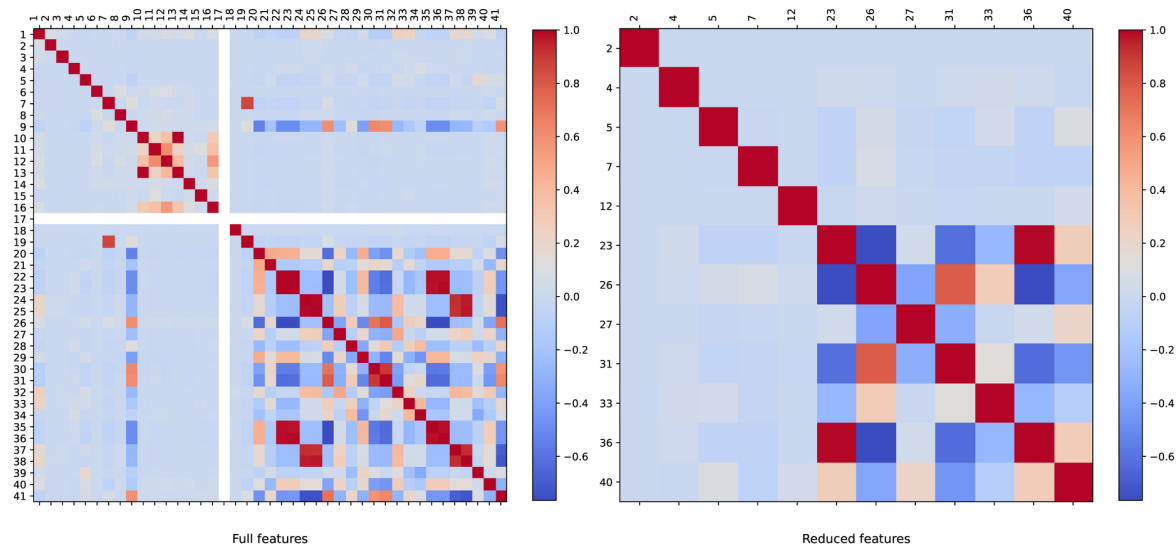


Fig 3. Original vs Reduced features Correlation heatmap of the NSL-KDD dataset
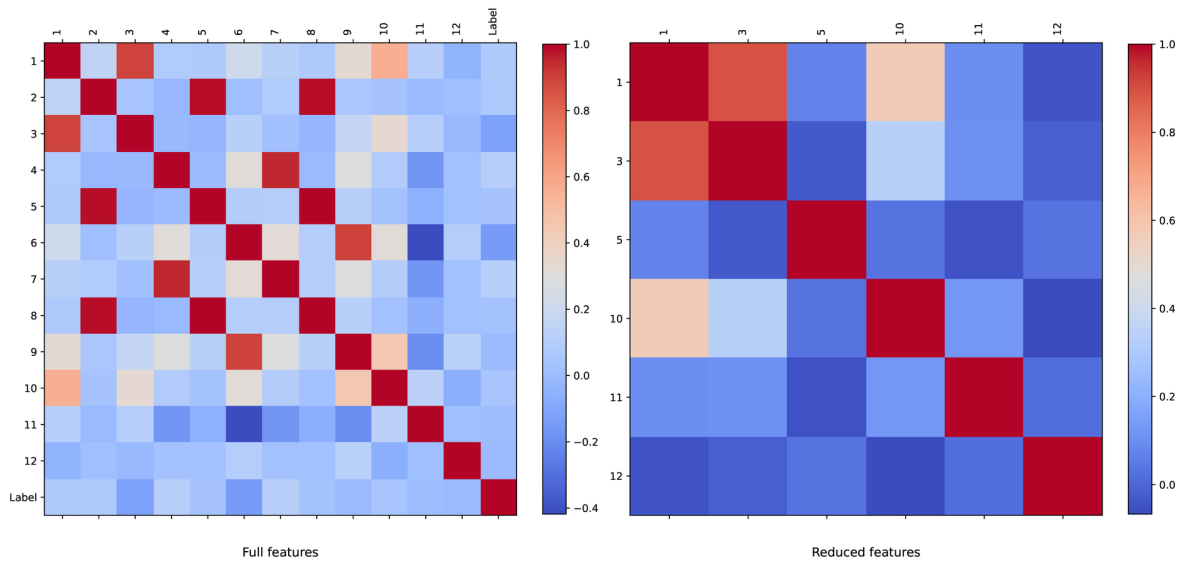


Fig 4. Original vs Reduced features Correlation heatmap of the DS2OS dataset
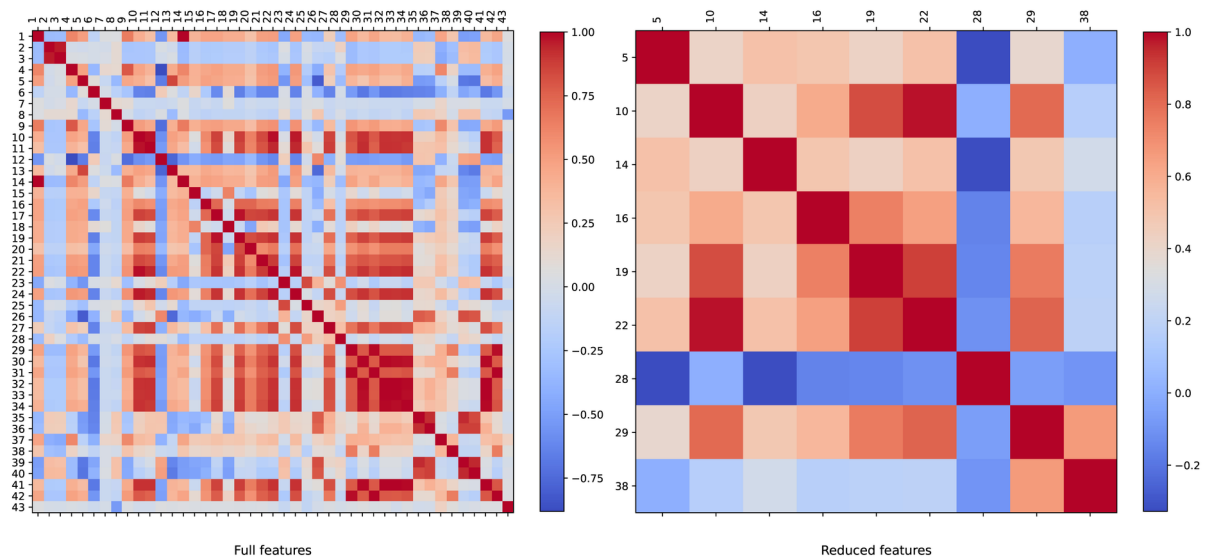
*Fig 5. Original vs Reduced features Correlation heatmap of the BoTIoT dataset*
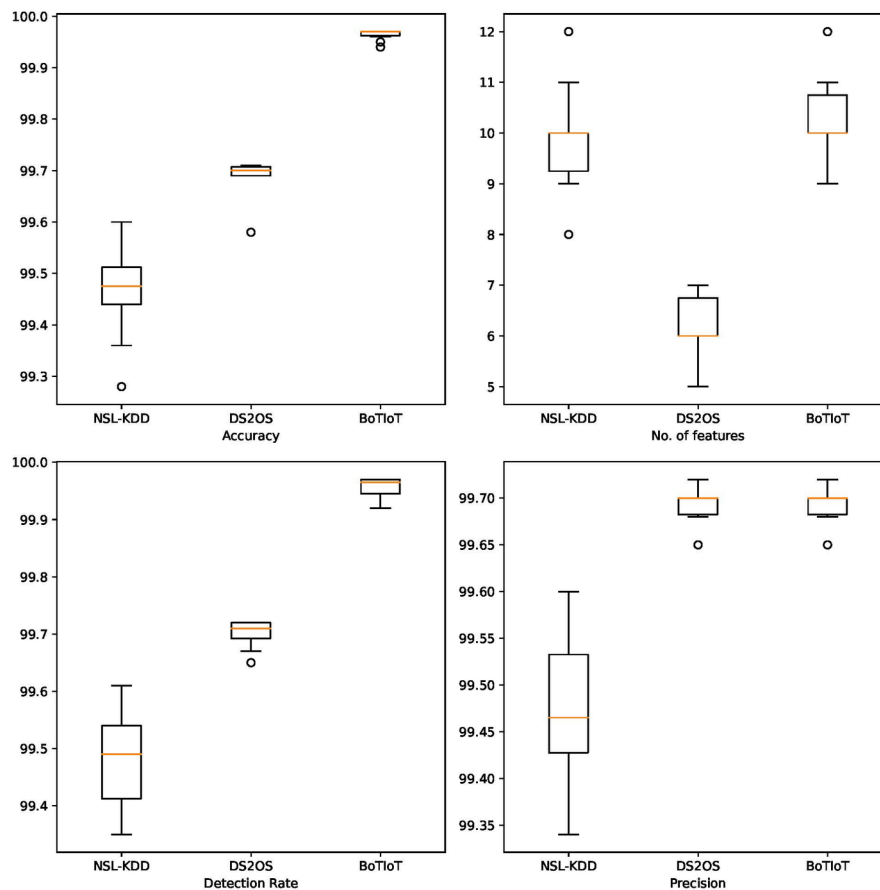


*Fig 6. Boxplots showing variation of experimental results for NSL-KDD, DS2OS and BoTIOT*

### 4.4. Correlation of original dataset vs reduced dataset

Figures 3, 4 and 5 shows the correlation heatmap for NSL-KDD, DS2OS and BoTIoT datasets respectively. The figures show both the correlation matrix of the original feature set vis-a-vis reduced feature set. Correlation is an important measure of finding the similarity between the features in pair. For optimizing the length of feature vector, similar or correlated features needs to be removed as they do not add any additional characterstic for classification task. The figures mentioned above shows how the final feature set outputted by the proposed algorithm does not contain the strongly correlated features i.e

the algorithm is largely able to remove the correlated features. Thus this test can be taken as statistical test for validating the proposed approach.

### 4.5. Consistency of experimental results through box plots

The values shown in the tables in this paper are the results of the best run during 10 runs of the algorithm. However, an algorithm's robustness and efficiency can be measured by its consistency over several runs. Fig 6 depicts the median and the interquartile range of the different runs of experimental results graphically. For NSL-KDD dataset

Table 10. DR -Classwise comparison of existing algorithm with proposed algorithm for NSL-KDD dataset

| Work | Normal | Probe | DoS | U2R | R2L |
|---|---|---|---|---|---|
| Pajouh et al. [25] | 94.56 | 79.76 | 84.68 | 67.16 | 34.81 |
| Wu et al. [43] | 95.31 | 83.4 | 83.49 | 11.55 | 24.69 |
| Gao et al. [7] | 94.33 | 87.11 | 84.37 | 25 | 55.27 |
| Yang et al. [44] | 97.38 | 73.94 | 81.09 | 6.5 | 17.25 |
| Zhang et al. [47] | 97.03 | 80.38 | 83.95 | 32.84 | 11.26 |
| Pajouh et al. [26] | 94.43 | 87.32 | 88.2 | 70.15 | 42 |
| Tian et al. [39] | 94.9 | 98.26 | 98.04 | 75.5 | 95.24 |
| Su et al. [34] | 97.5 | 85.76 | 87.55 | 20.95 | 44.25 |
| Mahfouz et al. [20] | 99.5 | 91.6 | 99.2 | 39.3 | 55.1 |
| RF [18] | 98.01 | **99.41** | 99.71 | 93.22 | **97.08** |
| Xgboost [18] | 99 | 96.86 | 93.68 | 83.92 | 80.11 |
| KNN [18] | 98 | 98.25 | 99.69 | **93.36** | 95.23 |
| Proposed work | **99.76** | 98.75 | **99.89** | 66.67 | 88.58 |

Table 11. DR-Classwise comparison of existing algorithm with proposed algorithm for DS2OS dataset

| Works | DoS | Probe | MC | MO | Scan | Spying | WS | Normal |
|---|---|---|---|---|---|---|---|---|
| Pahl et al. [24] | 68 | **100** | 32 | 66 | 54 | 13 | **100** | 95 |
| Hasan et al. [11] | 66 | 92 | 97 | **100** | 95 | 93 | **100** | **100** |
| Hasan et al. [11] | 66 | 92 | 92 | 56 | 71 | 4 | **100** | **100** |
| RF [18] | 66 | **100** | 97 | **100** | 97 | **100** | **100** | **100** |
| Xgboost [18] | 66 | **100** | 99 | **100** | **100** | **100** | **100** | **100** |
| KNN [18] | 66 | **100** | 99 | **100** | 98 | **100** | **100** | **100** |
| Proposed work | **99.45** | 99.75 | **99.6** | 99.8 | 98.73 | 99.84 | 99.74 | 99.81 |

Table 12. DR-Classwise comparison of existing algorithm with proposed algorithm for BoTIoT dataset

| Work | Normal | DDoS | DoS | Reconnaissance | Theft |
|---|---|---|---|---|---|
| Shafiq et al.[29] | 75 | 98 | **100** | 81 | 93 |
| Soe et al. [33] | 0 | **100** | **100** | **100** | 0 |
| RF [18] | **100** | **100** | 99 | **100** | 93 |
| Xgboost [18] | **100** | **100** | **100** | **100** | 93 |
| Proposed Work | 99.2 | 99.98 | 99.97 | 99.89 | **96.2** |

Table 13. Comparative analysis of works reported in the literature for NSL-KDD

| IDS | No. of features | Accuracy | Ratio |
|---|---|---|---|
| GHSOM [46] | 41 | 96.02 | 2.34 |
| A-GHSOM [15] | 41 | 96.63 | 2.35 |
| Kayacik [16] | 41 | 90.04 | 2.19 |
| DT-EnSVM2 [10] | 41 | 99.41 | 2.42 |
| NB-SVM2 [9] | 41 | 99.36 | 2.42 |
| S-NDAE+RF [31] | 28 | 85.42 | 3.05 |
| PCA+GHSOM-pr [13] | 30 | 87.23 | 2.90 |
| FDR+ kernel PCA [14] | 23 | 90 | 3.91 |
| NGSA +GHSOM-pr [41] | 27 | 98.61 | 3.65 |
| GWOSVM-IDS [28] | 12 | 96 | 8 |
| NNIA+GHSOM-pr [41] | 24 | 99.47 | 4.14 |
| kNN+GR+RFMDA [18] | 18 | 98.67 | 5.48 |
| Decision tree based [38] | 16 | 98.38 | 6.15 |
| SIGMOID_PIO [1] | 18 | 86.9 | 4.83 |
| RF+PSO [19] | 10 | 99.32 | 9.93 |
| Proposed PrGWO-SK(b=0.001) | 8 | 99.361 | 12.42 |
| Proposed PrGWO-SK(b=.0001) | 12 | 99.60 | 8.30 |

Table 14. Comparative analysis of works reported in the literature for DS2OS

| IDS | No. of features | Accuracy | Ratio |
|---|---|---|---|
| Pahl et al. [24] | 12 | 96.3 | 8.025 |
| Hasan et al. [11] | 11 | 99.35 | 9.031 |
| RF [18] | 6 | 99.4 | 16.566 |
| Xgboost [18] | 6 | 99.43 | 16.571 |
| KNN [18] | 6 | 99.4 | 16.566 |
| Proposed work | 6 | 99.71 | 16.618 |

Table 15. Comparative analysis of works reported in the literature for BoTIoT

| IDS | No. of features | Accuracy | Ratio |
|---|---|---|---|
| Shafiq et al. [29] | 43 | 98.35 | 2.287 |
| Soe et al. [33] | 8 | 99.1 | 12.387 |
| RF [18] | 10 | 99.99 | 9.999 |
| Xgboost [18] | 10 | 99.99 | 9.999 |
| KNN [18] | 10 | 85.92 | 8.592 |
| Proposed work | 9 | 99.97 | 11.107 |

the median value of accuracy is 99.475%, while the minimum value and maximum values are 99.6% and 99.28% respectively. However, this minimum value is the outlier as shown in the box plot and the mean is 99.463%. Similarly for DS2OS, the median, mean, minimum and maximum values are 99.7%, 99.688%, 99.58% and 99.71% respectively. For BoTIoT these values are 99.97%, 99.964%, 99.94% and 99.97%. The above results for all the datasets shows the consistency in performance of the algorithm except one or two outliers. Considering the number of features, NSL-KDD has median of 10, mean of 9.8 , minimum as 8 and maximum as 12. For DS2OS dataset these are 6, 6.2, 5 and 7 and for BoTIoT these are 10, 10.2, 9 and 11 respectively in order of median, mean, minimum and maximum. Similarly for DR, NSL-KDD has median of 99.58%, mean of 99.48%, minimum as 99.35% and maximum as 99.61%. For DS2OS median is 99.71%, mean is 99.70, minimum is 99.65% and maximum is 99.72%. For BotIoT median is 99.94%, mean is 99.96%, minimum is 99.92% and maximum is 99.97%. Regarding Precision values for NSLKDD median is 99.58%, mean is 99.47%, minimum is 99.34% and maximum is 99.6%. For DS2OS median is 99.7%, mean is 99.69%, minimum is 99.65% and maximum is 99.72%. Lastly for BoTIoT median is 99.96%, mean is 99.96%, minimum is 99.95% and maximum is 99.97%.

### 4.6. Comparison of Detection rates achieved by different algorithms with proposed algorithm for different classes:

For measuring effectivess of any algorithm it is important to compare it with other existing algorithms in terms of common performance metrics. Table 10, 11 and 12 shows the classwise performance of the proposed algorithm compared to other algorithms. Table 10 shows that the proposed approach outperforms for 'Normal' and 'DoS' classes giving 99.76% and 99.89% while for other classes different algorithms performs better. Table 11 shows that the proposed approach performs best for the DoS and MC attacks. For DoS it gives DR as 99.45% which is much better than the second best of 66%. For MC it gives 99.6%. Similarly, table 12 shows that the proposed approach performs best in terms of detecting "theft" attack type giving DR as 96.2% while the second best was 93%.

### 4.7. Comparison in terms of Accuracy and Length of feature set of different algorithms with proposed algorithm:

Tables 13, 14 and 15 shows the comparison of the proposed approach with others' work in terms of accuracy, number of features and the ratio. Table 13 shows that in terms of the accuracy two best performing algorithms are NNIA+GHSOM-pr (99.47%) and DT-EnSVM2

(99.41%). However, proposed approach has been able to outperform these efficient algorithms measuring 99.60%. In terms of length of feature vector the best performing algorithm is RF+PSO giving output feature vector of length 10. In this case also the proposed work outperforms RF+PSO giving vector of length 8. However, in this case the accuracy reduces to 99.36%.

Table 14 shows that most of the algorithms gave accuracy close to 99.43% with length of feature vector as low as 6. Though PrGWO-SK was not able to reduce the length of feature vector further below 6 but the accuracy showed substantial improvement measuring 99.71%.

Similarly table 15 shows the results for BoTIoT dataset. Here the best accuracy was achieved by Kumar et al. as 99.99% with feature length as 10. Considering the optimal feature length the best was achieved as 8 by Soe et al. However, the accuracy measured in this case was only 99.1%. Proposed approach was able to measure accuracy as 99.97% with length of feature vector as 9. Thus the proposed approach was able to outperform these algorithms - first algorithm in terms of length without compromising much on accuracy and second algorithm in terms of accuracy by a substantial margin.

## 5. Conclusions and limitations

This paper aimed to find reduced optimal feature vector of a traffic dataset to reduce the computational complexity by removing the data dimensionality curse. A two layered structure was used - first layer employed the SVM technique while second layer used the kNN as wrapper technique. For searching the solution space swarm intelligence based modified form of GWO called PrGWO was proposed alongwith two different fitness functions. The effectiveness of the experimental results were established through use of metrics - Accuracy, Detection rate, FPR, TPR, Precision. Through extensive experiments it was found that the proposed methodology and algorithm performed better for several individual classes -Normal, DoS for NSl-KDD, DoS and MC for DS2OS datasets and Theft for BoTIoT dataset. In terms of overall accuracy, the PrGWO-SK performed better for NSL-KDD at the same time length of feature vector was also reduced. For DS2OS, though the length of feature vector could not be reduced still notable increase in accuracy was witnessed. In case of BoTIoT the combination of accuracy and length of feature vector was effectively optimal. Although the present work performs better as mentioned above, it has shown limitations in detecting classes like U2R , R2L where the DR was not found to be better than some of the existing algorithms. Hence as a future work classification of classes like U2R and R2L can be pursued futher. Moreover, the authors intend to take up parameter tuning as further research objective in future.

## References

1. Alazzam H, Sharieh A, Sabri KE. A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. Expert systems with applications 2020; 148: 113249, https://doi.org/10.1016/j.eswa.2020.113249.
2. Almiani M, AbuGhazleh A, Al-Rahayfeh A, Atiewi S, Razaque A. Deep recurrent neural network for IoT intrusion detection system. Simulation Modelling Practice and Theory 2020; 101: 102031, https://doi.org/10.1016/j.simpat.2019.102031.
3. Antunes Rodrigues J, Torres Farinha J, Mendes M, Mateus R, Marques Cardoso A. Short and long forecast to implement predictive

maintenance in a pulp industry. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2022; 24(1): 33–41, http://doi.org/10.17531/ein.2022.1.5.

4. Baranowski J. Predicting IoT failures with Bayesian workflow. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2022; 24 (2): 248–259, http://doi.org/10.17531/ein.2022.2.6.

5. Emary E, Zawbaa HM, Hassanien AE. Binary grey wolf optimization approaches for feature selection. Neurocomputing 2016; 172: 371-381, https://doi.org/10.1016/j.neucom.2015.06.083.

6. Gao H, Qiu B, Barroso RJ, Hussain W, Xu Y, Wang X. TSMAE: a novel anomaly detection approach for internet of things time series data using memory-augmented autoencoder. IEEE Transactions on Network Science and Engineering(Early Access) 2022: 1, https://doi.org/10.1109/TNSE.2022.3163144.

7. Gao X, Shan C, Hu C, Niu Z, Liu Z. An adaptive ensemble machine learning model for intrusion detection. IEEE Access 2019; 7: 82512-82521, https://doi.org/10.1109/ACCESS.2019.2923640.

8. Golrang A, Golrang AM, Yildirim Yayilgan S, Elezaj O. A novel hybrid IDS based on modified NSGAII-ANN and random forest. Electronics 2020; 9(4): 577, https://doi.org/10.3390/electronics9040577.

9. Gu J, Lu S. An effective intrusion detection approach using SVM with naïve Bayes feature embedding. Computers & Security 2021; 103: 102158, https://doi.org/10.1016/j.cose.2020.102158.

10. Gu J, Wang L, Wang H, Wang S. A novel approach to intrusion detection using SVM ensemble with feature augmentation. Computers & Security 2019; 86: 53-62, https://doi.org/10.1016/j.cose.2019.05.022.

11. Hasan M, Islam MM, Zarif MI, Hashem MM. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet of Things 2019; 7: 100059, https://doi.org/10.1016/j.iot.2019.100059.

12. Hodo E, Bellekens X, Hamilton A, Dubouilh PL, Iorkyase E, Tachtatzis C, Atkinson R. Threat analysis of IoT networks using artificial neural network intrusion detection system. 2016 IEEE International Symposium on Networks, Computers and Communications (ISNCC), Hammamet, IEEE 2016: 1-6, https://doi.org/10.1109/ISNCC.2016.7746067.

13. Hoz ED, Hoz ED, Ortiz A, Ortega J, Martínez-Álvarez A. Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps. Knowledge Based Systems 2014; 71: 322-338, https://doi.org/10.1016/j.knosys.2014.08.013.

14. Hoz ED, Ortiz A, Ortega J, Hoz ED. Network anomaly classification by support vector classifiers ensemble and non-linear projection techniques. 2013 International Conference on Hybrid Artificial Intelligence Systems, Berlin, Springer 2013: 103-111, https://doi.org/10.1007/978-3-642-40846-5_11.

15. Ippoliti D, Zhou X. An adaptive growing hierarchical self organizing map for network intrusion detection. 2010 IEEE 19th International Conference on Computer Communications and Networks, Zurich, IEEE 2010: 1-7, https://doi.org/10.1109/ICCCN.2010.5560165.

16. Kayacik HG, Zincir-Heywood AN, Heywood MI. A hierarchical SOM-based intrusion detection system. Engineering applications of artificial intelligence 2007; 20(4): 439-51, https://doi.org/10.1016/j.engappai.2006.09.005.

17. Koroniotis N, Moustafa N, Sitnikova E, Turnbull B. Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-Iot dataset. Future Generation Computer Systems 2019; 100: 779-796, https://doi.org/10.1016/j.future.2019.05.041.

18. Kumar P, Gupta GP, Tripathi R. Toward design of an intelligent cyber attack detection system using hybrid feature reduced approach for iot networks. Arabian Journal for Science and Engineering 2021; 46(4): 3749-3778, https://doi.org/10.1007/s13369-020-05181-3.

19. Kunhare N, Tiwari R, Dhar J. Particle swarm optimization and feature selection for intrusion detection system. Sādhanā 2020; 45(1): 1-4, https://doi.org/10.1007/s12046-020-1308-5.

20. Mahfouz AM, Venugopal D, Shiva SG. Comparative analysis of ML classifiers for network intrusion detection. 2020 4th International Congress on Information and Communication Technology, Singapore, Springer 2020: 193-207, https://doi.org/10.1007/978-981-32-9343-4_16.

21. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Advances in Engineering Software 2014; 69: 46-61, https://doi.org/10.1016/j.advengsoft.2013.12.007.

22. Modi C, Patel D. A feasible approach to intrusion detection in virtual network layer of Cloud computing. Sādhanā 2018; 43(7): 1-6, https://doi.org/10.1007/s12046-018-0910-2.

23. Pahl MO, Aubet FX. DS2OS traffic traces . [https://www.kaggle.com/francoisxa/ds2ostraffictraces].

24. Pahl MO, Aubet FX, Liebald S. Graph-based IoT microservice security. 2018 IEEE/IFIP Network Operations and Management Symposium(NOMS), Taipei, IEEE 2018: 1-3, https://doi.org/10.1109/NOMS.2018.8406118.

25. Pajouh HH, Dastghaibyfard G, Hashemi S. Two-tier network anomaly detection model: a machine learning approach. Journal of Intelligent Information Systems 2017; 48(1): 61-74, https://doi.org/10.1007/s10844-015-0388-x.

26. Pajouh HH, Javidan R, Khayami R, Dehghantanha A, Choo KK. A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks. IEEE Transactions on Emerging Topics in Computing 2016; 7(2): 314-23, https://doi.org/10.1109/TETC.2016.2633228.

27. Rathore S, Park JH. Semi-supervised learning based distributed attack detection framework for IoT. Applied Soft Computing 2018; 72: 79-89, https://doi.org/10.1016/j.asoc.2018.05.049.

28. Safaldin M, Otair M, Abualigah L. Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing 2021; 12(2): 1559-76, https://doi.org/10.1007/s12652-020-02228-z.

29. Shafiq M, Tian Z, Sun Y, Du X, Guizani M. Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city. Future Generation Computer Systems 2020; 107: 433-42, https://doi.org/10.1016/j.future.2020.02.017.

30. Shams EA, Rizaner A, Ulusoy AH. Trust aware support vector machine intrusion detection and prevention system in vehicular adhoc networks. Computers & Security 2018; 78: 245-254, https://doi.org/10.1016/j.cose.2018.06.008

31. Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. IEEE transactions on Emerging topics in Computational Intelligence 2018; 2(1): 41-50, https://doi.org/10.1109/TETCI.2017.2772792.

32. Sivapalan G, Nundy KK, Dev S, Cardiff B, John D. ANNet: a lightweight neural network for ECG anomaly detection in IoT edge sensors. IEEE Transactions on Biomedical Circuits and Systems 2022; 16(1): 24-35, https://doi.org/10.1109/TBCAS.2021.3137646.

33. Soe YN, Feng Y, Santosa PI, Hartanto R, Sakurai K. Towards a lightweight detection system for cyber attacks in the IoT environment using corresponding features. Electronics 2020; 9(1): 144, https://doi.org/10.3390/electronics9010144.

34. Su T, Sun H, Zhu J, Wang S, Li Y. BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset. IEEE Access 2020; 8: 29575-29585, https://doi.org/10.1109/ACCESS.2020.2972627.

35. Tama BA, Comuzzi M, Rhee KH. TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. IEEE Access 2019; 7: 94497-94507, https://doi.org/10.1109/ACCESS.2019.2928048.

36. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the KDD CUP 99 data set. 2009 IEEE symposium on Computational Intelligence for Security and Defense Applications, Ottawa, IEEE 2009: 1-6, https://doi.org/10.1109/CISDA.2009.5356528.

37. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. The NSL-KDD data set. [https://www.unb.ca/cic/datasets/nsl.html].

38. Teng S, Wu N, Zhu H, Teng L, Zhang W. SVM-DT-based adaptive and collaborative intrusion detection. IEEE/CAA Journal of Automatica Sinica 2017; 5(1): 108-18, https://doi.org/10.1109/JAS.2017.7510730.

39. Tian Q, Han D, Li KC, Liu X, Duan L, Castiglione A. An intrusion detection approach based on improved deep belief network. Applied Intelligence 2020; 50(10): 3162-3178, https://doi.org/10.1007/s10489-020-01694-4.

40. Vijayanand R, Devaraj D, Kannapiran B. Intrusion detection system for wireless mesh network using multiple support vector machine classifiers with genetic-algorithm-based feature selection. Computers & Security 2018; 77: 304-314, https://doi.org/10.1016/j.cose.2018.04.010.

41. Wei W, Chen S, Lin Q, Ji J, Chen J. A multi-objective immune algorithm for intrusion feature selection. Applied Soft Computing 2020; 95: 106522, https://doi.org/10.1016/j.asoc.2020.106522.

42. Wisanwanichthan T, Thammawichai M. A Double-Layered Hybrid Approach for Network Intrusion Detection System Using Combined Naive Bayes and SVM. IEEE Access 2021; 9: 138432-138450, https://doi.org/10.1109/ACCESS.2021.3118573.

43. Wu K, Chen Z, Li W. A novel intrusion detection model for a massive network using convolutional neural networks. IEEE Access 2018; 6: 50850-50859, https://doi.org/10.1109/ACCESS.2018.2868993.

44. Yang Y, Zheng K, Wu C, Niu X, Yang Y. Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks. Applied Sciences 2019; 9(2): 238, https://doi.org/10.3390/app9020238.

45. Yao H, Wang Q, Wang L, Zhang P, Li M, Liu Y. An intrusion detection framework based on hybrid multi-level data mining. International Journal of Parallel Programming 2019; 47(4): 740-758, https://doi.org/10.1007/s10766-017-0537-7.

46. Yu Z, Tsai JJ, Weigert T. An adaptive automatically tuning intrusion detection system. ACM Transactions on Autonomous and Adaptive Systems 2008; 3(3): 1-25, https://doi.org/10.1145/1380422.1380425.

47. Zhang C, Ruan F, Yin L, Chen X, Zhai L, Liu F. A deep learning approach for network intrusion detection based on NSL-KDD dataset. 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), Xiamen, IEEE 2019: 41-45, https://doi.org/10.1109/ICASID.2019.8925239.
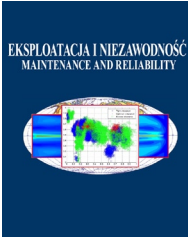
# Optimization synthesis of technological parameters during manufacturing of the parts

Indexed by:
Web of Science Group

Yaroslav Kusyi[a], Vadym Stupnytskyy[a], Oleh Onysko[b], Egidijus Dragašius[c], Saulius Baskutis[c], Rafal Chatys[d,*]

[a] Lviv Polytechnic National University, Department of Robotics and Integrated Mechanical Engineering Technologies, Bandera st. 12, 79013 Lviv, Ukraine
[b] Ivano-Frankivsk National Technical University of Oil and Gas, Department of Computerized Mechanical Engineering, Karpatska st. 15, 76000 Ivano-Frankivsk, Ukraine
[c] Kaunas University of Technology, Faculty of Mechanical Engineering and Design, Department of Production Engineering, Studentu st. 56, 51424 Kaunas, Lithuania
[d] Kielce University of Technology, Faculty of Mechatronics and Mechanical Engineering, Department Mechatronics and Armament Engineering, al. Tysiąclecia Panstwa Polskiego 7, 25-314 Kielce, Poland

## Highlights

- The developed synthesis method for optimizing the parameters of manufacturing processes made it possible to improve manufacturing efficiency, processing quality, extend the service life of cutting tools, reduce the consumption of energy, and all this created added value for the production process.

- The mathematical model of the technical system is written using the system of Chapman-Kolmogorov differential equations.

- Markov chains were used to study the possible states of objects and to analyse their transition to other states.

- Proposed optimization technique during the manufacturing of parts has been successfully integrated and tested in the real industry.

## Abstract

Technological ensuring the reliability of machine parts is realized by failing to reach the limited state of the elements of the technological system: machine – clamping device – metal-cutting tool-part. A method of optimization synthesis of parameters of technological processes of manufacturing machine parts has been developed. Testing the developed methodology, it was found that the metal cutting tool is Meanwhile, research has shown that metal cutting machine has the least influence on the formation of detailed quality-adjustable parameters from all the the weakest element of the technological system in terms of reliability and has the greatest impact on the quality of machined parts. elements of the process media "machine – clamping device – cutting tool". Finally, a concrete example is provided to demonstrate the effectiveness of the proposed method. The proposed technique has been successfully tested for the manufacturing process of the reduction-gear housing.

## Keywords

Reliability Engineering, FMEA, FMECA, technological damage, technological graph of reliability, Markov chain.

## 1. Introduction

Currently, operational safety is the most important characteristic of modern technical systems [19, 3, 48], which involves the analysis of failures and their consequences [21, 4, 30, 20]. The concept of risk is interpreted as a probabilistic criterion for hazards of a particular type, or the number of possible losses (damages) caused by an adverse event, or a combination of these values [4, 29].

At the same time, the development and implementation of the integrated information systems management for technological processes of manufacturing products in the practice of mechanical engineering enterprises is the main mover of economic growth of industrialized countries in the world [18, 23, 32, 55]. These systems give a competitive advantage to manufacturing companies in a global business environment.

The changing priorities of modern mechanical engineering require a thorough study of complex engineering problems and a mathematical apparatus to solve them [14]. In addition, systematic theoretical and experimental investigations are necessary. The appearance of a new fundamental problem – protecting facilities to prevent their failures, accidents, catastrophes – determines new parameters, criteria, risks, and regulatory and technical documents, rethinking the attitude toward providing the products' reliability indicators and operational characteristics [21, 29, 10].

Investigation of the regulated initial quality parameters of machines parts during their manufacturing to provide their operational

characteristics and indicators of reliability is an essential task for the formation of available products [21, 18, 28, 38]. In order to guarantee the reliability of the final produced products, the cornerstone is the relationship between the reliability of the technological system and the reliability of the manufactured parts.

Technological ensuring the reliability of machine parts is realized by failing to reach the limited state of the elements of the technological system: machine – clamping device – metal cutting tool – part. A method of optimization synthesis of parameters of technological processes of manufacturing machine parts has been developed. To evaluate the presented method, a series of modelling experiments were performed, during which various machining modes were tested and their influence on the cutting edge of the cutting tool during machining of the part was determined. A case study was used to illustrate how to apply the method in real manufacturing conditions. The proposed technique has been successfully tested for the manufacturing process of the gearbox housing.

## 2. Literature review

A typical approach to international theory and practice on system reliability and risk minimization [15] combines constructive and analytical (quantitative and qualitative) methods [4, 12, 36]. These methods do not replace each other but interact with each other. Potential failures are identified, and measures are taken to eliminate them to reduce their impact or probability of occurrence using the FMEA (Failure Mode and Effects Analysis) and FMECA (Failure Mode Effects and Criticality Analysis) techniques [5, 8, 11]. Reliability parameters are calculated using quantitative methods, methods using the criterion of achieving the product's limit state or its elements [4, 15].
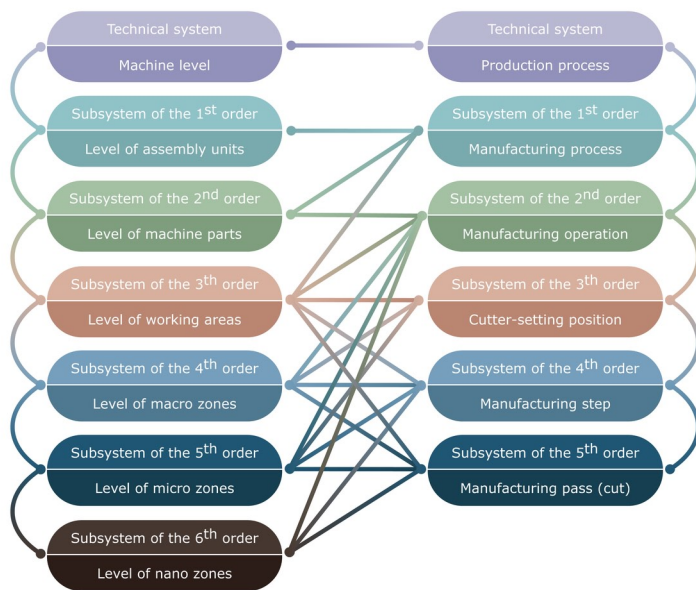


*Fig.1. Formation of subsystems of different hierarchical levels according to the theory of technical systems for function-oriented design of technological processes*

According to the standard approach to providing the reliability of technical systems [4] to calculate the indicators of reliability and durability, it is necessary to establish the regularities of distribution for the element's failures. The reliability of the technical system is determined based on analysis and study of the established frequencies of distribution for the failures of technical system elements [21, 4]. Mathematical operations for the obtained distribution regularities provide system reliability indicators within the framework of the system

(structural, mathematical) theory of reliability [29, 44]. The priority task of the system theory of reliability is the transition from indicators of elements reliability to indicators of system reliability, which complicates the practical solution of modern mechanical engineering problems [56, 43].

The difference in the calculations of the real reliability of the technical system using a system approach consists in the analysis of the reliability of elements and techniques in the general multilevel simulation process using physical and structural models [4, 52]. Physical models of element reliability are used at the lower levels, and structural models – at higher ones [44, 25]. This approach leads to the analysis of reliability indicators in relationship with safety and risk, providing the quality of the technical system. At the same time, ALARA (as low as risk acceptable), currently adopted in most countries, recognizes a particular economic character of providing the reliability and safety of technical systems with minimization of the degree of risk [53, 17].

A hybrid modelling framework for production systems combine machine and system-level models of separate and continuous dynamics to examine the effects of different control variables on production efficiency, reliability, quality, and power consumption [39].

The theory of reliability of technical systems is widely used in the calculations of electrical and electronic circuits. Such circuits have standard components with known reliability characteristics [51, 6].

In the general case, according to the principles of functionally oriented design, a product as part of a technical system is developed as an engineering pyramid from a set of subsystems of different hierarchical levels: the machine level (the final element of satisfying social needs); the level of components that make up the machine - assembly units (subsystem of the 1st order); level of machine parts (subsystem of the 2nd order); the level of the working areas of the part (subsystem of the 3th order); the level of macrozones in the working areas (macroscopic level of research; subsystem of the 4th order); the level of microzones (microscopic level of research; subsystem of the 5th order) and the level of nanozones (submicroscopic/nanoscopic level of research; subsystem of the 6th order) (Fig.1).

Any subsystem is a technical system in relation to a subsystem of the lower level, corresponding to the general principles of systems theory [21, 23].

Analytical or simulation models of algorithms for the functional behaviour of various machines and mechanisms for determining performance or reliability indicators are developed at the stage of their system-engineering design [7, 33, 34]. Complex analytical models are designed (developed) to analyze the functional characteristics and reliability indicators of the research object [41, 47, 50].

The technology of analytical modelling of machines, mechanisms and technological systems makes it possible to use Markov chains to study the possible states of objects and analyse their transition to other states. Such a model of the object of study is represented in the form of a graph of states and transitions, which is actually an intermediary model. The graph of states and transitions is described by the system of Chapman-Kolmogorov differential equations [7, 24, 9, 35].

At the same time, for example, when designing parts and systems of the radio-electronic industry, it is assumed that a specific technical system fails with a certain intensity. The duration of stay in good condition is described by an exponential law with the parameter $\lambda(t)$ = const, where $\lambda(t)$ is the failure rate [6, 7].

The failure rate of basic events in this case:

$$\lambda_i = \frac{1}{T_i} \qquad (1)$$

where: $T_i$ is the average value of the duration of the $i$-th operational block of behavior equivalent to the algorithm.

The technical system is restored after a failure, and the duration of the restoration is distributed according to the Erlang distribution law [6, 7].

The intensity of restoration of the technical system is described by the formula [7]:

$$\mu_{er}(t) = \frac{\alpha \cdot (\alpha \cdot t)^{n-1}}{(n-1)! \sum\limits_{k=0}^{n-1} \frac{(\alpha \cdot t)^k}{k!}} \qquad (2)$$

where: $n$ is the order of the distribution law (shape parameter), $\alpha$ is the scale parameter.

Two events are considered in the general case for a technical system: the "technical system failure" event and the "technical system restoration" event. The graph of states and transitions of the technical system is shown in Fig. 2. The states of such a technical system are characterized by the following probabilities: $P_0(t)$ – the probability that the technical system is ready for work (capable of working) in state $S_0$; $P_1(t)$ is the probability of the technical system being in a faulty state (being under repair or failure) $S_1$ [6, 7].
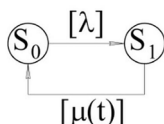


*Fig. 2. Graph of states and transitions of a technical system*

The mathematical model of such a technical system is presented as a system of differential equations with variable coefficients [7]:

$$\begin{aligned}
\frac{dP_0(t)}{dt} &= -\lambda \cdot P_0(t) + \mu_{er}(t) \cdot P_1(t), \\
\frac{dP_1(t)}{dt} &= \lambda \cdot P_0(t) - \mu_{er}(t) \cdot P_1(t).
\end{aligned} \qquad (3)$$

The state $S_1$ is replaced by a chain to bring the mathematical model (3) to a homogeneous Markov model, according to the Erlang's method of stages. The number of states in the new chain is determined by the order of the chosen Erlang distribution law. The equivalent graph of states and transitions for the chain ($n = 2$) is shown in Fig. 3 for the transformation of the technical system shown in Fig. 1. On Fig. 3 are denoted by $P_{10}$, $P_{11}$, that is, the probability of the technical system staying in fictitious states $S_{10}$ and $S_{11}$, describing the recovery procedure [5, 6, 43].
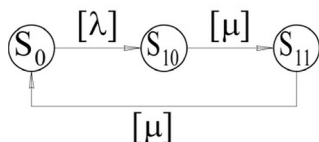


*Fig. 3. Graph of states and transitions of a technical system*

The mathematical model of a technical system is written using a system of Chapman-Kolmogorov differential equations with constant coefficients after the formation of a graph of states and transitions [7]:

$$\begin{aligned}
\frac{dP_0(t)}{dt} &= -\lambda \cdot P_0(t) + \mu \cdot P_1(t), \\
\frac{dP_{10}(t)}{dt} &= \lambda \cdot P_0(t) - \mu \cdot P_{10}(t), \\
\frac{dP_{11}(t)}{dt} &= -\mu \cdot P_0(t) + \mu \cdot P_{11}(t).
\end{aligned} \qquad (4)$$

The method based on the implementation of the Laplace transformation is used to solve system (4) with a conditionally small number of equations [7]:

$$\begin{aligned}
s \cdot p_0(s) - C_0(0) &= -\lambda \cdot p_0(s) + \mu \cdot p_{11}(s), \\
s \cdot p_{10}(s) - C_1(0) &= \lambda \cdot p_0(s) - \mu \cdot p_{10}(s), \\
s \cdot p_{11}(s) - C_2(0) &= \mu \cdot p_{10}(s) - \mu \cdot p_{11}(s).
\end{aligned} \qquad (5)$$

The values of the constants take the form $C_0(t=0)=1$, $C_1(t=0)=0$ and $C_2(t=0)=0$ for a capable of working technical system at the initial moment of time. In this regard, the solution of the system of linear equations (5) takes the following form:

$$\begin{aligned}
p_0(s) &= \frac{(\mu+s)^2}{s^3 + (2 \cdot \mu + \lambda) \cdot s^2 + (\mu + 2 \cdot \lambda) \cdot \mu \cdot s}, \\
p_{10}(s) &= \frac{\lambda \cdot (\mu+s)}{s^3 + (2 \cdot \mu + \lambda) \cdot s^2 + (\mu + 2 \cdot \lambda) \cdot \mu \cdot s}, \\
p_{11}(s) &= \frac{\lambda \cdot \mu}{s^3 + (2 \cdot \mu + \lambda) \cdot s^2 + (\mu + 2 \cdot \lambda) \cdot \mu \cdot s}.
\end{aligned} \qquad (6)$$

However, in the analysis of mechanical systems, the indicators of reliability and durability change due to a large number of possible design options and manufacturing technology, different operating conditions of a particular element in various machines and mechanisms. This complicates their prediction of operating conditions [25].

According to the principle of resource-dependent behaviour, elements of a real mechanical system are interdependent. Their failures or limit states are related to each other and are determined by the influence of common factors [15]. Modern approaches to the calculation of technical system reliability indicators are based on applying the concept of the limit state of the system element [4, 25]. A review of the literature has shown that due to the lack of maintenance of the functional characteristics of production systems, the reliability of the technological system is problematic to evaluate exactly. For this purpose, a method of optimization synthesis of parameters of technological processes was developed to fully describe the state of the technological system, taking in to account the machining modes, physical-mechanical characteristics of the processed material, and the desired machining accuracy and surface quality of the part.

## 3. Research methodology

The output parameters are formed as a result of sequential processing of the workpiece, which is analysed by the technological inheritability of the part properties. As a rule, regulated indicators are set for each operation. Each operation, as a rule, has its own regulated indicators. They are provided due to technological transitions in the processing of parts [23, 37].

The probability of failure to reach the limit state of the workpiece part $P_{Xij}(r;k,....t)$ for the $j$-th parameter in the $i$-th technological operation of the technological process of manufacturing, taking into account the theorem of adding incompatible events, is determined by the formula:

$$P_{X_{ij}}\left(r,k,.....t\right)=\prod_{j=1}^{q_V}\left(1-F_{X_{ij}}\left(r,g,.....t\right)\right) \qquad (7)$$

where: $X_{ij}$ $(j\in[1;q_i])$ is defined the $j$-th output parameter $X$ for the $i$-th technological operation; $F_{X_{ij}}$ $(r;k,....t)$ is the formation of the limit state of the workpiece by the $j$-th parameter in the $i$-th technological operation of the technological process during production.

Output parameters after the last technological operation: $X_{nj}=X_j(j\in[1;m])$, where: $m$ – the total number of output parameters. The output parameters $q_i$ for the $i$-th technological operation form a set of input parameters for $(i+1)$ operation. However, only some of them will be included in the set of output parameters of the final product according to the synergistic approach: $q_{0i}<q_i$ [13].

The output parameters of mechanical engineering parts are classified into three groups [21].

The output parameters of group I for mechanical engineering parts include a few of the output parameters of intermediate operations (for example, physical-mechanical characteristics of parts material). The finishing and strengthening operations of parts manufacturing provide a lot of output parameters according to accuracy, quality of surfaces, operational characteristics and reliability indicators of mechanical engineering objects (the output parameters of group II for mechanical engineering parts). The output parameters of group III for mechanical engineering objects are functionally related to previous intermediate operations parameters according to technological inheritability and technological inheritance during parts manufacturing [22].

Then the total number of output parameters is:

$$m=q_n+\sum_{i=0}^{n-1}q_{oi} \qquad (8)$$

where: $q_n$ is the number of output parameters obtained in the last operation and functionally related with the technological inheritability and technological inheritance during parts manufacturing (the output parameters of group II and III); $q_{oi}$ is the number of output parameters for the $i$-th technological operation which form the output parameters of group I.

The probability of the working state of the part $P(t)$ in the general case for the technological process, independently forming each initial parameter, is determined as follows:

$$P\left(r,k,.....t\right)=\prod_{i=1}^{n}P_i\left(t\right)=\prod_{i=0}^{n}\prod_{j=1}^{m}P_{X_{ij}}\left(r,g,.....t\right) \qquad (9)$$

where: $P_{X_{ij}}$ $(r,g,....t)$ is the probability of failure to reach the limit state of the part by the $j$-th parameter in $i$-th technological operation of the technological process of its manufacture.

Analysis of the probability of providing regulated quality parameters of the workpiece is appropriate, taking in to account the influence of elements of the technological system in the $i$-th technological operation, using the mathematical apparatus for implementation of Markov processes [26, 27]. The technological graph of reliability for the $i$-th technological operation, described by Markov chains, is shown in Fig. 4.

Vertices of the technological graph of reliability in the $i$-th technological operation (Fig. 4) describe the possibility of the technical system in $(k+3)$ possible states taking into account the influence of the technological environment elements: 1 are aspects of the metal-working technological system: metal-cutting machine (MCM), clamping device located on the metal-cutting machine (MCMD), metal cutting tool (MCT) during the processing of the workpiece in the $i$-th technological operation is not in the limit state that provides the formation of regulated quality parameters of the part according to with the require-
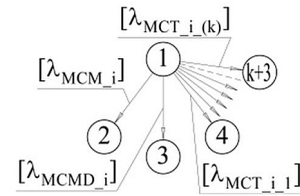


*Fig. 4 Technological graph of the reliability for the i-th technological operation that is described by Markov chains to predict the probability of providing regulated quality parameters of the workpiece, taking in to account the influence of the technological system elements*

ments of design and technological documentation; 2, 3, 4,…, $(k+3)$ are metal-cutting machine, clamping device on the metal-cutting machine, the first / $k$-th metal-cutting tool in general or their elements, in particular, are in the limit state in the metal-working technological system during processing of the workpiece in the $i$-th technological operation, respectively. In Fig. 2 $\lambda_{MCM\_i}$, $\lambda_{CD\_i}$, $\lambda_{MCT\_i\_1}$, …., $\lambda_{MCT\_i\_(k)}$ are failure rates of the metal-cutting machine, the clamping device on the metal-cutting machine and the first / $k$-th metal-cutting tool or their elements in the $i$-th technological operation during workpiece processing.

According to Fig. 4, notation of probabilities of events is introduced: $P_1(t)$ is the probability of the technical system in state 1 that characterizes the ensuring of regulated parameters of the workpiece during its processing in the $i$-th technological operation in case of not reaching the limit states of the metal-cutting machine, clamping device and tool (or tools) in general and their elements in particular; $P_2(t)$, $P_3(t)$, $P_4(t)$, …., $P_{(k+3)}(t)$ is the probability of the technical system in states 2, 3, 4, ….., $(k+3)$, characterizes the failure to provide the regulated parameters of the workpiece during its processing in the $i$-th technological operation, provided that the metal-cutting machine reaches the limit state, a clamping device, the first / $k$-th metal-cutting tool in general and (or) their elements in particular, respectively.

The system of Chapman-Kolmogorov differential equations for the technological graph of reliability (Fig. 4), taking in to account the influence of elements of the metal-working technological system on providing the regulated quality parameters of machine product for the $i$-th technological operation, looks like:

$$\frac{dP_1(t)}{dt}=-\lambda_{MCM\_V}\cdot P_1(t)-\lambda_{MCMD\_V}\cdot P_1(t)-$$
$$\sum_{k=1}^{z}\lambda_{MCT\_(k)i}\cdot P_1(t),$$
$$\frac{dP_2(t)}{dt}=\lambda_{MCM\_V}\cdot P_1(t),$$
$$\frac{dP_3(t)}{dt}=\lambda_{MCMD\_V}\cdot P_1(t),$$
$$\frac{dP_4(t)}{dt}=\lambda_{MCT\_i\_1}\cdot P_1(t),$$
$$.............................$$
$$\frac{dP_{z+3}(t)}{dt}=\lambda_{MCT\_i\_z}\cdot P_1(t),$$
$$P_1(t)+P_2(t)+P_3(t)+P_4(t)+..........P_{z+3}(t)=1,$$
$$t\in[0;t], \qquad (10)$$

where: $k$ is the $k$-th metal-cutting tool in the $i$-th technological operation ($i=1,… z$).

The last equation in the system (10) represents the addition theorem of incompatible events. This equation is presented in the form from the viewpoint of the physical essence of prediction of the regulated workpiece parameters during its processing in the $i$-th technological

operation, taking in to account the influence of elements of the metal-working technological system:

$$P_{W(P)i}(r,g,....t) + F_{MCM\,i}(r,g,....t) +$$
$$+F_{MCMD\,i}(r,g,....t) + \sum_{k=1}^{z} F_{MCT\,(k)i}(r,g,....t) = 1 \qquad (11)$$

where: $P_{W(P)i}(r, g, ....t)$ is the probability of providing regulated quality parameters of the workpiece in the $i$-th technological operation of the technological process of the product manufacturing in case of failure to reach the limit state of the metal-cutting machine, clamping device located on the metal-cutting machine and metal-cutting tool (or tools) in general and their elements in particular; $F_{MCMi}(r, g, ....t)$ is the probability of failure to provide regulated quality parameters in the $i$-th technological operation of the technological process of product manufacturing, when the limit state is reached by the metal-cutting machine in general or its elements in particular; $F_{MCMDi}(r, g, ....t)$ is the probability of failure to provide the regulated quality parameters in the $i$-th technological operation of the technological process of product manufacturing, when the limit state is reached by the clamping device on a metal-cutting machine in general or its elements in particular; $F_{MCT\,i}(r, g, ....t)$ is the probability of failure to provide regulated quality parameters in the $i$-th technological operation of the technological process of product manufacturing when the limit state is reached by the metal-cutting tool (or tools) in general or its (their) elements in particular; $k$ is the $k$-th metal cutting tool in the $i$-the technological operation, $[k \in 1;z]$.

The probability of providing the regulated quality parameters of the workpiece in the $i$-th technological operation of the technological process of product manufacturing in case of not reaching the limit state of the metal-cutting machine, clamping device on the metal-cutting machine and metal-cutting tool (or tools) in general and their elements in particular, is determined by (12):

$$P_{W(P)i}(r,g,....t) = 1 - F_{MCM\,i}(r,g,....t) -$$
$$F_{MCMDi}(r,g,....t) - \sum_{k=1}^{z} F_{MCT\,(k)i}(r,g,....t) \qquad (12)$$

The probability of providing the regulated quality parameters of the workpiece in the $i$-th technological operation of the technological process of manufacturing the product is determined by the possibilities of reaching the limit state by the metal-cutting machine $F_{MCMi}(r, g, ....t)$, the clamping device on metal-cutting machine $F_{MCMDi}(r, g, ....t)$ and metal-cutting tool (or tools) $F_{MCTi}(r, g, ....t)$ in general and their elements in particular.

In addition, the elements of the metal-working technological system have specific reliability parameters according to the technical requirements for their manufacturing and operation that provide regulated quality parameters of machine parts during their formation, control and assembly.

In this case, the conditions for providing serviceability to technological problems are determined by:

$$P_{W(P)i}(r,g,....t) \geq \left[P_{W(P)}\right];$$
$$P_{MCM\,i}(r,g,....t) \geq \left[PMCM\right];$$
$$P_{MCMD\,i}(r,g,....t) \geq \left[P_{MCMD}\right]; \qquad (13)$$
$$P_{MCT(k)i}(r,g,....t) \geq \left[PMCT\right],$$

where: $P_{W(P)i}(r, g, ....t)$, $P_{MCMi}(r, g, ....t)$, $P_{MCMDi}(r, g, ....t)$, $P_{MCT(k)i}(r, g, ....t)$ are the probability of providing the regulated quality parameters in the $i$-th technological operation of technological process of the

manufacturing, when the limit state is not reached by the metal-cutting machine, a clamping device on the metal-cutting machine and $k$-th metal-cutting tool, respectively, in general or its (their) elements in particular; $[P_{W(P)}]$, $[P_{MCM}]$, $[P_{MCMD}]$, $[P_{MCT}]$ is regulated probability of providing a certain indicator of reliability for the part, metal-cutting machine, clamping device and metal-cutting tool.

In this case, the number of technological steps within the content and the number of technological operations within the technological process of the part manufacturing is optimized; that is, the number of subsystems at different levels of the technical system is optimized. This allows us in the self-organized systems to control their adaptability and reliability of the conditions of formation and fluctuations by changing the number of subsystems [13, 54]. Entropy increases in closed independent subsystems [21, 13]. To provide this condition, the distribution of probabilities and the introduction of isolated reserves are used in disconnected open subsystems that exchange resources and information with the external environment.

## 4. The study of the failure rate for the elements of the technological system "machine – clamping device - tool – part"

### 4.1. General provisions

The failure rate is a priority indicator of the reliability of non-repaired and non-restorated facilities. The failure rate is the relative density of the probability of an object failure, which is determined under the condition that a failure does not occur by the considered point in time [16]. In the general case [42]:

$$\lambda(t) = \frac{f(t)}{P(t)}, \qquad (14)$$

where: $f(t)$ is the distribution density of uptime, $P(t)$ is the probability of non-failure operation of an object of a technical system, subsystem, etc.

### 4.2. Determination of the failure rate for the main technological equipment (metal-cutting machines) in the system "machine – clamping device - tool - part" in the manufacture of parts

The operation of metal-cutting equipment implies that the system of maintenance and repair of the machine is implemented clearly in accordance with its repair cycle. This minimizes the impact of wear of components and parts of metal-cutting machines. In this regard, the exponential distribution law of the parameters of metal-cutting machines is used to conduct research in the process of machining parts, including reduction-gear housing as example.

The restoration of the technical parameters of the main technological equipment lost during operation is carried out in accordance with its repair cycle. The duration of the repair cycle directly depends on the degree of accuracy of the machine, its operating conditions, cutting modes during the machining of parts, the type of production and the service life of rapidly wearing machine parts.

According to the production experience of the Lviv plant of milling machines, the average duration of the repair cycle is 30,000 working hours. At the same time, the repair cycle is about 90 months, the overhaul period is 9 months, and the period between inspections is 5 months for the work in two shifts.

The metal-cutting machine is the strongest element in the process media "machine – clamping device – tool". Since the overhaul period is set in the repair cycle of a metal-cutting machine in accordance with the need to restore its technical characteristics lost during operation, the failure rate of metal-cutting machine with an exponential law of reliability can be defined as follows:

$$\lambda_{MCM} = \frac{1}{T_{IT\_MCM}}, \qquad (15)$$

where: $T_{IT\_MCM}$ is the average overhaul period in the repair cycle of metal-cutting machine, min.

$$T_{IT\_MCM} = \frac{F_{r.} \cdot \left[ T_{IT\_MCM} \right]}{12} \cdot 60, \qquad (16)$$

where: $[T_{IT\_MCM}]$ – the average overhaul period in the repair cycle of metal-cutting equipment in months, $F_{r.}$ – the actual annual fund of the equipment operation time in hours; for two-shift operation $F_r$=4055 h [21, 46]. From production experience, we take $[T_{IT\_MCM}]$ = 9 months. Then:

$$T_{IT\_MCM} = \frac{4055 \cdot 9}{12} \cdot 60 = 182475 \text{ min.}$$

Based on production experience, the average overhaul period in the repair cycle of metal-cutting equipment is assumed to be the same for universal metal-cutting machines, i.e. $T_{IT\_MCM}$=182475 min.

$$\lambda_{MCM} = \frac{1}{182475} = 5.48 \cdot 10^{-6} \text{ 1/min.}$$

Hence $\lambda_{MCM}$=5.48·10⁻⁶ 1/min.

### 4.3. Calculation of the failure rate of clamping device in the system "machine – clamping device - tool - part" in the manufacture of parts

As stated earlier, during the operation of metal-cutting equipment, we accept that the system of maintenance and repair of technological equipment is implemented clearly in accordance with its repair cycle. The failure rate of clamping device is determined by the exponential law of reliability according to the formula [21]:

$$\lambda_{MCMD} = \frac{1}{T_{MCMD}}, \qquad (17)$$

where: $T_{MCMD}$ – established trouble-free time between failures of the clamping device.

Based on production experience for mechanized clamping devices on universal machines, it can be stated that $T_{MCMD}$ =1500 h. Then:

$$\lambda_{MCMD} = \frac{1}{1500 \cdot 60} = 1.111 \cdot 10^{-5} \text{ 1/min.}$$

Consequently, the failure rate of clamping devices on universal metal-cutting machines $\lambda_{MCMD}$= 1.111·10⁻⁵ 1/min.

### 4.4. Determination of the failure rate of a metal-cutting tool in the system "machine – clamping device - tool - part" in the manufacture of parts

The metal-cutting tool is the weakest element of the process media "machine – clamping device - tool" in terms of reliability [40]. The exponential distribution is used among the distribution laws of random variables to assess the reliability parameters of a metal-cutting tool, which is justified by the following criteria [21, 40]:

1) The probability of failure-free operation is always underestimated with an exponential distribution in comparison with other statistical laws, which determines more stringent requirements for the reliability parameters of a metal-cutting tool;

2) The actual duration of the metal-cutting tool is much less than its stability when analyzing reliability indicators for a certain manufacturing step (operation) when processing with this tool. This makes it possible not to take in to account its wear, and the failure rate over a short period of time can be assumed to be a constant value.

The rejection of the worst-quality samples of metal-cutting tools occurs during their operation during the machining of workpieces. Therefore, the failure rate remains practically unchanged for a batch of metal-cutting tools [2].

The failure rate is assumed to be equal to the destruction rate of the metal-cutting tool $\lambda_{MCT\_0}$ in the calculations of the reliability indicators of the elements of the technological system "machine – clamping device – tool – part" for the exponential distribution [40]:

$$\lambda_{MCT} = \lambda_{MCT\_0} = \lambda_{MCT\_bas.} \cdot k_{f.}, \qquad (18)$$

where: $\lambda_{MCT\_bas.}$ – basic failure rate of the tool according to exponential reliability law, 1/min, $k_f$ – dimensionless correction factor that takes in to account the proportion of sudden failures for a given type of tool.

The basic failure rate of the tool is determined by the formula [40]:

$$\lambda_{MCT\_bas.} = \frac{1}{T}, \qquad (19)$$

where: $T$ – cutting tool life, min.

After substituting (19) into (18), we get:

$$\lambda_{MCT} = \lambda_{MCT\_0} = \frac{k_{f.}}{T}. \qquad (20)$$

Based on the recommended values of durability and production experience, the failure rate $\lambda_{MCT}$ is calculated for metal-cutting tools that are used for processing the gearbox housing, in particular [21]: for face mills – 2.5·10⁻³-3·10⁻³; for lathe tools – 2.7·10⁻²; for drills – 2.22·10⁻²-6.67·10⁻²; for core drills and countersinks – 3.33·10⁻²; for reamers – 2.22·10⁻²-3.33·10⁻²; for taps – 1.11·10⁻².

## 5. Mathematical modeling of technological support of quality parameters in the manufacture of parts during their machining operations

Optimization synthesis involves the following problems [21, 49, 1]: substantiation of specific indicators of reliability of machine parts relative to the regulated indicators of machine reliability; establishment of the limit value of the machining time at a particular technological step (technological operation) of mechanical processing of the part according to specific reliability indicators; determination of optimal modes of part processing according to the limit value of the machining time by the calculated reliability indicators.

If condition (13) is not met, the parameters of the technological operation, in particular the elements of the cutting modes that determine the central machining time or (and) the resistance of the tool at the limiting technological step, are optimized.

Determination of the regulated reliability indicators of products concerning the regulated indicators of reliability of the machine is not the problem of this paper. The value of the gamma-percentile operating time to failure $P(t_\gamma)$ = 1.0; 0.99; 0.95; 0.90; 0.80; 0.50 is regulated

by standards depending on the responsibility of the element and its cost [48].

The provision of technological reliability parameters, taking in to account the influence of a specialised system of metal-processing, is analysed in technological operations and steps during the manufacture of mechanical engineering part–a reduction-gear housing (enterprise "Agromashproekt", Lviv, Ukraine).

Casting of the cast iron 30B (USA Standard ASTM A 48) is used for a workpiece of the reduction-gear housing. This workpiece's technological route of machining consists of 10 technological machining operations. Metal-cutting and measuring tools, and technological equipment correspond to this type of production.

Technological operation 005 (horizontal milling) will be analysed in detail (Fig. 5, Table 1).

The reduction-gear housing represents one of the most difficult-to-cut classes of the mechanical engineering parts – "Cases". It serves as a datum's part for the location of shafts with gears, bearings, and other gearbox elements.
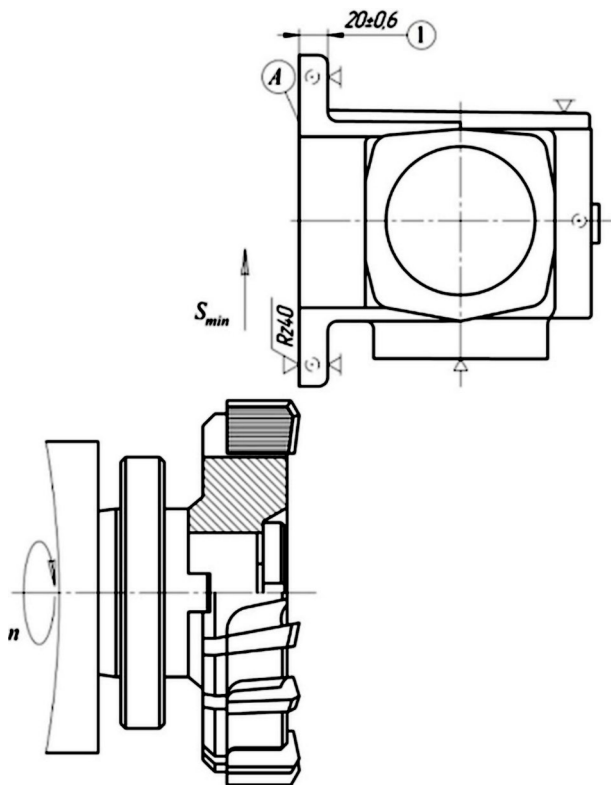


*Fig. 5. Sketch map of the operation 005 (horizontal-milling) for machining of reduction-gear housing*

A system of the Chapman-Kolmogorov differential equations was developed for operation 005 of the technological process of manufacturing the reduction-gear housing for the technological graph of reliability (Fig. 6):

$$\frac{dP_1(t)}{dt} = -\lambda_{MCM\_005} \cdot P_1(t) - \lambda_{\square CMD\_005} \cdot P_1(t)$$
$$-\lambda_{\square CT\_005} \cdot P_1(t),$$
$$\frac{dP_2(t)}{dt} = \lambda_{MCM\_005} \cdot P_1(t),$$
$$\frac{dP_3(t)}{dt} = \lambda_{MCMD\_005} \cdot P_1(t),$$
$$\frac{dP_4(t)}{dt} = \lambda_{MCT\_005} \cdot P_1(t),$$
$$P_1(t) + P_2(t) + P_3(t) + P_4(t) = 1,$$
$$t \in [0;t].$$

(21)

The partial solution of the system of differential equations (21), which is necessary for the optimization synthesis, as:

$$P_1(t) = e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t} + C_1.$$

(22)

If $P_1$ (t = 0) = 1, $C_1$ (t = 0) = 0.
At that time $P_1(t)$ as element of equation (12) is equal:

$$P_1(t) = e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t}$$
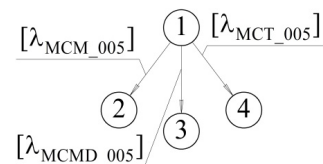
(23)



*Fig. 6. Technological graph of reliability for operation 005 for machining of reduction-gear housing*

After the substitution $\lambda_{MCM\_005}$=5.48·10⁻⁶ min⁻¹; $\lambda_{MCMD\_005}$=1.111·10⁻⁵ min⁻¹; $\lambda_{MCT\_005}$=2.5·10⁻³ min⁻¹ (see 4.2-4.4) in expression (22), is obtained $P_1(t) = e^{-2.517 \cdot 10^{-3} \cdot t}$.

$P_2(t), P_3(t), P_4(t)$ are calculated after solving the system of differential equations (21):

$$P_2(t) = \frac{\lambda_{MCM\_005} \cdot e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t}}{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right)} +$$
$$+\lambda_{MCM\_005} \cdot C_1 \cdot t + C_2;$$

$$P_3(t) = \frac{\lambda_{MCMD\_005} \cdot e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t}}{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right)} +$$
$$+\lambda_{MCMD\_005} \cdot C_1 \cdot t + C_3;$$

*Table 1. Technological operation of machining a workpiece "Reduction-gear housing"*

| Number of the operation | The name of the operation and the description of steps | Elements of the metal-treatment technological systems | | | Machining time, min | |
|---|---|---|---|---|---|---|
| | | MCM | MCT | MCMD | in steps | in operation |
| 005 | Horizontal-milling. To mill a reference plane A, providing the specified dimension 1 (Fig. 2) | Universal milling machine | Face milling cutter | Clamping device | 1.48 | 1.48 |

$$P_4(t) = \frac{\lambda_{MCT\_005} \cdot e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t}}{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right)} +$$
$$+ \lambda_{MCT\_005} \cdot C_1 \cdot t + C_4.$$

After substituting $C_1$ (t = 0) = 0, formulas in symbolic form can be written as follows:

$$C_2 = \frac{\lambda_{MCM\_005}}{\left(\lambda_{MCM\_005} + \lambda_{MCMB\_005} + \lambda_{MCT\_005}\right)};$$

$$C_3 = \frac{\lambda_{MCMD\_005} \cdot 1}{\left(\lambda_{MCM\_005} + \lambda_{MCM\square\_005} + \lambda_{MCT\_005}\right)};$$

$$C_4 = \frac{\lambda_{MCT\_005}}{\left(\lambda_{MCM\_005} + \lambda_{MCM\square\_005} + \lambda_{MCT\_005}\right)}.$$

Quantitative values of constants: $C_2 = 2.178 \cdot 10^{-3}$; $C_3 = 4.415 \cdot 10^{-3}$; $C_4 = 0.993$.

The components of equation (12) are presented as:

$$F_2(t) = \frac{\lambda_{MCM\_005} \cdot e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t}}{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right)} + 2.178 \cdot 10^{-3};$$

$$F_3(t) = \frac{\lambda_{MCMD\_005} \cdot e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t}}{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right)} + 4.415 \cdot 10^{-3};$$

(24)

$$F_4(t) = \frac{\lambda_{MCT\_005} \cdot e^{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right) \cdot t}}{-\left(\lambda_{MCM\_005} + \lambda_{MCMD\_005} + \lambda_{MCT\_005}\right)} + 0.993.$$

The probabilities of ensuring the regulated parameters of the product workpiece are calculated in the process of its manufacturing by steps (subsystems of the 2nd order) and by technological operations (subsystems of the 1st order) if the limit states of the elements of the technological environment are not reached during the development of the technological process (technical system) for the manufacture of the reduction-gear housing when solving a system of differential equations. The calculated probabilities are grouped by surface treatment methods within a certain operation. After that, the probabilities of ensuring the regulated parameters of the workpiece of the part are determined if the limiting states of the elements of the process media "metal-cutting machine – clamping device – tool" are not reached in total for the surface (set of surfaces). Chamfering for internal and external cylindrical surfaces, threaded elements, etc. is interpreted under the totality of surfaces. Determining the probabilities of ensuring regulated quality parameters in the process of machining parts have their own characteristics for the principles of object-oriented and functionally-oriented design. The probability of ensuring the regulated parameters of product quality when the limit states of the elements of the process media "metal-cutting machine – clamping device – tool" for a certain technological operation is not reached is calculated as the probability of a sequence of independent events when multiplying the probabilities by manufacturing steps for the principle object-oriented design of technological processes for the manufacture of parts. In contrast, the probability of ensuring the regulated quality parameters of the product when the limiting states of the elements of the process media "metal-cutting machine – clamping device – tool" for a certain technological operation is not reached is defined as the probability

of ensuring the quality parameters of the datum surface with high requirements for accuracy and quality of its processing within the technological operation for the principle of functionally oriented design of technological processes for manufacturing parts. The probability of ensuring the regulated parameters of product quality for a technological operation is defined as the probability of a sequence of independent events for one surface (set of surfaces) when processing surfaces of the same type with the same technological routes. The value of the probability of ensuring the regulated parameters of product quality in total for technological operations according to the principles of object-oriented and function-oriented design of technological processes for manufacturing parts will be the same for an operation with one manufacturing step.

The results of computer simulation for operation 005 of processing the reduction-gear housing are presented in Fig. 7 and Fig. 8.

Numerical values of probabilities: $P_1(t) = 0.996$; $F_2(t) = 8.09 \cdot 10^{-6}$; $F_3(t) = 1.64 \cdot 10^{-5}$; $F_4(t) = 3.69 \cdot 10^{-3}$ calculated after substituting the values $\lambda_{MCM\_005} = 5.48 \cdot 10^{-6}$ min$^{-1}$; $\lambda_{MCMD\_005} = 1.111 \cdot 10^{-5}$ min$^{-1}$; $\lambda_{MCT\_005} = 2.5 \cdot 10^{-3}$ min$^{-1}$ (see 4.2-4.4), $t_0 = 1.48$ min for machining the reduction-gear housing at operation 005.

The probability of ensuring the regulated parameters of the product workpiece during its processing at operation 005 of the reduction-gear housing when the limit states of the elements of the process media "metal-cutting machine – clamping device – tool" are not reached is shown in Fig. 7. The limiting values of the direct manufacturing time of the workpiece at operation 005 are calculated for the regulated value of the gamma-percentile operation time to failure.
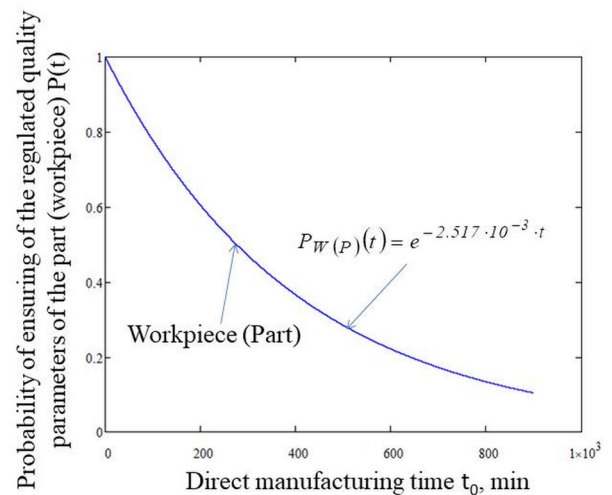


*Fig.7. The probability of ensuring the regulated parameters of the workpiece of the reduction-gear housing during its processing at operation 005 if the limit states of the elements of the process media are not reached*

When testing the developed methodology, it was found that the metal-cutting tool is the weakest element in the technological system in terms of reliability and has the greatest impact on the quality of parts. At the same time, the durability of a metal-cutting tool serves as its average time to failure. The metal-cutting machine has the least influence on the formation of regulated parameters for the quality of parts from the elements of the process media "machine – clamping device – tool".

## 6. Optimization synthesis of the technological process of manufacturing the reduction-gear housing

The limit value of the machining time of the technological transition (operations, etc.) is determined during the design and technological preparation of production in according with the established regulated reliability parameters and the law of their distribution.

The machining time maximum in a specific technological operation is the objective function of the problem of optimization and the process of the synthesized parameters:

$$t_{0\max}(t,S,n(V)) \to \max, \qquad (25)$$

where: $t$, $S$, $n$, $V$ are elements of the cutting modes: the cutting depth, the feed rate, the spindle rotation speed, and the cutting speed, respectively.

The system of equations for the problem optimization in the general case for regulated gamma-percentile operating time to failure is written as:

$$\begin{cases} (t) = e^{-(A \cdot t)}; \\ P(t) = P(t_{\gamma 1}); P(t) = P(t_{\gamma 2}); P(t) = P(t_{\gamma 3}); \\ P(t) = P(t_{\gamma 4}); P(t) = P(t_{\gamma 5}); P(t) = P(t_{\gamma 6}); \end{cases} \qquad (26)$$

where: $P(t) = e^{-(A \cdot t)}$ is a reliability function obtained as a result of the solution of the system of the Chapman-Kolmogorov differential equations for a particular technological operation; $P(t_{\gamma 1})$, ..., $P(t_{\gamma 6})$ are regulated gamma-percentile operating times to failure.

The solution to the optimization problem for planning the technological process of parts manufacturing is of the form:

$$t_{0k} \le [t_0], \qquad (27)$$

where: $t_{0k}$ is the machining time for implementing the $k$-th technological step ($k = 1, \ldots, n$), $[t_0]$ is the limit value of the machining time according to the regulated reliability indicator regularity of its change for a particular workpiece.

The machining time for machining operations:
– for feed per spindle revolution of the metal-cutting tool (workpiece of a part):

$$t_0 = L_{mach.st.} / (S_{0\,pasp.} \cdot n_{pasp.}), \qquad (28)$$

– for feed per minute of the metal-cutting tool (workpiece of a part):

$$t_0 = L_{mach.st.} / S_{\min.pasp.}, \qquad (29)$$

where: $L_{mach.st.}$ is the calculated full required length of the machining step of the metal-cutting tool, $S_{0pasp.}$, $S_{min.pasp.}$, $n_{pasp.}$ are values of feed per spindle revolution, feed per minute and rotary speed according to the certified data of the metal-cutting machine, respectively.

$$L_{mach.st.} = \ell + \ell_1 + \ell_2, \qquad (30)$$

where: $\ell$ is the length of the surface to be machined, mm, $\ell_1$ is the length of travel required for cutting tool approach, mm, $\ell_2$ is the length of overtravel of the cutting tool, mm.

The calculated length of the machining step of the metal-cutting tool remains constant while processing a specific surface of the workpiece at the technological step of a particular technological operation.

The following conditions will be the optimization criteria for a constant value of the working stroke of the metal-cutting tool ($L_{mach.st.}$ = const.):

$$\begin{cases} S_0 \ge [S_0], \\ n \ge [n], \\ S_{\min.} \ge [S_{\min.}], \\ T_{MCT} \le [T_{MCT}], \end{cases} \qquad (31)$$

where: $[S_0]$, $[n]$, $[S_{min}]$ are the limit values of feed per spindle revolution, rotary speed and feed per minute for a particular method of processing and providing the required accuracy, quality of the surface layer of the workpiece and reliability indicators; $T_{MCT}$, $[T_{MCT}]$ are the actual and limit value of the tool resistance, respectively.

The results of solving the optimization problem (27) for a specified gamma-percentile operating time to failure ($t_\gamma$) in the manufacture of the reduction-gear housing in operation 005 is shown in Fig. 9. The gamma-percentile time to failure for a specific part is determined on the basis of the calculation of dimensional chains based on the gamma-percentile operating time to failure of the machine regulated by the official purpose.

This article considers a variant of a complex solution to the problem for all possible regulated values of the gamma-percentile time to failure (Fig. 9). The dependence graph of the probability of providing regulated parameters in the process of machining the workpiece of the reduction-gear housing at operation 005 $P_1(t) = e^{-2.517 \cdot 10^{-3} \cdot t}$ is drawn for the results of mathematical modelling. The regulated parameters of the gamma-percentile time to failure (1.00; 0.99; 0.95; 0.90; 0.80; 0.50) are plotted along the y-axis. After that, lines parallel to the x-axis are built until they intersect with the resulting graph. The intersection points will give the desired values of the limiting values of the direct manufacturing time for the regulated parameters of the gamma-percentage time to failure in operation 005.

The system of equations (26) for operation 005 under machining of the reduction-gear housing is presented in general by:

$$\begin{cases} Đ(t) = e^{-2.517 \cdot 10^{-3} \cdot t}; \\ P(t) = 1.00; P(t) = 0.99; P(t) = 0.95; \\ P(t) = 0.90; P(t) = 0.80; P(t) = 0.50. \end{cases} \qquad (32)$$

The limit values of the machining time that is inversely proportional to the cutting modes (feed and rotary speed) are set by the results of calculations in Mathcad: for $P(t_\gamma) = 1$ $[t_0] = 0$ min; for $P(t_\gamma) = 0.99$ $[t_0] = 3.99$ min; for $P(t_\gamma) = 0.95$ $[t_0] = 20.38$ min; for $P(t_\gamma) = 0.9$ $[t_0] = 41.87$ min; for $P(t_\gamma) = 0.8$ $[t_0] = 88.67$ min; for $P(t_\gamma) = 0.5$ $[t_0] = 275.43$ min (Fig. 8).

It is necessary to consider in optimizing the cutting modes [31] that the increase in the rotary speed and feed reduces the machining time. The metal-cutting tool resistance also decreases under such conditions [21, 45].

The calculations were performed for $P(t_\gamma) = 0.99$ and $[t_0] = 3.99$ min (Fig. 8).

Condition (26) is fulfilled according to table 1 and Fig. 8:

$$t_{0\ 005} = 1.48 < [t_0] = 3.99 \text{ min.}$$

The following cutting modes are accepted for the selected mode of the technological process during milling of a basic plane in operation 005 for machining the reduction-gear housing according to the technical documentation: $t = 3$ mm; $S_{min.pasp.} = 200$ mm/min; $n_{pasp.} = 125$ min$^{-1}$, the calculated length of the working stroke of the metal-cutting tool $L_{mach.st.} = 296$ mm [21].

To provide the specified gamma-percentile ope-rating time to failure $P(t_\gamma) = 0.99$ for the limit machining time $[t_0] = 3.99$ min, the limit value of the feed per minute is determined by (29):
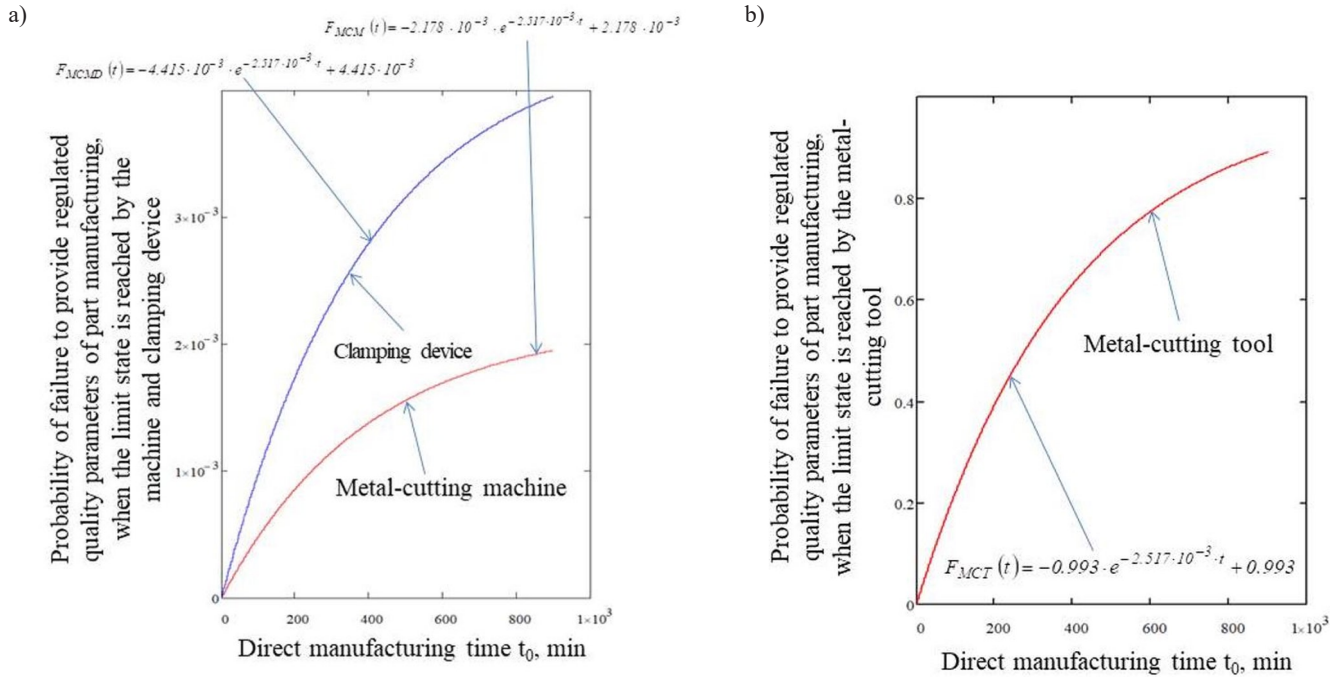
a)

$F_{MCM}(t) = -2.178 \cdot 10^{-3} \cdot e^{-2.517 \cdot 10^{-3} \cdot t} + 2.178 \cdot 10^{-3}$

$F_{MCMD}(t) = -4.415 \cdot 10^{-3} \cdot e^{-2.517 \cdot 10^{-3} \cdot t} + 4.415 \cdot 10^{-3}$

Clamping device

Metal-cutting machine

b)

Metal-cutting tool

$F_{MCT}(t) = -0.993 \cdot e^{-2.517 \cdot 10^{-3} \cdot t} + 0.993$

*Fig.8. Probabilities of failure to ensure the regulated quality parameters at operation 005 machining of the reduction-gear housing when the limit state is reached: a) for metal-cutting machine and clamping device, b) for metal-cutting tool*

$$\left[ S_{\min. \, pasp.} \right] = L_{mach.st.}/[t_0], \tag{33}$$

$$\left[ S_{\min. \, pasp.} \right] = 296/3.99 = 74.19 \text{ mm/min}.$$

$$S_{\min.nacn.} = 200 \; S_{\min.\ddot{\imath}\grave{a}\tilde{n}\grave{\imath}.} = 200 \text{ mm/min} \geq \left[ S_{\min.} \right] = 74.19 \text{ mm/min}.$$

Fulfilment of condition (31) according to the results (33) is provided by the nearest more certificate value $S_{\min.pasp.} = 80$ mm/min and the following (ascending) value of a feed rate for the horizontal milling machine used in operation 005.

In particular, for operation 005 in the technological process of manufacturing of the reduction-gear housing $S_{\min.pasp.} = 200$ mm/min that provides the fulfilment of condition (32):

## 7. Conclusions

The main conclusions have been drawn based on the research results:

The process of providing the reliability for the metal-processing technological system in a certain specialised operation of the technological process of a workpiece manufacturing is checked by optimiz-
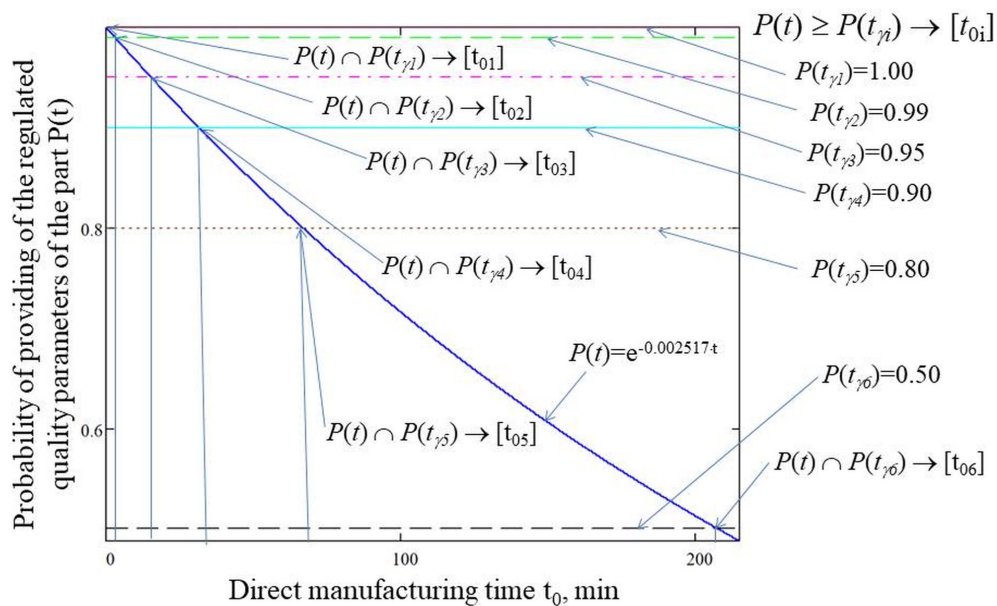


*Fig. 9. Checking the condition of ensuring workability of a workpiece according to the regulated reliability parameters (gamma-percentile operation time to failure ($t_\gamma$)) in technological operation 005 of machining of the reduction-gear housing*

ing the failure rate of the metal-cutting machine, clamping device, metal-cutting tool and rational choice of cutting modes. The process of providing the reliability for the elements of the technological system at a specific technological step during the workpiece manufacturing is checked by optimizing of the failure rate of the metal-cutting tools and rational choice of cutting modes. An increase in the rotary speed and (or) feed rate reduces the machining time and increases the probability of providing regulated quality parameters of the part workpiece. However, the intensification of the cutting modes of the workpiece surfaces reduces the tool life. It increases the probability of failure to provide regulated quality parameters of the workpiece when the metal-cutting tool reaches its limit state.

Optimization synthesis of parameters of manufacturing processes improves manufacturing efficiency, guarantee processing quality, prolongs the life of cutting tools, reduces the consumption of energy resources and all this adds value to the production process. In addition, the method is open for further modelling and optimization.

After testing the developed methodology, it was found that in terms of reliability, the metal cutting tool is the weakest element in the process media „machine-clamping device-cutting tool" and has the greatest influence on the formation of adjustable quality parameters of the machined parts. Meanwhile, metal cutting machine has the least impact on the quality parameters of the machined parts in the technological system.

Finally, the proposed optimization technique during the manufacturing of parts has been successfully integrated and tested in the real industry. As a result of optimization of cutting modes of the technological operation 005 (horizontal-milling) to manufacture the reduction-gear housing, the limit value of direct manufacturing time and feed per minute are $[t_0]$ = 3.99 min and $[S_{min}]$ = 74.19 mm/min, respectively for the regulated gamma-percentile operation time to failure $P(t_\gamma)$=0.99. Rational cutting modes for this technological operation provide working capacity according to technical requirements.

# References

1. Abbas, M., ElMaraghy, H. 2018. Synthesis and optimization of manufacturing systems configuration using co-platforming, CIRP Journal of Manufacturing Science and Technology 20, 51-65. https://doi.org/10.1016/j.cirpj.2017.09.006

2. Arévalo-Ruedas, J.H., Espinel-Blanco, E., Florez-Solano, E. 2021. Statistical analysis of cutting tool wear in machining centers, Journal of Physics: Conference Series 2139 012019. https://doi.org/10.1088/1742-6596/2139/1/012019

3. Bazaluk, O.; Slabyi, O.; Vekeryk, V.; Velychkovych, A.; Ropyak, L.; Lozynskyi, V. 2021. A technology of hydrocarbon fluid production intensification by productive stratum drainage zone reaming, Energies 14 (12): 3514. http://dx.doi.org/10.3390/en14123514

4. Bertsche B. 2008. Reliability in Automotive and Mechanical Engineering. Berlin Heidelberg: Springer-Verlag, 492p. https://doi.org/10.1007/978-3-540-34282-3

5. Birolini A. 2014. Reliability Engineering: Theory and Practice. Berlin Heidelberg: Springer-Verlag. 626p. https://doi.org/10.1007/978-3-662-05409-3

6. Bobalo, Y.; Seniv, M.; Yakovyna, V.; Symets, I. 2018. Method of Reliability Block Diagram Visualization and Automated Construction of Technical System Operability Condition, Advances in Intelligent Systems and Computing III, 871: 599-610. http://dx.doi.org/10.1007/978-3-030-01069-0_43

7. Bobalo, Yu.,Volochiy, B., Lozynsky, O., Mandzyy, B., Ozirkovskyy, L., Fedasyuk, D., Shcherbovskykh, S., Yakovyna, V. 2013. Mathematical Models and Methods of Reliability Analysis of Radioelectronic, Electrical and Software Systems. Lviv, Lviv Polytechnic National University. 300p. [in Ukrainian]

8. Braband, J. 2003. Improving the risk priority number concept, Journal of System Safety 39 (2): 21-23

9. Cai, K.-Y., Hu, D.-B., Bai, C.-G., Hu, H., Jing, T. 2008. Does software reliability growth behavior follow a non-homogeneous Poisson process, Information and Software Technology 50, 1232-1247. https://doi.org/10.1016/j.infsof.2007.12.001

10. Chen, Z., Chen, Z., Zhou, D., Xia, T., Pan, E. 2021. Reliability evaluation for multi-state manufacturing systems with quality-reliability dependency, Computers & Industrial Engineering 154, 107166. https://doi.org/10.1016/j.cie.2021.107166

11. Crawley, F. 2020. Failure modes and effects analysis (FMEA) and failure modes, effects and criticality analysis (FMECA). In: A guide to hazard identification methods. 2th.ed. 103-109. Elsevier Inc. https://doi.org/10.1016/B978-0-12-819543-7.00012-4

12. Di Bona, G.; Silvestri, A.; Forcina, A.; Petrillo, A. 2018. Total efficient risk priority number (TERPN): a new method for risk assessment, Journal of Risk Research 21 (11), 1384-1408. http://dx.doi.org/10.1080/13669877.2017.1307260

13. Haken H. 2006. Information and Self–Organization. A Macroscopic Approach to Complex Systems: Third Enlarged Edition. Berlin: Springer. 258p. http://dx.doi.org/10.1007/3-540-33023-2

14. Hu, Z.; Du, X. 2017. System reliability prediction with shared load and unknown component design details, AI EDAM 31(3): 223-234. http://dx.doi.org/10.1017/S0890060417000130

15. International standard IEC 60812. 2006. Analysis techniques for system reliability - *Procedure for failure mode and effects analysis* (*FMEA*). 93p.

16. ISO 14224:2016, Third Edition (2016-09-15). 2016. Petroleum, petrochemical and natural gas industries - Collection and exchange of reliability and maintenance data for equipment. ISO copyright office, Vernier, Geneva, Switzerland, 280p. 10/06/2016 22:31:11 MDT

17. Kim, S. I.; Lee, H. Y.; Song, J. S. 2018. A study on characteristics and internal exposure evaluation of radioactive aerosols during pipe cutting in decommissioning of nuclear power plant, Nucl. Eng. Technol. 50 (7): 1088-1098. http://dx.doi.org/10.1016/j.net.2018.06.010

18. Klocke F. 2011. Manufacturing Processes 1: Cutting. Berlin: Springer-Verlag. 504p. https://doi.org/10.1007/978-3-642-11979-8

19. Kopei, V. B.; Onysko, O. R.; Panchuk, V. G. 2019. Component-oriented acausal modeling of the dynamical systems in Python language on the example of the model of the sucker rod string, PeerJ Computer Science 10: 227. http://dx.doi.org/10.7717/peerj-cs.227

20. Korba, P., Hunady, R., Hovanec, M., Racek, B., Pavelka, P. 2021, Fatigue life analysis of an aircraft brake component to prevent damage and ensure operational safety, Engineering Failure Analysis 129, 105653. https://doi.org/10.1016/j.engfailanal.2021.105653.

21. Kusiy Ya. 2021. Scientific and applied bases of technological inheritability of quality parameters for providing of operational characteristics of products: Thesis of Doctor of technical sciences. Lviv: Lviv Polytechnic National University. 432p https://lpnu.ua/sites/default/files/2021/

dissertation/16474/dysertdsckusyiyaroslav.pdf [in Ukrainian].

22. Kusyi, Y. M.; Kuk, A. M. 2020. Investigation of the technological damageability of castings at the stage of design and technological preparation of the machine Life Cycle, Journal of Physics: Conference Series 1426. http://dx.doi.org/10.1088/1742-6596/1426/1/012034.

23. Kusyi, Y.; Stupnytskyy, V. 2020. Optimization of the Technological Process Based on Analysis of Technological Damageability of Casting. In: V. Ivanov, J. Trojanowska, I. Pavlenko, J. Zajac, D. Perakovič (eds). Advances in Design, simulation and manufacturing III, 276-284. Springer Nature Switzerland AG. http://dx.doi.org/10.1007/978-3-030-50794-7_27

24. Lai, R., Garg, M. 2012. A detailed study of NHPP software reliability models, Journal of Software 7(6), 1296-1306. doi:10.4304/jsw.7.6.1296-1306

25. Latinovic, T.; Preradović, D.; Barz, C. R.; Pop Vadean, A.; Todić, M. 2019. Big Data as the basis for the innovative development strategy of the Industry 4.0, IOP Conference Series: Materials Science and Engineering 477(1): 012045. https://iopscience.iop.org/article/10.1088/1757-899X/477/1/012045

26. Lee, C., Park, J., Choi, J., Ha, J., Lee, S. 2021. Control logic synthesis for manufacturing systems using Markov decision processes, IFAC-Papers On Line 54 (20), 495-502. https://doi.org/10.1016/j.ifacol.2021.11.221

27. Li, X., Fang, Z., Yin, C. 2020. A machine tool matching method in cloud manufacturing using Markov decision process and cross-entropy, Robotics and Computer Integrated Manufacturing 65, 101968. https://doi.org/10.1016/j.rcim.2020.101968

28. Li, Y., Liu, Q., Tong, R., Ciu, X. 2015. Shared and service-oriented CNC machining system for intelligent manufacturing process, Chinese Journal of Mechanical Engineering, 28 (6), 1100-1108. DOI: 10.3901/CJME.2015.1010.119

29. Nachlas J. A. 2017. Reliability Engineering: Probabilistic Models and Maintenance Methods, 2nd ed. CRC Press, Taylor & Francis Group. 378p. https://doi.org/10.1201/9781315307596

30. Nakagawa T. 2005. Maintenance Theory of Reliability. London Limited: Springer-Verlag. 269p. https://doi.org/10.1007/1-84628-221-7.

31. Ostasevicius, V.; Paulauskaite-Taraseviciene, A.; Paleviciute, I., Jurenas, V., Griskevicius, P., Eidukynas, D., Kizauskiene, L. 2022. Investigation of the Robotized Incremental Metal-Sheet Forming Process with Ultrasonic Excitation. Materials 15(3): 1024. http://dx.doi.org/10.3390/ma15031024.

32. Paraschos, P.D., Xanthopoulos, A.S., Koulinas, G.K., Koulouriotis, D.E. 2022. Machine learning integrated design and operational management for resilient circular manufacturing systems, Computers & Industrial Engineering 167, 107971. https://doi.org/10.1016/j.cie.2022.107971.

33. Pham, H. 2003. Handbook of Reliability Engineering. Springer-Verlag London Limited. 663p. https://doi.org/10.1007/b97414

34. Pham, H. 2006. System software reliability. Springer-Verlag London Limited. 442p. https://doi.org/10.1007/1-84628-295-0

35. Pham, H. 2003. Software reliability and cost models: Perspectives, comparison, and practice, European Journal of Operational Research, 149 (3), 475-489. https://doi.org/10.1016/S0377-2217(02)00498-8

36. Roci, M., Salehi, N., Amir, S., Shoaib-ul-Hasan, S., Asif, F.M.A., Mihelič, A., Rashid, A. 2022. Towards circular manufacturing systems implementation: A complex adaptive systems perspective using modelling and simulation as a quantitative analysis tool, Sustainable Production and Consumption 31, 97-112 https://doi.org/10.1016/j.spc.2022.01.033

37. Ropyak, L. Y.; Pryhorovska, T. O.; Levchuk, K. H. 2020. Analysis of materials and modern technologies for PDC drill bit manufacturing, Progress in Physics of Metals 21 (2): 274-301. http://dx.doi.org/10.15407/ufm.21.02.274.

38. Sabri-Laghaie, K., Fathi, M., Zio, E., Mazhar, M. 2022. A novel reliability monitoring scheme based on the monitoring of manufacturing quality error rates, Reliability Engineering and System Safety 217, 108065. https://doi.org/10.1016/j.ress.2021.108065

39. Saez, M., Barton, K., Maturana, F., Tilbury, D.M. 2022. Modelling framework to support decision making and control of manufacturing systems considering the relationship between productivity, reliability, quality, and energy consumption, Journal of Manufacturing Systems, 62, 925-938. https://doi.org/10.1016/j.jmsy.2021.03.011

40. Salonitis, K., Kolios, A. 2013. Reliability assessment of cutting tools life based on advanced approximation methods, Procedia CIRP 8, 397-402. https://doi.org/10.1016/j.procir.2013.06.123

41. Shakhovska, N., Yakovyna, V., Kryvinska, N. 2020. An Improved Software Defect Prediction Algorithm Using Self-organizing Maps Combined with Hierarchical Clustering and Data Preprocessing. In: Hartmann, S., Küng, J., Kotsis, G., Tjoa, A.M., Khalil, I. (eds) Database and Expert Systems Applications. DEXA 2020. Lecture Notes in Computer Science(), vol 12391. Springer, Cham. https://doi.org/10.1007/978-3-030-59003-1_27.

42. Signoret, J.-P., Leroy, A. 2021. Reliability Assessment of Safety and Production Systems. Analysis, Modelling, Calculations and Case Studies, Springer Nature Switzerland AG. 887p. https://doi.org/10.1007/978-3-030-64708-7

43. Sonsino, C. M.; Heim, R.; Melz, T. 2016. Lightweight-structural durability design by consideration of variable ampli- tude loading, International Journal of Fatigue 91: 328-336. http://dx.doi.org/10.1016/j.ijfatigue.2015.07.030.

44. Sosnovskiy L.;·Sherbakov S. 2016. Mechanothermodynamics. Springer: Cham, Switzerland, 155p. DOI 10.1007/978-3-319-24981-0

45. Sun, H., Liu, Y., Pan, J., Zhang, J., Ji, W. 2020. Enhancing cutting tool sustainability based on remaining useful life prediction, Journal of Cleaner Production 244, 118794. https://doi.org/10.1016/j.jclepro.2019.118794

46. Tönissen, S., Rey, J., Klocke, F. 2015. Economic efficiency of manufacturing technology integration, Journal of Manufacturing Systems 37, 173-181. https://doi.org/10.1016/j.jmsy.2015.07.003

47. Volochiy, B., Yakovyna, V., Mulyak, O., Kharchenko, V. 2018. Availability model of critical nuclear power plant instrumentation and control system with non-exponential software update distribution. In:, et al. Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2017. Communications in Computer and Information Science, vol 826. Springer, Cham. https://doi.org/10.1007/978-3-319-76168-8_1

48. Xie, L., Habrekke, S., Liu, Y., Lundteigen, M.A. 2019. Operational data-driven prediction for failure rates of equipment in safety instrument systems: A case study from the oil and gas industry, Journal of Loss Prevention in the Process Industries 60, 96-105. https://doi.org/10.1016/j.jlp.2019.04.004.

49. Yakovenko, I.; Permyakov, A.; Prihodko, O.; Basova, Y.; Ivanova, M. 2020. Structural Optimization of Technological Layout of Modular Machine Tools. In: , et al. Advanced Manufacturing Processes. InterPartner-2019, Lecture Notes in Mechanical Engineering. Springer, Cham.: 352-363. https://doi.org/10.1007/978-3-030-40724-7_36.

50. Yakovyna, V., Seniv, M., Symets, I., Sambir, N. 2020. Algorithms and software suite for reliability assessment of complex technical systems, Radio Electronics, Computer Science, Control 4, 163-177. https://doi.org/10.15588/1607-3274-2020-4-16

51. Yakovyna, V., Symets, I. 2021. Reliability assessment of CubeSat nanosatellites flight software by high-order Markov chains, Procedia Computer Science 192, 447-456. https://doi.org/10.1016/j.procs.2021.08.046

52. Yang, X., He, Y., Liao, R., Cai, Y., Ai, J. 2022. Integrated mission reliability modelling based on extended quality state task network for intelligent multistate manufacturing systems, Reliability Engineering and System Safety 223, 108495. https://doi.org/10.1016/j.ress.2022.108495.

53. Yeung, A. W. K. 2019. The "As Low As Reasonably Achievable" (ALARA) principle: a brief historical overview and a bibliometric analysis of the most cited publications, Radioprotection 54(2): 103-109. https://doi.org/10.1051/radiopro/2019016.

54. Yoshimura, M. 2007. System Design Optimization for Product Manufacturing, Concurrent Engineering 15 (4): 329-343. http://dx.doi.org/10.1177/1063293x07083087.

55. Zhang, X., Ming, X. 2021. An implementation for smart manufacturing information system (SMIS) from an industrial practice survey. Computers & Industrial Engineering 151, 106938. https://doi.org/10.1016/j.cie.2020.106938

56. Zio, E. 2009. Reliability engineering: old problems and new challenges, Reliability Engineering and System Safety 94: 125-149.
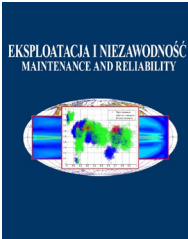
# Multi-criteria optimization of the turning parameters of Ti-6Al-4V titanium alloy using the Response Surface Methodology

Indexed by:
Web of Science Group

Rafał Kluz[a,*], Witold Habrat[b], Magdalena Bucior[a], Krzysztof Krupa[c], Jarosław Sęp[a]

[a] Rzeszow University of Technology, Department of Manufacturing and Production Engineering, Faculty of Mechanical Engineering and Aeronautics, al. Powst. Warszawy 8, 35-959 Rzeszow, Poland

[b] Rzeszow University of Technology, Department of Manufacturing Techniques and Automation, Faculty of Mechanical Engineering and Aeronautics, al. Powst. Warszawy 8, 35-959 Rzeszow, Poland

[c] Rzeszow University of Technology, Department of Materials Science, Faculty of Mechanical Engineering and Aeronautics, al. Powst. Warszawy 12, 35-959 Rzeszów, Poland

## Highlights

- The RSM method is an effective tool for modeling the turning process of the Ti6Al4V titanium alloy.

- The use of a polycrystalline diamond cutting insert for dry turning enables environmentally friendly machining.

- Too high temperature during the cutting process leads to the release of chemical compounds.

- Multi-criteria optimization of the turning process enables to obtain cutting parameters that meet the quality requirements.

## Abstract

The paper depicts an application of Response Surface Methodology (RSM) for predicting selected parameters in turning of Ti-6Al-4V titanium alloy using polycrystalline diamond tool. Response surface plots that are generated by the model helps in determining the optimum combination of input factors (cutting speed $v_c$ and feed rate $f$) for best possible surface roughness ($Sa$), cutting force ($F_c$) and temperature ($T$) for dry and cooling turning. The methodology of multi-criteria optimization was used to establish the interaction between input parameters and given responses.

## Keywords

Response Surface Methodology, titanium alloy, multi-criteria optimization, turning process.

## 1. Introduction

Titanium alloys have a wide range of applications [28]. This is due to properties such as high specific strength (ultimate strength to density ratio), resistance on high temperature and corrosive environment.This kind of material is used to manufacture elements of airplanes, helicopters such as the construction of blades in engines, window framesin the cockpit [28] or for landing gear [1]. Plate and frame heat exchangers made with titanium alloy are used in power plants, refineries, air conditioning systems, chemical plants, offshore platforms, surface ships and submarines [28]. Knee prostheses, trauma fixation devices (nails, plates, screws), surgical devices, pacemakers or implants [16] made with this alloys are also used in medicine. Objects produced from titanium alloys are generally characterized by high requirements for accuracy and surface roughness, and therefore they are shaped by machining processes. The processes for titanium alloys are difficult due to the properties, which is hard-to-machine materials [6]. This leads to occurrence of high value of temperature and the rapid wear of the cutting tools. The available studies in this scientific field [3, 11, 18] indicate the main causes of these difficulties, including: low thermal conductivity, high chemical reactivity with most tool materi-als, thermoplastic instability during machining, tendency for chips to stick totool, or finally the tendency to build up edge, which promotes chipping of the cutting edge [27].

Due to the low value of the thermal conductivity of titanium, the heat generated in the cutting zone is not dissipated, but this is concentrated on the edges and tips of the cutting blade. High temperature (in the treatment zone exceeding even 1100° C) leads to intensive wear of the blade and its plastic deformation. With respect to the high temperature in the machining zone, the phenomenon of softening takes place and the cutting resistance increases. The strong chemical affinity of titanium to tool materials promotes adhesion and the formation of built-up edges on the tips of cutting tools, and, consequently, their rapid blunting and chipping. In turn, the high elasticity of titanium causes the formation of elastic deformations and vibrations during machining, which affects the changes in the instantaneous values of the depth of cut. At small depths, there is no longer cutting with a blade, but only plastic deformation, which also strengthens the processed material and increases its ultimate strength and hardness [14, 30].

Improvement in the machinability of titanium alloys can be obtained by lowering temperature value of the process. The recommendations

(*) Corresponding author.
E-mail addresses: R. Kluz (ORCID: 0000-0001-6745-294X): rkktmiop@prz.edu.pl, W. Habrat (ORCID: 0000-0002-9010-8175): witekhab@prz.edu.pl,
M. Bucior (ORCID: 0000-0002-1081-5065): magdabucior@prz.edu.pl, K. Krupa (ORCID: 0000-0003-1822-7230): krupa@prz.edu.pl,
J. Sęp (ORCID: 0000-0003-2544-2211): jsztmiop@prz.edu.pl

for machining titanium alloys [22] indicate on the following features: low cutting speed, high feed and depth of cut (compared to steel machining), with the use of very sharp tools reducing value of loading and the phenomena of workpiece crushing. Taking into account the possibilities for optimization of production costs, research are currently directed towards the processes of shaping titanium parts with methods ensuring maximum efficiency, while maintaining the desired surface geometric structure and the required tool life [3, 5, 13, 24].

In articles and studies on the quality of surface obtained as a result of machining, a lot of attention is paid to design of experiments (DOE) techniques[17]. Experimental procedures are mainly used in the early stages of developing technological processes. In practice, there are many factors that influence the key properties of a part. The technologist's task is to determine factors significant and insignificant in further stages of technological process. On the other hand, without prior knowledge, it is difficult to determine the significance of individual factors. In such circumstances experimental techniques are often used. Despite the determination of an appropriate set of examined factors under study, it is usually difficult to find the optimal conditions in which the technological process is to be carried out. In such a case, one of the experimental techniques should be used to determine such process conditions for which the surface quality is considered satisfactory. One of them are experimental techniques grouped under the common name of the Response Surface Method (RSM). The RSM method is a combination of mathematical and statistical methods especially useful when the process under study is influenced by several factors [4, 10, 29].To build an empirical model, this method uses the quantitative data available from experiments to solve multivariate equations. The solution of these equations is taken as the optimum combination of input parameters. In RSM the equations are represented graphically which are indicative of relation between the input and output factors [19, 23, 26].

The selection of optimal turning parameters is crucial to enhance the quality of the industry's product and economy. There have been few attempts to optimize process parameters in turning Ti-6Al-4V alloy with various process parameters. Artificial neural networks (ANN) [7], Analysis of variance (ANOVA) [12], Fuzzy rule [12, 20] and Taguchi method [9] are among various optimization techniques applied to find the optimal turning process parameters for better results. A brief review of the literature on the optimization of input factors for the turning process of titanium alloys is presented here.

The authors of the articles [21, 22] conducted studies on the effect of changing cutting parameters on the components of the cutting force and surface roughness of Ti-6Al-4V ELI titanium alloy parts with the use of a roughing cutting insert. The research were carried out in the range of cutting speed 25 ÷ 40 m/ min and feed 0.1 ÷ 0.35 mm/ rev. During the tests, the lowest roughness equal to Ra = 0.32 μm was ob-

tained for the feed rate of 0.1 mm / rev and the cutting speed of 25 m/ min. In turn, in [22] it was shown that an increase in the cutting depth from 0.5 mm to 2 mm, at a cutting speed of 35 m/ min and a feed rate of 0.25 mm / rev, increases the surface roughness Ra from 0.3 μm to 2 μm. In the articles [25, 15], the Response Surface Methodology was used to optimize the turning process of the Ti-6Al-4V titanium alloy. The author of the work [25] optimized the turning process using a tool made of a TiAlN coated carbide. He conducted the tests by changing the cutting speed in the range of 30 ÷ 113 m/ min, the feed rate of 0.2 ÷ 0.4 mm/ rev and the depth of cut in the range of 0.5 ÷ 1 mm. The analysis showed that the lowest surface roughness can be obtained at a cutting speed of 72 m/ min and a feed rate of 0.2 mm/ rev. Mia et al. [15] used RSM to study surface roughness and cutting forces in cryogenic turning of a titanium alloy with a WC coated tool. The following parameters were adopted for the tests: cutting speed in the range 78 ÷ 156 m/ min and feed in the range 0.12 ÷ 0.16 mm/ rev. As a result of the analysis, it was shown that the most favorable cutting parameters, allowing to obtain a roughness of Ra = 1.05 μm, is the cutting speed of 78 m/ min at a feed rate of 0.16 mm/ rev, a cutting depth of 1.0 mm and a feed force f = 208 N.

The present study is an investigation the influence of different cutting parameters of the Ti-6Al-4V titanium alloy at different the conditions of dry and cooling turning with a polycrystalline diamond tool. The statistical manner for determining the Response Surface Method was used to model the relationship between the parameters of the cutting process and the components of the cutting forces as well as the parameters of the surface topography. Such a multi criteria optimization technique can offer a reliable solution and a balance among all included outputs. It also helps to provide different solutions that can be very useful to select the most appropriate cutting conditions based on the desired objectives. Properly selected processing parameters are important for ensuring the required reliability in the future exploitation of the shaped products. They ensure shaping of the cutting forces in the range that prevents their excessive growth, which causes deformation of the surface layer material and its strengthening. In addition, reducing the temperature value of the cutting process and the morphology of the phase components of the workpiece surface can prevent surface damage and fatigue fracturing.

## 2. Devices, material and process parameters

The cutting process was examined on a test stand at the Research and Development Laboratory for Aerospace Materials (Rzeszow, Poland) on a CNC lathe Gildemeister NEF 600. The measuring system consists of a Kistler 9257B piezoelectric dynamometer - attached to the turret with a VDI holder, it enables the measurement of the cutting force components: cutting force ($F_c$), feed force ($F_f$) and passive force ($F_p$) (Fig. 1). The signal from the dynamometer is amplified by
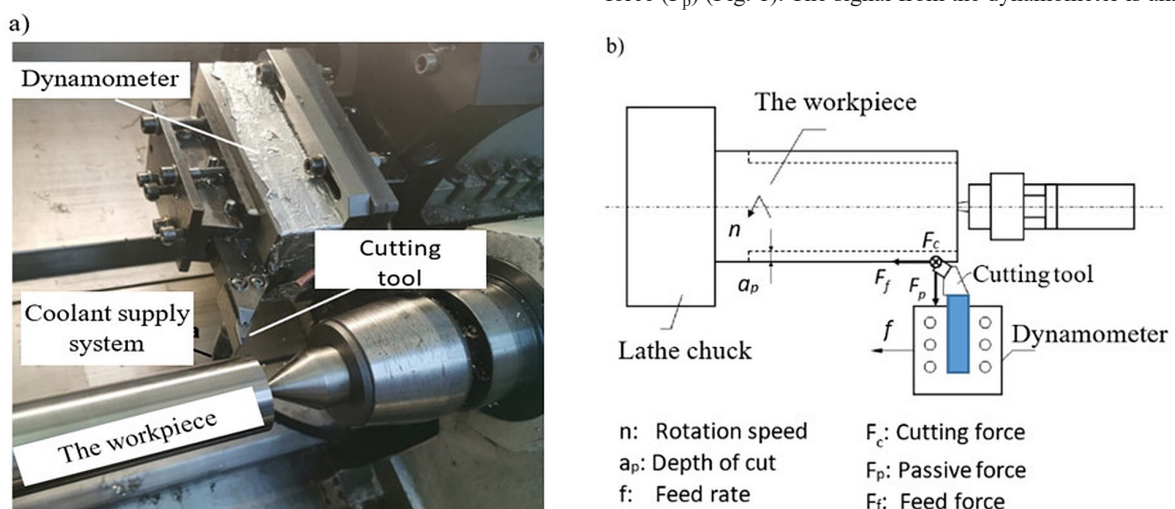


Fig. 1.View of the working space of the NEF600 lathe: a) fixing the cutting tool in the dynamometer, b) kinematic diagram of the stand

a 5070 type charge amplifier and then transmitted to a computer via USB using the National Instruments 16-bit NI 9215 analog-to-digital converter. Visualization, processing and saving of the signal is carried out using a program developed in the LabVIEW environment. The sampling frequency of the signal was set at 1 kHz (Fig. 2).

The X6540sc thermal imaging camera by FLIR (Fig. 2) with a resolution of 640 ×512 pixels and a recording speed of 126Hz: 640×512 to 4011Hz: 64×8 was used to measure values of temperature during the turning process.
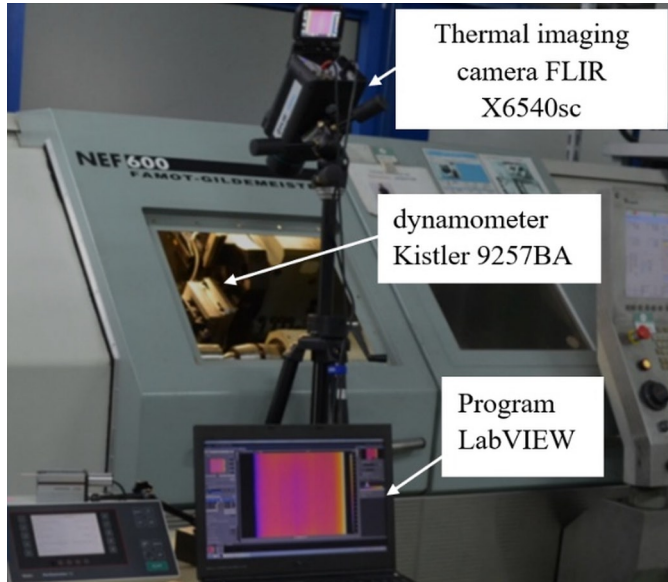


*Fig. 2. Configuration of the stand with thermovision measurement during turning process*



*Fig. 3. Cutting insert VCGT 160404: a) dimensions, b) photo*

The statistical method of determining the Response Surface Method was used to model the relationship between the parameters of the cutting process and the components of the cutting force as well as the parameters of the surface topography. Design-Expert 12 software by Stat-Ease (*Minneapolis*, Minnesota) was used for modeling.

A central, wall-centered composition plan was adopted during the research. The order of runs were generated by the program Design-Expert 12 (Table 2). The range of variability of the input quantities included:

- cutting speed $v_c$ in the range of 120 ÷ 240 m/min
- feed rate $f$ in the range of 0.1 ÷ 0.3 mm/rev,
- constant depth of cut $a_p$= 0.25 mm

To determine the optimal cutting parameters, the weighted sum method was used, because this is one of the best-known methods of multi-criteria optimization. Due to the fact that the objective func-

*Table 1. Chemical composition of the Ti-6Al-4V alloy (wt. %) [2]*

| Ti | C max | Fe max | N max | Al | O max | V | H max | Y max | Rest |
|---|---|---|---|---|---|---|---|---|---|
| Variable | 0.08 | 0.03 | 0.05 | 5.50-6.75 | 0.20 | 3.5-4.5 | 0.015 | 0.005 | 0.40 |

The 3D topography measurements of the treated surface were made using the optical method Alicona Infinite Focus focal differentiation microscope. The measurement included 3D topography on an area of 1.6 × 1.6 mm. This method included noise removal, shape profile filtering, topography imaging with 3D maps, determination of selected surface topography parameters and their statistical evaluation. The parameters of the surface topography were determined in accordance with ISO 25178-2 [8].

The surface observation was conducted using a HITACHI S-3400N Scanning Electron Microscope (SEM) equipped with Energy Dispersive X-ray Spectrometer (EDS) and Wavelength Dispersive X-ray Spectrometer (WDS) systems for the analysis of chemical composition and an Electron Backscattered Diffraction (EBSD) system enabling the determination of the texture of materials and the identification of phase components and morphology of their microstructure. Microstructural examination of the dry turning surface was carried out with using Nikon Epiphot 300 light microscope with NIS-Elements V2.3 software.

The material used in this study was Ti-6Al-4V titanium alloy. Due to its good properties at elevated temperatures, it is one of the species commonly used in the aviation and electrical industries. The chemical composition is provided in Table 1.

A folding tool with a polycrystalline diamond insert was used for the tests. The configuration of the toolkit included:

- lampholder: SVJBL2525M16 JET (Sandvik, Sweden)
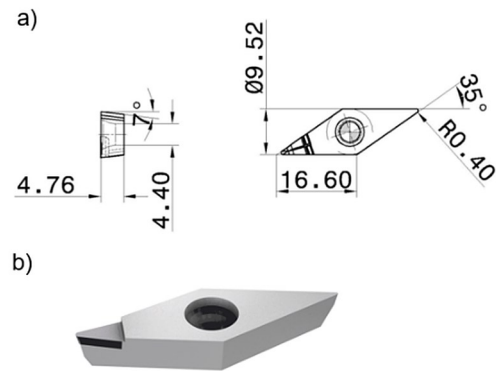- the cutting insert: VCGT 160404 ID5 (Iscar, Israel) (Fig. 3).

*Table 2. Cutting parameters for Ti-6Al-4V titanium alloy, selected by mean of the RSM method*

| No. | Run | Cutting speed $v_c$, m/min | Feed rate $f$, mm/rev |
|---|---|---|---|
| 1 | 8 | 120 | 0.1 |
| 2 | 7 | 240 | 0.1 |
| 3 | 5 | 120 | 0.3 |
| 4 | 9 | 240 | 0.3 |
| 5 | 6 | 120 | 0.2 |
| 6 | 4 | 240 | 0.2 |
| 7 | 3 | 180 | 0.1 |
| 8 | 1 | 180 | 0.3 |
| 9 | 10 | 180 | 0.2 |
| 10 | 2 | 180 | 0.2 |

tions are expressed on various scales of values, their transformation was performed to a dimensionless form assuming values in the range [0, 1].

Taking into account the features of turning, surface roughness was a critical parameter ensuring the correct course of the process. Additionally, in the case of turning with cooling, the passive force responsible for shape errors was measured. In the case of dry turning, the

temperature values were measured. The values should be the range that does not cause the release of chemical compounds adhesively adhering to the surface of the part and the phenomena of graphitization.

## 3. Results and discussion

### 3.1. Turning with tool cooling

During the precision turning of the Ti-6Al-4V titanium alloy, the courses of the $F_{c\,(cutting\,force)}$, $F_f$ (feed force), $F_p$ (passive force) components of the cutting force were collected using the measurement path presented in Fig. 1b. Cooling and lubricating fluid Ecocool Global 10 was used for cooling. In the conducted research, an aqueous concentrate solution with a concentration of 8% was used. The liquid was fed using a special nozzle that was an integral part of the HPC system holder. The liquid pressure was 8MPa. The analysis of forces in the finishing turning process provides information on the mechanics of the cutting process, taking into account the properties of the workpiece material and tool geometry. Table 3 presents the results of measurements of the cutting force components during turning with cooling through the tool, for the adopted research range.

*Table 3. Cutting force components according to the RSM methodology*

| No. | Cuttingspeed, m/min | Feedrate, mm/rev | $F_c$, N | $F_p$, N | $F_f$, N |
|-----|------|------|--------|-------|-------|
| 1 | 120 | 0.1 | 65.92 | 24.10 | 20.56 |
| 2 | 240 | 0.1 | 63.32 | 22.91 | 19.28 |
| 3 | 120 | 0.3 | 151.11 | 41.74 | 23.67 |
| 4 | 240 | 0.3 | 144.92 | 39.75 | 22.08 |
| 5 | 120 | 0.2 | 111.16 | 32.96 | 23.18 |
| 6 | 240 | 0.2 | 105.27 | 28.18 | 19.90 |
| 7 | 180 | 0.1 | 64.28 | 22.63 | 18.72 |
| 8 | 180 | 0.3 | 14647 | 42.56 | 22.78 |
| 9 | 180 | 0.2 | 107.35 | 30.25 | 21.68 |
| 10 | 180 | 0.2 | 106.26 | 27.82 | 21.62 |

Based on the ANOVA analysis of variance, appropriate models were selected and adequate regression equations were obtained for the values of the $F_c$, $F_f$, $F_p$ force components (1-3).

$$F_C = 25.82 - 0.0407 v_c + 475.303 f - 150.805 f^2 \qquad (1)$$

$$F_p = 17.14 - 0.022 v_c + 90.66 f \qquad (2)$$

$$F_f = 21.10 - 0.017 v_c + 16.60 f \qquad (3)$$

Figure 4 shows the graphs of the components of the cutting forces as a function of the cutting speed and feed rate. The analysis of Fig. 4 shows that the cutting forces components have a very similar course. Both the main cutting force $F_c$ and the passive force $F_p$ practically do not change with increasing cutting speed. Their increase occurs with the increase in feed rate. Increasing the feed rate from f = 0.1mm/ rev to f = 0.2mm/ rev increases the cutting force $F_c$ by 69.23%. Increasing the feed rate to f = 0.3mm/ rev results in a further increase in cutting force by 36.03%. It should be noted that the cutting force $F_c$ and the passive force $F_p$ are strongly correlated with each other. This is evidenced by the value of the Pearson correlation coefficient, r = 0.98. The feed force $F_f$ is correlated to a lesser degree with the cutting force $F_c$ (r = 0.84). Its value decreases slightly with the increase of the cutting speed, while, similarly to the other components, it increases with the increase of the feed rate.
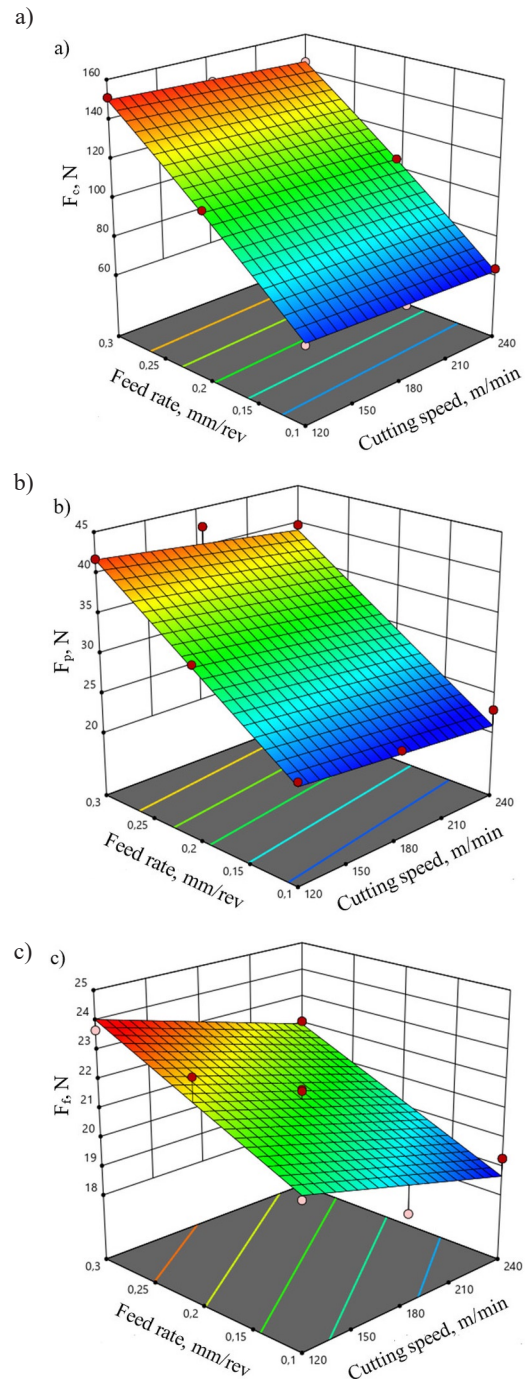


*Fig. 4. Graph of the components of cutting forces: a) $F_c$, b) $F_p$, c) $F_f$*

On the basis of the final turning tests of the Ti-6Al-4V alloy with the use of a polycrystalline diamond cutting insert, the influence of the cutting parameters on the selected parameters of the surface topography was determined. Table 4 presents the results of measurements of the surface topography parameters after turning with the tool cooling.

The analysis of the measurement results showed that the Sa (arithmetical mean of the height deviations of the surface) parameter is strongly correlated with the Sz parameter (the maximum height of surface). Pearson's correlation coefficient is r = 0.99 (Fig. 5).Therefore, the further part of the analysis focuses on one of these parameters (Sa) most often used in industrial practice.
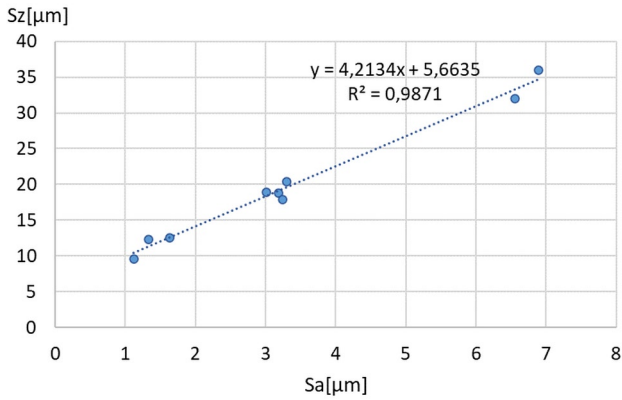
*Fig. 5. Graph of the correlation of the Sa and Sz parameters*

Regardless of the adopted cutting speed, it can be noticed that the roughness Sa increases with the increase of the feed rate value. In the case of a feed of f = 0.3 mm/ rev, an increase in the cutting speed from $v_c$ = 120 m/ min to $v_c$ =180 m/ min causes an increase in roughness by 5.03% to a value of Sa = 6.59 μm. In this case, a further increase in the cutting speed does not significantly affect the surface roughness. For a feed rate of f = 0.1 mm/ rev, an increase in the cutting speed from $v_c$ =120 m/ min to $v_c$ = 180 m/ min results in a slight increase in surface roughness from Sa = 1.33 μm to Sa = 1.63 μm, and then causes it to decrease by 31.28% at a cutting speed of $v_c$ = 240 m/ min (Fig. 6a).
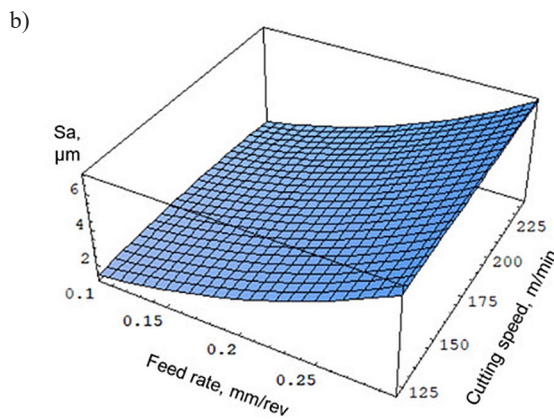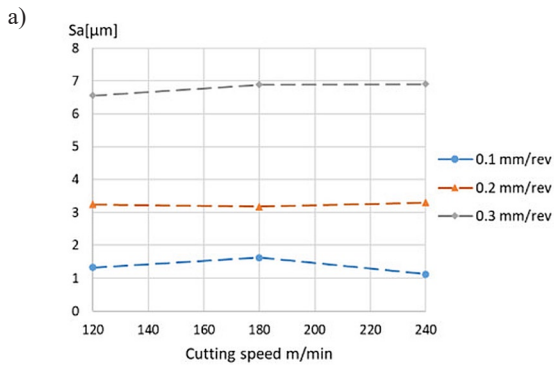


*Fig. 6. Dependence of the Sa parameter on the cutting speed a) test results, b) plot of the regression function*

Based on the ANOVA analysis of variance, an adequate regression model of the Sa parameter was developed (Fig. 6b), (4):

$$S_a = -0.65226 + 9.02652f + 35.9614f^2 + 76.4532f^3 + 0.01230v_c - 0.105fv_c$$
$$+0.00833f^2v_c + 0.000028v_c^2 + 0.0003472fv_c^2 - 2.62091 \cdot 10^{-7}v_c^3 \quad (4)$$

## 3.2. Dry turning process

The temperature tests in the cutting zone were to verify whether the increase in cutting speed and feed rate resulted in higher temperatures in the chip forming zone. Too high temperature may lead to unfavorable phenomena in the structure of the surface layer and limit the use of a polycrystalline diamond tool due to the possibility of graphitization. Fig. 7 shows the results of the temperature measurement and the plot of the regression function.
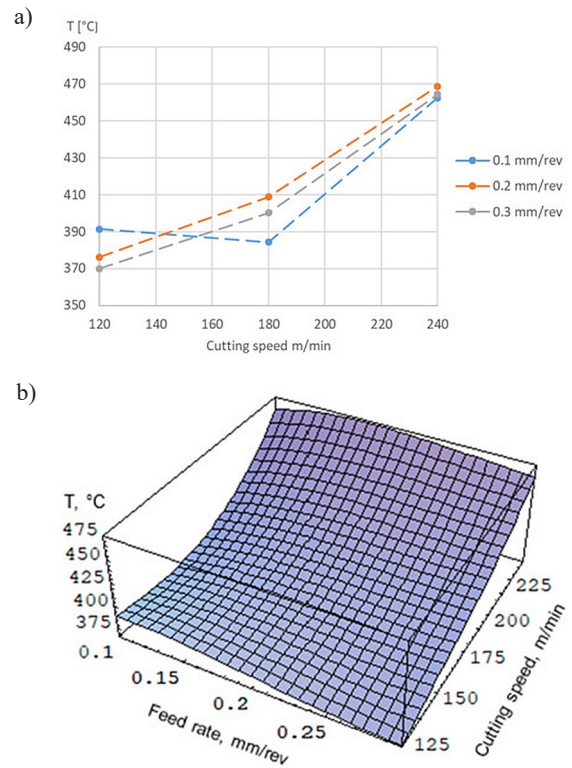


*Fig. 7. The influence of cutting parameters on the temperature value; a) test results, b) plot of the regression function*

The adequate equation of the regression function, statistically verified at the significance level of α= 0.05, describing the influence of the cutting process parameters on the temperature value, takes the form (5):

$$T = 540.887 - 873.187f - 3056.4f^2 + 6563.45f^3 - 1.37693v_c + 16.95fv_c. \quad (5)$$

At a feed rate of f = 0.1 mm/ rev, an increase in the cutting speed from $v_c$ = 120 m/ min to $v_c$ = 180 m/ min causes a slight decrease in temperature by 1.79%. The temperature in the considered range of cutting speed changes seems to be stabilized much below the value at which the graphitization phenomenon occurs. A further increase in cutting speed causes the temperature to rise sharply by 20.39% to 462°C. During the machining of the shaft with both the feed rate f = 0.2 mm/ rev and f = 0.3 mm/ rev, a similar trend of temperature increase was observed with the increase of the cutting speed.

On the basis of the final turning tests of the Ti-6Al-4V alloy with the use of a polycrystalline diamond cutting insert, the influence of the cutting parameters on the selected parameters of the surface topography was also determined. As in the case of turning with tool cooling, the Sa parameter is also strongly correlated with the Sz parameter. Pearson's correlation coefficient is r = 0.999. Therefore, in the further part of the analysis, the regression equation was estimated also only for the Sa parameter.
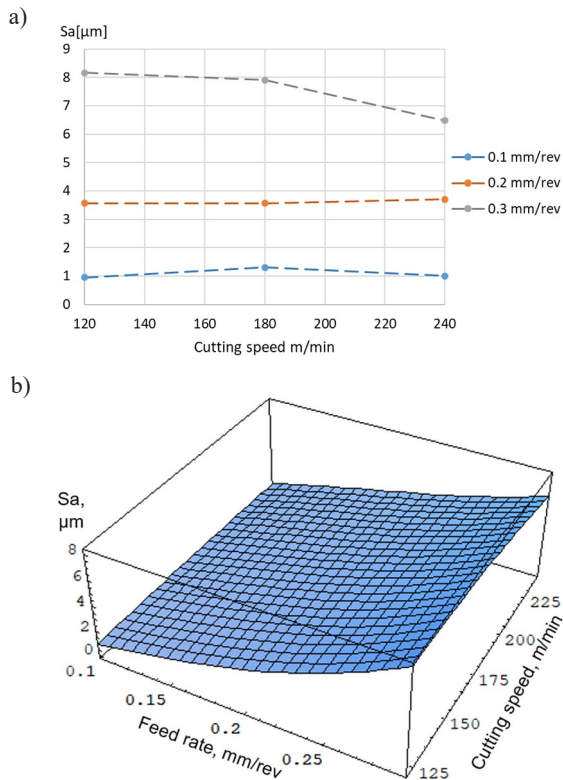
a)

b)

Fig. 8. Dependence of the Sa parameter on the cutting speed and feed rate a) test results, b) plot of the regression function

Table 4. Surface topography parameters in accordance with the methodology of the RSM

| Lp. | Cutting speed, m/min | Feed rate, mm/rev | Sa, µm | Sz, µm |
|-----|----------------------|-------------------|--------|--------|
| 1 | 120 | 0.1 | 1.33 | 12.37 |
| 2 | 240 | 0.1 | 1.12 | 9.59 |
| 3 | 120 | 0.3 | 6.56 | 32.04 |
| 4 | 240 | 0.3 | 6.91 | 31.18 |
| 5 | 120 | 0.2 | 3.24 | 17.89 |
| 6 | 240 | 0.2 | 3.3 | 20.35 |
| 7 | 180 | 0.1 | 1.63 | 12.56 |
| 8 | 180 | 0.3 | 6.89 | 35.99 |
| 9 | 180 | 0.2 | 3.18 | 18.78 |
| 10 | 180 | 0.2 | 3.01 | 18.90 |

The adequate equation of the regression function, statistically verified at the significance level of $\alpha = 0.05$, describing the influence of the cutting process parameters on the temperature value, takes the form (6):

$$S_a = 0.44594 - 13.4516f + 9.33451f^2 + 340.137f^3 - 0.009491v_c + 0.37375fv_c$$
$$0.795833f^2v_c - 0.000035v_c^2 - 0.000354v_c^2 + 5.79579 \cdot 10^{-8}v_c^3 \quad (6)$$

When analyzing the results of roughness measurements during dry turning and with cooling through the tool, similar trends in the roughness change depending on the cutting parameters can be noticed (Fig. 8a).In both the first and the second case, when turning with a feed rate of f = 0.1 mm/ rev, an initial increase in the value of the Sa parameter can be observed, which decreases after exceeding the cutting speed of $v_c$ = 180 m/ min. In the case of turning with a feed rate
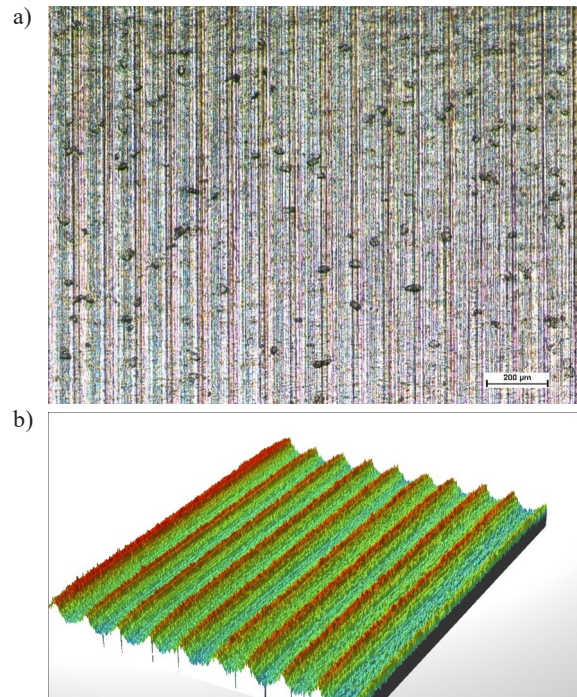
a)



b)



Fig. 9. Surface obtained after dry turning at the cutting speed $v_c$ = 120m / min and feed rate f = 0.3 mm/ rev; a) view of surface, b) three-dimensional surface topography

of f = 0.3mm/ rev, the value of the Sa parameter remains at a similar level in the cutting speed range of $v_c$ = 120÷ 180m/ min, and then it rapidly decreases. It should also be noted that the increase in feed rate and the associated temperature rise leads to the appearance of a chemical reaction causing the release of a compound that adhesively adheres to the surface. On an optical microscope image, it takes the form of small spots, 25 ÷ 40 µm wide (Fig. 9a). The thickness of the layers of the adhering compound are so small that it is difficult to observe them on the three-dimensional surface topography (Fig. 9b). When measuring surface topography, they can be considered noise and filtered out. Their removal is extremely difficult and requires time-consuming and costly technological operations. Inaccurate cleaning of the treated surface causes a change in the properties of the top layer and deterioration of the tribological properties of the parts.

Figure 10 shows a photo of the surface of the part made with the use of a scanning electron microscopefor feed rate f = 0.3 mm/rev and $v_c$ = 120 m/min. Figure 11 shows the chemical composition of the Ti6Al4V titanium alloy (11a) and the compound adhesively adhering to the surface (11b). Analysis of the chemical composition shows that the adhesive compound is formed as a result of a thermal chemical re-
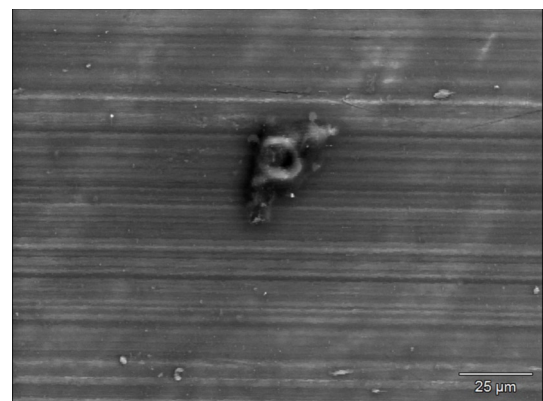


Fig. 10. Scanning electron microscope photo of the surface after dry turning for feed rate f = 0.3 mm/rev and $v_c$ = 120 m/min

action of the residual cooling lubricant used during roughing. The use of a liquid at this stage of treatment is necessary due to the tendency of the Ti6Al4V titanium alloy to harden the surface layer. Removal of the remaining cooling lubricant is very difficult and expensive, and the process itself does not guarantee the possibility of cleaning the surface of the remaining chemicals. It should be noted here that the chemical reaction only takes place at increased feed rate and at elevated temperatures. In the case of turning with a low feed rate value, no precipitated chemical compounds were observed on the surface of the part. This opens the possibility of machining without the use of a cooling lubricant, in an ecological manner, while ensuring a surface roughness similar to that in the case of using a coolant, and with the appropriate selection of cutting parameters, even smaller.
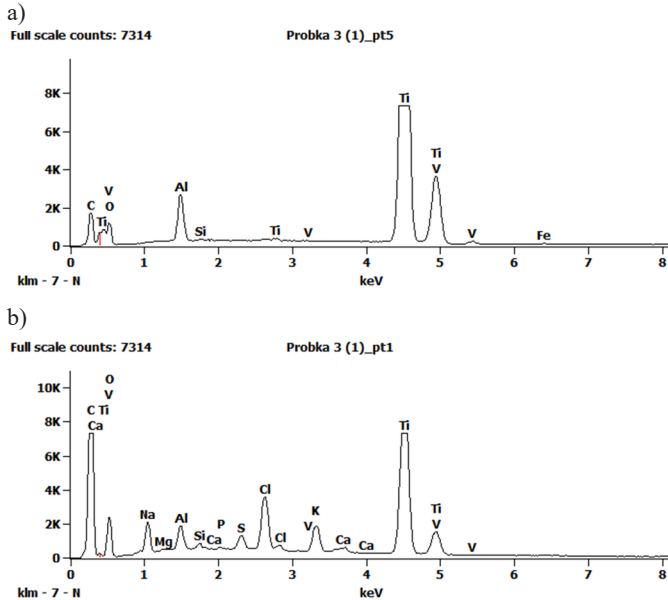
a)



b)



Fig. 11. Chemical composition of a) titanium alloy Ti6Al4V, b) compound adhesively adhering to the surface

## 4. Multi-criteria optimization

Due to the physical properties of the Ti6Al4V titanium alloy, the hardening of the surface layer as a result of permanent deformation and low thermal conductivity, finishing turning of the Ti-6Al-4V alloy with the use of a polycrystalline diamond cutting insert requires the selection of optimal process parameters. This selection must guarantee not only the required parameters of the top layer, but also high stability and efficiency of the process. This requires multi-criteria optimization of the process and finding a compromise solution that meets the above presented conditions. In order to write the multi-criteria problem, the following designations were adopted:

- $E \subset R^m$, a set of permissible solutions (a range of process setting parameters);

- $z = (z_1, z_2, \ldots, z_m) \in E$, acceptable solution;
- $f_i : E \to R$, $i$-th objective function ($i = 1, 2, \ldots, k$);
- $(z) = (f_1(z), f_2(z))$, theobjective function for a multi-criteria problem.

The problem of multi-criteria optimization of the selection of turning process parameters with tool cooling can be written in the form of (7):

$$\begin{cases} f_1(z) = f \to max, \\ f_2(z) = v_c \to max, \\ f_3(z) = F_c \to min \\ f_4(z) = S_a \to min \\ \quad z \in E \end{cases} \tag{7}$$

However, in the case of dry turning (8):

$$\begin{cases} f_1(z) = f \to min, \\ f_2(z) = v_c \to max, \\ f_3(z) = T \to min \\ f_4(z) = S_a \to min \\ \quad z \in E \end{cases} \tag{8}$$

One-criteria problem (9) is an $i$-th partial problem, where the vector $z^{io} \in E$, in which the $i$-th objective function achieves the extremum searched. The vector (10) is a vector called the ideal (utopian) solution in the space of evaluation, while (11) is the ideal solution to the function (7) or (8).

$$f_i(z) \to extremum, z \in E \tag{9}$$

$$\varphi^o = (f_1(z^{1o}), f_2(z^{2\,o})) \tag{10}$$

$$z^o = (z^{1o}, z^{2o}) \tag{11}$$

The set of effective solutions usually contains many solutions. Therefore, the aim of the presented problem was to select one compromise (optimal) solution from a set of effective solutions. For this purpose, functions (7) and (8) has been reduced to a single-criterion form, with the scalarisation function $s : R^k \to R$ in the form (12):

$$\max(s(f_1(z), f_2(z)) : z \in E) \tag{12}$$

Scalarisation of the function was carried out using the method of weighting the grades. Values of weights $u_i > 0$ of particular criteria $f_i$ (fulfilling the condition $u_1 + u_2 + u_3 + u_4 = 1$), were assumed, and then the optimal solution of the problem was determined (13).

$$\max(\sum_{i=1}^{k} u_i f_i(z) : z \in E) \tag{13}$$

It is only possible to create a function $\varphi(z) = \sum_{i=1}^{k} u_i f_i(z)$ if all the values of objective functions are expressed in the same units and scales. Since the objective functions in this case were expressed in different scales of values, they were transformed into a dimensionless form (14).

$$f_i^u(z) = \frac{f_k(z) - \min(f_k(x) : x \in E)}{\max(f_k(x) : x \in E) - \min(f_k(x) : x \in E)} \tag{14}$$

Objective functions $f_i^u$ takes values for $z \in E$ from the interval [0, 1] and they are dimensionless. After the unitarisation, an optimal solution of the problem is determined according to (15):

$$\max (\sum_{i=1}^{k} u_i f_i^u(z) : z \in E) \tag{15}$$

The optimal solution of the function (15) is an effective solution to the multi-criteria problem. The form of the solution depends on the weight values $u_i$ adopted. If during the calculations, it is assumed that all parameters for turning with cooling by the tool, are of equal importance ($u_1 = 0.25$, $u_2 = 0.25$, $u_3 = 0.25$, $u_4 = 0.25$), the objective function reaches the maximum value for the cutting speed $v_c = 236$ m/ min and

feed rate f = 0.132 mm / rev. This enables reach the surface roughness to be Sa = 1.66 µm with the cutting force $F_c$ = 77.29 N. Unfortunately, the obtained solution cannot be accepted due to the too high surface roughness inadequate for finishing turning. Therefore, during the further search for the solution to the problem (15), it was assumed that the surface roughness is of the greatest importance for the correct course of the process ($u_1$=0.2 , $u_2$=0.2, $u_3$=0.2, $u_4$=0.4). Such an assumption makes it possible to obtain a compromise solution with the cutting speed $v_c$= 240 m / min and the feed f = 0.1 mm / rev (point $P_1$ Fig. 12). Machining with these parameters makes it possible to obtain a surface roughness of Sa = 1.16 µm with the cutting force $F_c$ = 63.26 N.
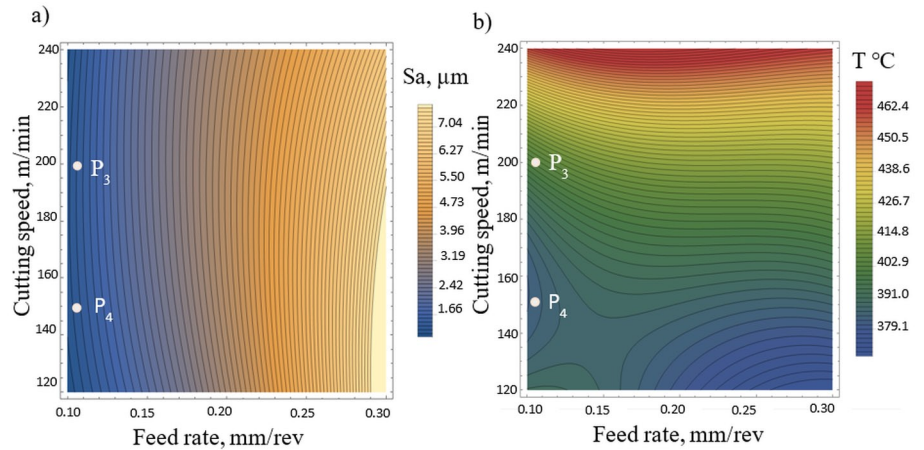


*Fig. 13. The results of solving the problem of multi-criteria optimization of dry turning a) contour plot of the regression function for Sa, b) contour plot of the regression function for T*
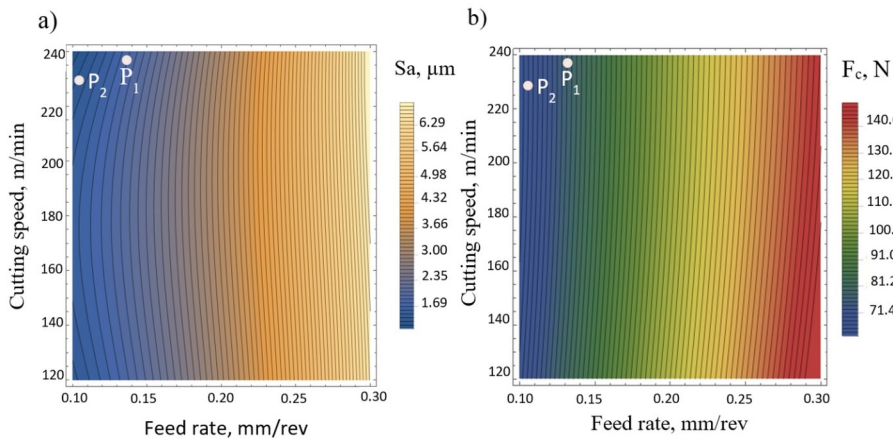


*Fig. 12. The results of solving the problem of multi-criteria optimization in turning with the tool cooling a) contour plot of the regression function for Sa, b) contour plot of the regression function for $F_c$*

• Surface roughness, both in turning with tool cooling and in dry turning, increases with increasing feed rate. During the tests, the lowest roughness value (Sa = 1.12 µm) in turning with cooling was obtained at a cutting speed $v_c$ = 240 m/ min and a feed rate f = 0.1 mm/ rev, while in dry turning at a cutting speed $v_c$ = 120 m/ min and a feed rate f = 0.1 mm / rev (Sa = 0.95 µm).

• The temperature in the cutting zone with a feed in the range of f = 0.2 ÷ 0.3 mm / rev increases continuously with increasing cutting speed. At a feed rate of f = 0.1 mm / rev, the increase in the cutting speed from $v_c$ = 120m / min to $v_c$ = 180 m / min doesn't cause a statistically significant increase in temperature.

• In dry turning, an increase in the feed rate leads to the appearance of a chemical reaction causing the release of a compound that adheres adhesively to the surface of the part. This phenomenon is difficult to observe when measuring roughness with a contact and optical profilometer, which leads to deterioration of the performance of the part.

• Multi-criteria optimization of the turning process enables to obtain compromise (optimal) solutions. In the case of turning with cooling by the tool, the use of a cutting speed of $v_c$ = 230 m/ min and a feed rate of f = 0.1 mm / rev allows to obtain a surface roughness of Sa = 1.16 µm with the cutting force ($F_c$ = 77.29 N), which does not significantly affect the accuracy of the part. During dry turning, a cutting speed of $v_c$ = 150m / min and a feed rate of f = 0.1mm / rev enables a surface roughness of Sa = 1.09 µm at a chip formation temperature of 382.3° C to prevent the graphitization process.

In the case of dry turning, assuming the same significance of the parameters ($u_1$=0.25, $u_2$=0.25, $u_3$=0.25, $u_4$=0.25) allows to obtain the optimal solution for the cutting speed $v_c$ = 198m / min and the feed rate of f = 0.1mm / rev (Point $P_3$, Fig.13). It is makes possible to obtain a surface roughness of Sa = 1.16 µm at a temperature in the chip area of 400 ° C, to prevent the graphitization process. After assuming that the most important parameters determining the structure of the surface layer of the processed material are the roughness Sa and temperature T, modifying the weight values ($u_1$=0.1 , $u_2$=0.1, $u_3$=0.4, $u_4$=0.4), it can be obtain another solution of the objective function (Point $P_4$ Fig. 13) ensuring surface roughness Sa = 1.09 µm at cutting speed $v_c$ = 150m / min and feed f = 0.1mm / rev and a slightly lower temperature with value T = 382.3 ° C.

## 5. Conclusions

This paper presents the analysis of turning process of Ti-6Al-4V titanium alloy with using a polycrystalline diamond tool. The research were conducted with using RSM method for dry and cooling turning. The machining experiment, model development and result analysis revealed the following conclusions:

## References

1. Abbas A.T., Sharma N., Anwar S., Luqman M., Tomaz I., Hegab H., Multi- response optimization in high-speed machining of Ti-6Al-4V using TOPSIS-fuzzy integrated approach. Materials 2020; 13 (5): 1104, https://doi.org/10.3390/ma13051104
2. ASTM B367-22 -Standard Specification for Titanium and Titanium Alloy Castings; ASTM International: West Conshohocken, PA, USA, 2022.
3. Ezugwu E.O., Bonney J., Da Silva R.B., Cakir O., Surface integrity of finished turned Ti6Al4V alloy with PCD tools using conventional and

high pressure coolant supplies. International Journal of Machine Tools and Manufacture 2007; 47 (6): 884-891, https://doi.org/10.1016/j.ijmachtools.2006.08.005

4.  Fan W., Shen W., Zhang Q., Alfredo H.-S. Ang.: A new response surface method based on the adaptive bivariate cut-HDMR. Engineering Computations 2021; 38 (3): 1402-1431, https://doi.org/10.1108/EC-06-2020-0343

5.  Gao Y., Wu Y., Xiao J., Lu D.: An experimental research on the machinability of a high temperature titanium alloy BTi-6431S in turning process. Manufacturing Rev. 2018; 5 (12): 1-7, https://doi.org/10.1051/mfreview/2018011

6.  Grzesik W, Niesłony P, Habrat W. Investigation of the tribological performance of AlTiN coated cutting tools in the machining of Ti6Al4V titanium alloy in terms of demanded tool life.Eksploatacja I Niezawodnosc- Maintenance and Reliability 2019; 21 (1): 153–158, http://dx.doi.org/10.17531/ein.2019.1.17

7.  Guo Z., Xu Ch., Wang X., Feng P., Zhang M. Determination of tool tip steady-state temperature in dry turning processbased on artificial neural network. Journal of Manufacturing Processes 2022; 79: 600-613, https://doi.org/10.1016/j.jmapro.2022.05.021

8.  ISO 25178-2 2012 Geometrical Product Specifications (GPS)- Surface texture: areal- Part 2: Terms, definitions and surface texture parameters (Geneva: International Organization for Standardization)

9.  Kechagias J. D., Aslani K. E., Fountas N. A., Vaxevanidis N. M., Manolakos D. E.. A comparative investigation of Taguchi and full factorial design formachinability prediction in turning of a titanium alloy. Measurement 2020; 151 107213, https://doi.org/10.1016/j.measurement.2019.107213

10. Khuri A.I.:A general overview of response surface methodology. Biometrics & Biostatistics International Journal 2017; 5(3): 87-93, DOI:10.15406/bbij.2017.05.00133

11. Kowalczyk M., Then chip compression ratio analysis in the aspect of TI-6AL-4V alloy turning with elevated cutting speeds. Czasopismotechniczne – Mechanika, 2012; 109 (8): 55-69.

12. Kuntoğlu M., Saglam H. ANOVA and fuzzy rule based evaluation and estimation of flank wear, temperature and acoustic emission in turning. CIRP Journal of Manufacturing Science and Technology 2021; 35 (3): 589-603,https://doi.org/10.1016/j.cirpj.2021.07.011

13. Lampropoulos A. D., Markopoulos A. P., Manolakos D. E. Modeling of Ti6Al4V Alloy Orthogonal Cutting with Smooth Particle Hydrodynamics: A Parametric Analysis on Formulation and Particle Density. Metals 2019; 9 (4), https://doi.org/10.3390/met9040388

14. Le Coz G., Fischer M., Piquard R., D'Acunto A., Laheurte P., Dudzinski D.: Micro cutting of Ti-6Al-4V parts produced by SLM Process. Procedia CIRP 2017; 58: 228 -232, https://doi.org/10.1016/j.procir.2017.03.326

15. Mia M., Khan M. A., Dhar N. R. Study of surface roughness and cutting forces using ANN, RSM and ANOVA in turning of Ti-6Al-4V under cryogenic jets applied at flank and rake faces of coated WC tool. The International Journal of Advanced Manufacturing Technology 2017; 93: 975–991, DOI 10.1007/s00170-017-0566-9

16. Mierzejewska Ż.,Kuptel P.,Sidun J. Analysis of the surface condition of removed bone implants. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18 (1): 65–72, http://dx.doi.org/10.17531/ein.2016.1.9.

17. Myers, R. H., Montgomery, D. C. Response Surface Methodology: Product and Process Optimization Using Designed Experiments. 2nd Edition, John Wiley & Sons, New York. 2002.

18. Narojczyk J., Moroz D., Siemiątkowski Z., Machining titanium alloy Ti-6Al-4V implanted carbide tools. Mechanik 2015; 3: 359-362, http://dx.doi.org/10.17814/mechanik.2015.3.152

19. Ranganath M.S., Vipin H., Optimization of process parameters in turning operation using response surface methodology: a review. International Journal of Emerging Technology and Advanced Engineering 2014; 4 (10): 351-360.

20. Sivam S. P. S. S., Rajendra Kumar S., Rajasekaran A., Karuppiah S. Prediction Model of Setting Input Parameters for Turning Operation TI-6AL-4V by Fuzzy Rule based Modeling. 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI) 2017; 1343-1349, doi:10.1109/ICPCSI.2017.8391929

21. Stachurski W., Midera S., D. Ostrowski. Influence of cutting parameters during turning process of aerospace industry alloy Ti-6Al-4V ELI (Grade 23) on cutting forces and surface roughness of the workpiece. Mechanik 2015; 8-9: 365-373, DOI: 10.17814/mechanik.2015.8-9.446

22. Stachurski W, Ostrowski D.: Influence of cutting parameters during turning process of aerospace industry alloy Ti-6Al-4V ELI (Grade 23) on cutting forces and surface roughness of the workpiece. Mechanik2016; 8-9: 1032-1033, http://dx.doi.org/10.17814/mechanik.2016.8-9.233

23. Sulaiman M.A. et al., Optimization of turning parameters for titanium alloy Ti-6Al-4V ELI using the response surface method (RSM). Journal of Advanced Manufacturing Technology(JAMT) 2013; 7 (2): 11-28.

24. Suresh R. Nipanikar: Effect of Process Parameters during Turning of Ti6Al4V-ELI in Dry and MQL Environments. International Journal of Engineering Research & Technology (IJERT) 2020; 8 (14): 190-194.

25. Surya M. S. Optimization of turning parameters while turning Ti-6Al-4V titanium alloy for surface roughness and material removal rate using response surface methodology. Materials Today: Proceedings 2022; 62: 3479-3484, https://doi.org/10.1016/j.matpr.2022.04.300

26. Surya M.S., Vepa K.S., Karanam M., Optimization of machining parameters using ANOVA and grey relational analysis while turning Aluminium 7075. International Journal of Recent Technology and Engineering 2019; 8 (2): 5682–5686, DOI: 10.35940/ijrte.B3038.078219

27. Tatar K., Sjoberg S., Andersson N.: Investigation of cutting conditions on tool life in shoulder milling of Ti6Al4V using PVD coated micro-grain carbide insert based on design of experiments. Heliyon 2020; 6: 1-7, https://doi.org/10.1016/j.heliyon.2020.e04217

28. Veiga C., Davim J. P., Loureiro A. Properties and applications of titanium alloys: A brief review. Reviews on Advanced Materials Science 2012; 32: 14-34.

29. Xiaobo Z. Comparison of response surface method and Kriging method for approximation modeling 2017; 2nd International Conference on Power and Renewable Energy (ICPRE),DOI:10.1109/ICPRE.2017.8390502

30. Ziberov M., Bacci da Silva M., Mark Jackson M., Wayne N.P. Hung. Effect of Cutting Fluid on Micromilling of Ti-6Al-4V Titanium Alloy. Procedia Manufacturing 2016; 5: 332-347, https://doi.org/10.1016/j.promfg.2016.08.029
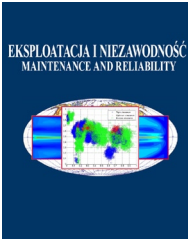
# Assessment of the modernized production system through selected TPM method indicators

**Indexed by:**
Web of Science Group

Edward Michlowicz[a]

[a] AGH University of Science and Technology, Faculty Mechanical Engineering and Robotics, al. Mickiewicza 30, 30-059 Kraków, Poland

## Highlights

- Modernization of a continuous production system.

- Proposed system evaluation methodology after the modernization.

- Evaluation criteria: selected TPM indices and MTBF and MTTR of production lines.

- Exemplification of the methodology for the zinc concentrate production system.

## Abstract

The subject of the studies is the evaluation of the operation of a production system after modernization. The analysed case concerns the modernization forced by the end of the product lifetime. The proposed methodology is that of a multicriterial evaluation of the system operation after modernization. The evaluation criteria are selected TPM indices: availability of machinery and equipment, production process capacity, product quality and overall equipment effectiveness (OEE). The additional criteria are reliability indices MTBF and MTTR of studied production lines and the MTTR of the most unreliable equipment in each analysed line. A yearly monitoring of production process was proposed for obtaining the statistical credibility of the evaluation results. Additionally, a fuzzy indicator of acceptability of the modernization assessment was proposed. The paper presents the results of studies of the system for production of zinc concentrate from post-production waste. The obtained values of OEE, MTBF and MTTR indicators for the three tested lines make it possible to state that the modernization carried out is acceptable.

## Keywords

continuous production system, system modernization, post-modernization evaluation, evaluation indices (OEE, MTTR, MTBF), process monitoring.

## 1. Introduction

Maintaining the continuous flow of materials and information is one of the most important tasks in production systems [3]. The main reasons for undesirable interruptions in systems operation are disruptions in supply of components and materials to relevant stations and downtime and failures of equipment in production lines. Numerous strategies and methods suggested by lean manufacturing and the operation and reliability theory are used to minimize such disruptions. The common assumption of the lean manufacturing-related methods is to ensure, maintain and improve the continuous flow of material in the production system [26, 27].

One of the methods to achieve this goal is to ensure the continuous operation of machines which is the main task of TPM (Total Productive Maintenance) [12, 13, 27]. Most publications on lean manufacturing include descriptions of various methods and options to improve the efficiency and productivity of production systems [1, 7, 15, 20]. The attempts to combine the lean manufacturing and the TMP into one consistent strategy of lean maintenance are increasingly often made [13, 22, 23]. An interesting review of literature on lean manufacturing can be found in [24, 25].

The analyses indicate that the lean monitoring level moves from the process evaluation to the company level. Conclusions presented in [29, 30] are even more far-reaching as the authors claim that the scope of lean evaluation has expended from the production process level to the supply chain level. Other methods are also suggested for a more detailed evaluation of the impact of the lean criteria on the leanness of processes, such as ANP (Analytic Network Process) [33], artificial intelligence methods [2], hybrid methods [34], and machine learning methods [4].

Another approach to improve the systems operation is Reconfigurable Manufacturing System (RMS). The principles of designing and the review of the RMS are presented in [14]. The selection of the production process in terms of maintaining the availability of the machines in the system is presented in [9]. Interested results of studies in Portuguese industrial production companies on the reconfigurability in are presented in [19].

Very significant is also the area of research on the improvement of the continuity of manufacturing processes by using the solutions from the operation and reliability theory. A systemic approach to the issues of prevention and predictable is presented in the extensive paper [36]. The states of production equipment capacity and states of quality of

manufactured products are important indices of the operational system evaluation, so the predictive strategy for multistate systems is very interesting [10]. The chain QR oriented to quality Q and reliability R is formulated in the strategy. An interesting paper [6] presents an integrated problem of choosing the production lot size, quality control and state-based maintenance for an imperfect production system which is subject to the reliability degradation [32]. The paper [32] includes a review of multicriteria models for solving the maintenance optimization problems. The review has identified 259 publication from the MCO (Multi Criteria Optimization) area and more than 100 universally used criteria.

An innovative predictive strategy for repairable complex systems is presented in [32]. The proposed long-term strategy includes the choice of degradation features and modules of the forecasting degradation models which allow obtaining accurate failure forecasts. Original solutions for the optimization process with redundancy with limitations with the use of an innovative algorithmic approach are presented in [11, 17]. Other analyses related to modelling and optimization of m-out-of-n backup systems are included [16, 18, 28].

There are a few papers on the evaluation of modernized systems. The issues of production process modernization in Russian industrial companies are presented in [35]. The conclusion presents a model of interrelation between the production modernization and the sustainable growth of a company. Finding the optimal design by a multicriterial evaluation is described in [31]. The evaluation uses a new measure of operational complexity of individual machines based on the number of parts, machines and operations [21]. Alternative design solutions are compared with each other using selected capacity criteria, followed by a multicriterial decision-making analysis based on the Analytic Hierarchy Process (AHP).

The review indicates only very few publications related to the use of TPM in evaluation of production systems after modernization. The TPM method is most often used to improve the efficiency of the operating systems. A novelty of the proposed is taking into account the evaluation method at the stage of making the decision to modernize. This is not an evaluation of the system modernization design, but an evaluation of the operation of the modernized system after a specific, longer time of operation (e.g. a year).

## 2. Model of a generalized production system

The generalized production system GPS is a certain ordered set of elements A and relations R between them:

$$GPS = \langle \{X, Y, T\}, R \rangle,$$
$$T : X \to Y.$$

where:

$X = \{X_1, X_2, ...X_i, ...X_M\}$; for $i = 1, ...M$ – set of external magnitudes describing input elements,

$Y = \{Y_1, Y_2, ... Y_j, ...Y_N\}$; for $j = 1, ...N$ – set of external magnitudes describing output elements,

$T = \{T_1, T_2, ... T_k, ...T_S\}$; for $k = 1, ...S$ – set of magnitudes describing the transformation of input vector into output vector,

$R = R_X \times R_Y \times R_T$ – material and information conjugations between the USP system elements and between the elements and the environment (most often close environment).

The diagram of the generalized production system is shown in Figure 1. Each system and system products have a specific, finite life, the so-called lifecycle. The continuous monitoring of selected features and system properties (acc. to Fig. 1) allows making the right decision at the right time – that is before the fourth lifecycle phase (decline of performance). In order to avoid the decision to decommission the system it is necessary to prepare the system modernization in advance (of course if it is possible and reasonable).

The reasons to make a decision to modernize the production system are usually:

- unsatisfactory economic indices,
- desire to modernize machine park,
- system adaptation to the requirements of Industry 4.0,
- approaching the fourth lifecycle phase (decline),
- limitation of availability for necessary raw materials and components,
- adaptation to changing environment requirements (e.g. EU directives), for instance in terms of environmental protection.
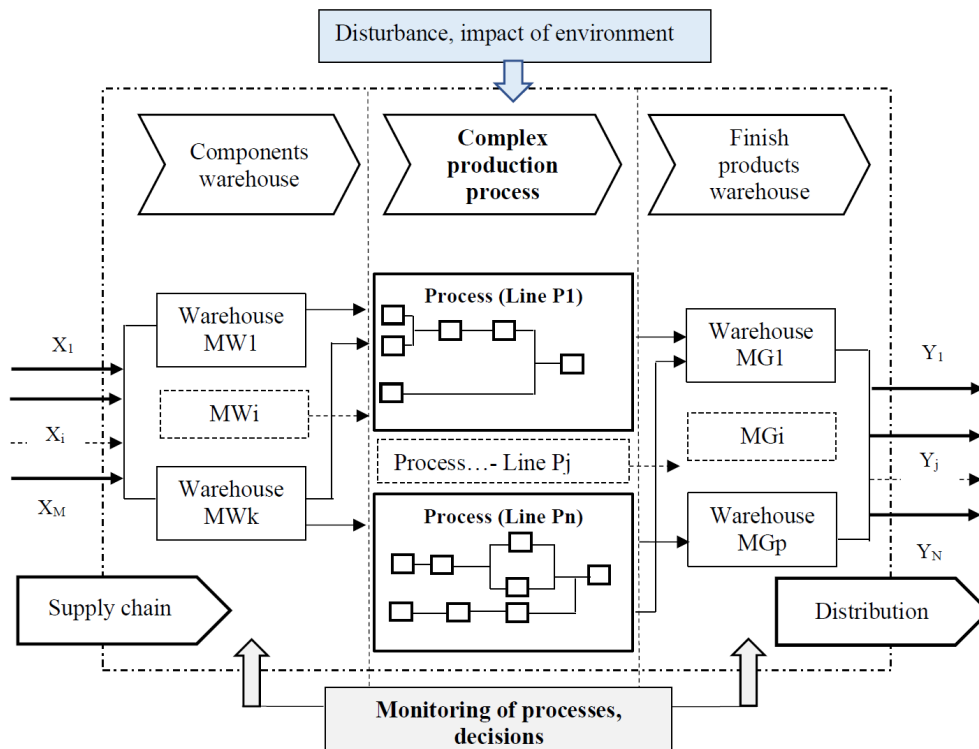


*Fig. 1. Diagram of a generalized production systems GPS (own study)*

The values determining thee specific character of the studied system include:
- continuous production (24 hours a day all year long),
- adaptation to changing environment protection requirements (e.g. EU directives),
- very significant limitation of supplies caused by exhaustion of current resources necessary for the production process,
- change of suppliers and components for production (an effect of the previous limitation).

The aforementioned reasons became an origin of the method for the evaluation of the production system modernization. The evaluation criteria are selected indices known from the TPM (Total Productive Maintenance).

Hence, the magnitudes describing the system outputs (acc. to Fig. 1) should include information typical to the evaluation of production systems (KPI – Key Performance Indicators) – about the process costs, achieved capacity, profitability, and also additional information on:
- availability of machines and equipment of production lines;
- product quality;
- mean time between failures;
- mean time to repair.

## 3. System modernization evaluation method

It was decided to modernize the system in a few stages:

**Stage I** – decision to modernize,

**Stage II** – system modernization (implementation),

**Stage III** – evaluation of production system after modernization.

Stage I should include a detailed identification of reasons for modernization. The identification result is the basis for three main tasks of the stage:
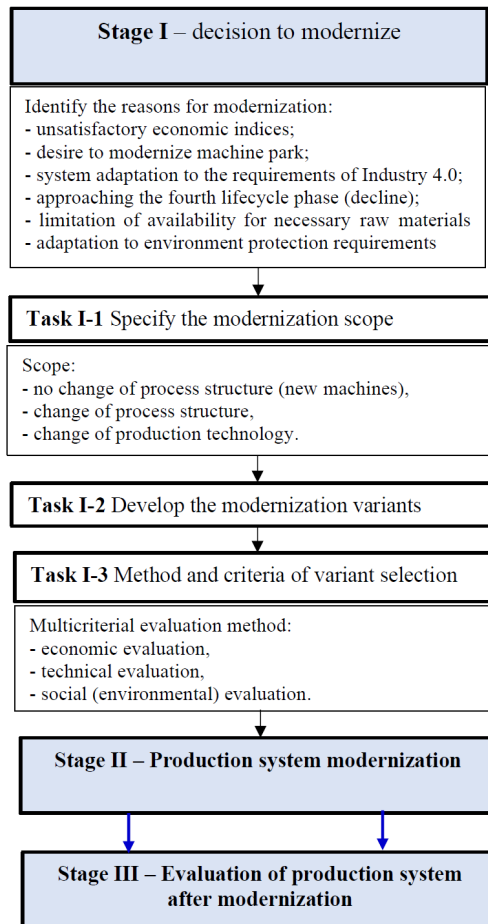


*Fig. 2. Diagram of variant selection algorithm (own study)*

- determine the scope of modernization;
- develop the variants of modernization;
- specify the criteria and method of selecting the variant.

A simplified variant selection algorithm is presented in Figure 2.

After the production system modernization (Stage II), it is necessary to evaluate the system operation.

The most important tasks in Stage III include:
- choose the post-modernization system evaluation criteria,
- monitor the process and collect the data about the process (over a longer time),
- process the data statistically,
- calculate the indices chosen for evaluation,
- analyse the results and make relevant decisions on further operation of the production line.

The post-modernization system evaluation algorithm is presented in Figure 3.



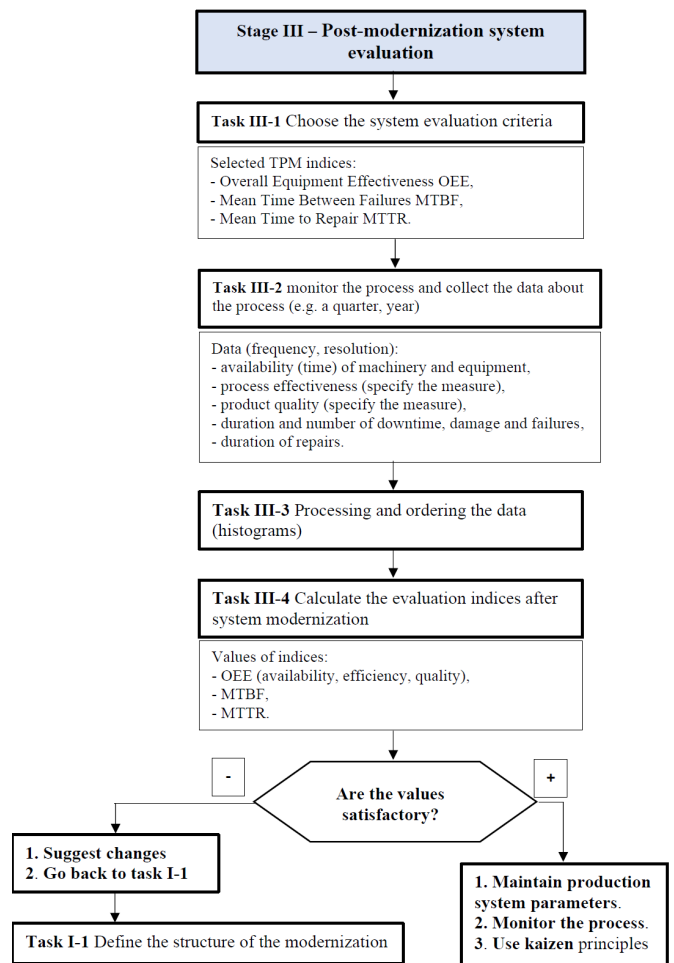*Fig. 3. Post-modernization system evaluation algorithm (own study)*

In terms of chosen evaluation criteria, the following conditions should be satisfied:
- overall equipment effectiveness OEE of the studies line PLi

$$OEE_{PLi} \geq OEE_{min} \; (eg.\, 0,80),$$

- availability of machines

$$A_{PLi} \geq A_{min} \; (eg.\, 0,90),$$

- process efficiency index

$$P_{PLi} \geq P_{min} \; (eg.\, 0,90),$$

- product quality index

$$Q_{PLi} \geq Q_{min} \ (eg. 0,95),$$

– mean time between failures

$$MTBF_{PLi} \geq MTBF_{min} \ (eg. \ TP \ hours),$$

– mean time to repair

$$MTTR_{PLi} \leq MTTR_{max} \ (eg. \ TN \ hours).$$

An important problem in this evaluation approach is a correct and accurate choice of limit, acceptable indices. Such choice can be based on the experience of engineering staff, expert knowledge, benchmarking, and also the knowledge of specificity of technological processes. Hence, the determination of an unambiguous final evaluation of the system after modernization is a complex problem. For example, in production process involving the use of post-production waste it is difficult to achieve the high product quality. For the purposes of the systems considered in the work, a simple fuzzy minimalist rule was proposed:

if

$$\bigwedge_{PLi} i = 1, .. n$$

$OEE_{PLi} \geq OEE_{min} \ \wedge \ MTBF_{PLi} \geq MTBF_{min} \ \wedge$

$MTTR_{PLi} \leq MTTR_{max} \rightarrow ACCEPTABLE \ ASSESSMENT$

In order to more accurately assess the operation of the system after the modernization, it is necessary to expand the proposed methodology using the principles of expert - fuzzy assessment.

## 4. Exemplification – zinc production from waste

### 4.1. System identification

The analysed company produces zinc concentrate necessary to make raw zinc. A rapid exhaustion of calamine resources forced a change of the processing technology in order to use another batch material. A decision has been made (Fig. 2) to make the zinc concentrate from zinciferous waste, particularly from dust from electric steel-melting shops and sludge from zinc electrolysis and industrial wastewater treatment plants. In order to prevent the contamination of the environment, such waste is subjected to zinc recovery during the pyrometallurgical processing in rolldown furnaces. This method is used all over the world.

The main problems linked to the system modernization included:

1. Ensuring continuous supply of new raw material (waste) which required a mechanism stimulating the feeding the system with waste generated by many suppliers.
2. Adaptation of the new technology to new environmental protection regulations, particularly new strict European regulations on allowed sulphur oxide emissions (IPPC Directive). The production from hazardous waste requires an Integrated Permit (PRTR - *Pollutant release and transfer registers* and EPER - European Pollutant Emission Register).
3. Achieving the process efficiency indices on European level.

As a result of modernization some lines were decommissioned, and the remaining lines received additional, new equipment. Three production line were modernized PL1, PL2 and PL3, which were analysed in detail.

The production process diagram is presented in Figure 4.

The assumption during the planning of modernization was that the production process efficiency will increase. Three basic characteris-
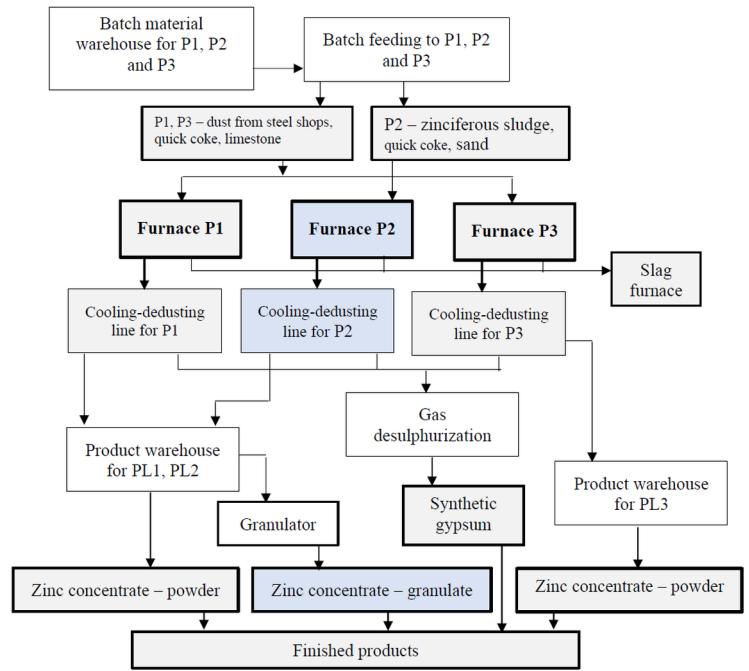


Fig. 4. Diagram of zinciferous waste processing line – after modernization (own study)

tics were used to determine the efficiency of utilization of resources after modernization:

– OEE - Overall Equipment Effectiveness,
– MTTR - Mean Time to Repair,
– MTBF - Mean Time Between Failures.

In accordance with the algorithm proposed during the modernization phase, a yearly monitoring of selected parameters was recommended (Fig. 3).

The data obtained for one year of operation were the basis of a detailed statistical analysis of:

– time between failures of individual lines and furnaces,
– time of failures, damage and micro-downtime of lines and the line equipment.

Histograms were made for three lines to show the system downtimes, with indication of the reasons (planned downtime/ failure). Examples of histograms for lines PL1 and PL2 are presented in figures 5 and 6.
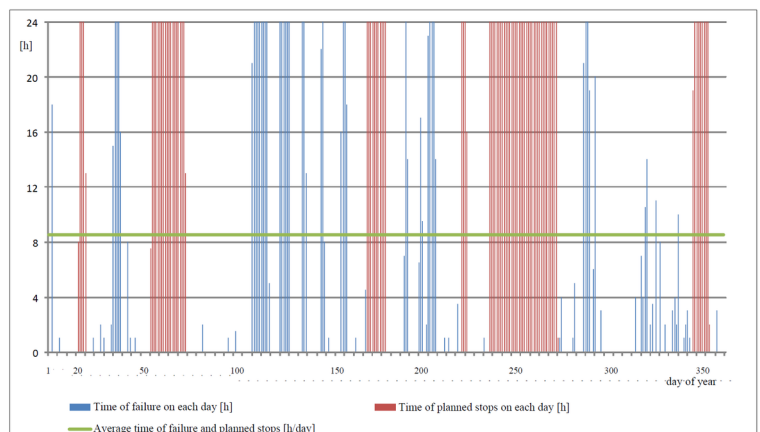


Fig. 5. Histogram of failures and planned downtime during one year for line PL1 (own study)

Figure7 presents the percent shares of times between failures, failures and planned downtime for line PL3.
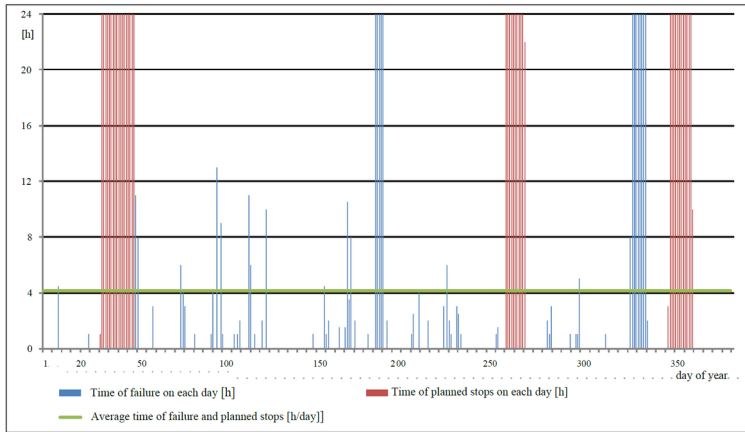
Fig. 6. Histogram of failures and planned downtime during one year for line PL2 (own study)
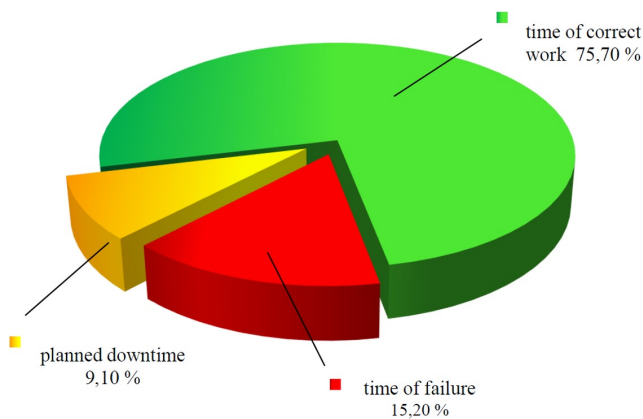


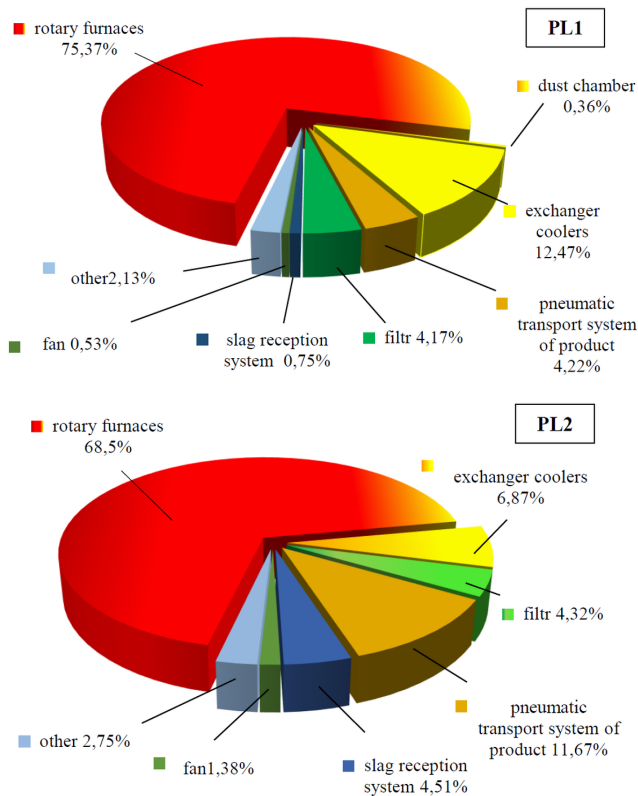Fig. 7. Availability of line PL3 in percent (own study)



Fig. 8. Percent shares of failures individual equipment in lines PL1 and PL2 (own study)

Due to the variety of failures of machinery and equipment in the process line, the downtimes were divided to the main sub-groups:
– slag reception systems,
– furnaces,
– dust chambers,
– heat exchanger coolers,
– filters,
– pneumatic product transport systems,
– fans.

Typical results for line PL1 are presented in Table 1.

The percent shares of failures of individual equipment in lines PL1 and PL2 are shown in Figure 8.

Table 1. Failure times for line PL1 by type

| Failure | time [h] | number |
|---|---|---|
| **Slag reception system** | | |
| removing lumps from rails | 1 | 1 |
| slag trap failure | 7,5 | 3 |
| **Furnace** | | |
| damping | 190 | 11 |
| slagging | 482 | 11 |
| heating | 159,5 | 11 |
| welding works on furnace end | 8 | 1 |
| problems with start-up –soft-start replacement | 2 | 1 |
| taring the mixture scale | 7,5 | 4 |
| Total: | 849 | 39 |
| **Dust chamber** | | |
| blocked screw feed under the chamber | 4 | 1 |
| **Heat exchanger coolers** | | |
| screw feeder failure | 22 | 2 |
| overhaul of screw under the hot cooler | 67 | 3 |
| cleaning of space between exchangers | 51,5 | 9 |
| **Filter** | | |
| high filter resistance – cleaning | 44 | 5 |
| filter check | 2 | 1 |
| failure od filter regeneration compressor | 1 | 1 |
| Pneumatic product transport system | | |
| oxide pump failure | 6 | 3 |
| failure of line-plate feeder | 41,5 | 10 |
| **Fan** | | |
| main fan cleaning | 1 | 1 |
| no compressed air | 5 | 1 |
| **Other** | | |
| no batch | 19 | 2 |
| defreezing of air for batch | 1 | 1 |
| no power on cooling-dedusting line 1 | 2 | 1 |
| failures of accompanying equipment | 2 | 2 |
| **Total events** | 1126,5 | 86 |

### 4.2. Analysis of results

The OEE index was determined in order to find the efficiency of use of resources using formulas (1), (2) and (3). The index defines the percent of theoretically achievable efficiency of line or equipment:

$$OEE = A \cdot P \cdot Q \cdot 100 \, [\%] \tag{1}$$

where:

A – production line availability, $A \in (0, 1)$,

P – process efficiency (performance), $P \in (0, 1)$,
Q – product quality, $Q \in (0, 1)$.

**Calculation of process availability index A**

$$A_{PLi} = \frac{A_{i2}}{A_{i1}} \cdot 100 \, [\%] \qquad (2)$$

$$A_{i1} = 365 \text{ days} \cdot 24 \text{ hours} - t_{ppi}$$

$$A_{i2} = 365 \text{ days} \cdot 24 \text{ hours} - t_{pp1} - t_{aw1}$$

where:

$A_{i1}$ [h] – planned work time (available time – planned downtime),

$A_{i2}$ [h] – operational time (planned work time – total downtime and failure time),

i – number of production line for which the index is calculated,

$t_{ppi}$ [h] – planned downtime of the ith production line,

$t_{awi}$ [h] – total failure time of the ith production line.

**Determination of process efficiency index P**

Two different approaches can be suggested for determination of the efficiency index [9].

On one hand, it can be assumed that if the furnace runs at the design capacity, and during the modernization all equipment and machines in the process line were designed for maximum design capacity of the rolldown furnace, the efficiency index for the whole process line can be taken as $P_{PLi} = 100 \%$.

On the other hand, due to the complexity of the process, difficulties in maintaining the technological discipline and varying quality

of batch materials, it is difficult to think that the capacity will always be kept at the highest level. It can be reduced, for instance, by growing deposits inside the furnaces. Due to the continuous character of the process, it is difficult to indicate accurate values of the efficiency index at a given time. Based on experience, it is however possible to assume that the reduction on the average should not exceed 5%. Such being the case, it is taken that $P_{PLi} = 95\%$.

**Calculation of quality index Q**

The quality index was determined by comparing the amount of batch material with the amount of concentrate received from it in relation to the zinc content in them, using the formula (3).

$$Q_{PLi} = \frac{Q_{i2}}{Q_{i1}} \cdot 100 \, [\%] \qquad (3)$$

where:

$Q_{i1}$ [Mg] – mass content of Zn in batch,
$Q_{i2}$ [Mg] – mass content of Zn in product.

The data from the year of production process monitoring in lines PL1, PL2 and PL3 are presented in Table 2.

The data on the mass zinc content in the batch material (dust from electric steel-melting shops and zinciferous sludge from zinc electrolysis) are presented in Table 3.

The calculations yielded the indices of availability, efficiency and quality. The results are presented in Table 4.

**MTBF and MTTF**

MTBF is the mean time between two failures or downtimes. The $MTBF_i$ was calculated according to the following formula:

Table 2. *Results of monitoring for production lines*

| | Production line PL1 | Production line PL2 | Production line PL3 |
|---|---|---|---|
| Available time [h/year] | 8760 | 8760 | 8760 |
| Planned downtime [h/year] | 1999,5 | 996 | 797 |
| Failure time [h/year] | 1126,5 | 509,5 | 1331,5 |
| Furnace failure time [h/year] | 849 | 349 | 994,5 |
| Time between failures [h/year] | 5634 | 7254,5 | 6631,5 |

Table 3. *Zinc concent in batch material and final product*

| Production line | | weight [Mg] | Zn content [%] | Zn amount [Mg] |
|---|---|---|---|---|
| Line **PL1** | Batch | 54 518,385 | 27,52 | 15 003,46 |
| | Production | 22 360,037 | 59,44 | 13 290,81 |
| Line **PL2** | Batch | 58 839,159 | 17,52 | 10 308,62 |
| | Production | 19 039,309 | 45,55 | 8 672,41 |
| Line **PL3** | Batch | 54 518,385 | 27,52 | 15 003,46 |
| | Production | 22 360,037 | 59,44 | 13 290,81 |

Table 4. *Partial indices for individual lines*

| Index | Production line PL1 | Production line PL2 | Production line PL3 |
|---|---|---|---|
| $A_{i1}$ [h/year]<br>$A_{i2}$ [h/year] | 6760,5<br>5634,0 | 7764,0<br>7254,5 | 7963,0<br>6631,5 |
| Availability – $A_{PLi}$ | 0,833 | 0,934 | 0,833 |
| Performance - $P_{PLi}$ | 0,95 | 0,95 | 0,95 |
| $Q_{i1}$ [h/year]<br>$Q_{i2}$ [h/year] | 15 003,46<br>13 290,81 | 10 308,62<br>8 672,41 | 15 003,46<br>13 290,81 |
| Quality – $Q_{PLi}$ | 0,886 | 0,841 | 0,886 |

$$\text{MTBF}_i = \frac{t_{ppri}}{n_{ppi}}$$

where:

$t_{ppri}$ – total correct operation time for the ith line [h]; i=1, 2, 3,
$n_{pp1}$ – number of evenings of correct operation of the ith line; i=1, 2, 3.

MTTR is the mean time needed to repair the equipment. Each "repair time" starts when the equipment fails and ends when the equipment starts to run according to its standard operation cycle. The MTTR is calculated according to the following formula:

$$\text{MTTR}_i = \frac{t_{awi}}{n_{ni}}$$

where:

$t_{awi}$ – total time of repair of the ith line [h]; i=1, 2, 3,
$n_{ni}$ – number of repairs of the ith line; i=1, 2, 3.

**Calculation of MTTR for furnaces P1, P2, P3**

The analyses indicate that the failures of furnaces take up the most failure time of the ith line. Consequently, this subassembly of each line was analysed in detail.

Calculation of MTTR for the ith furnace, i=1, 2, 3:

$$\text{MTTR}_{Pi} = \frac{t_{awPi}}{n_{nPi}}$$

where:

$t_{awPi}$ – total time of repair of furnace Pi [h],
$n_{nPi}$ – number of repairs of furnace Pi.

**Production line 1 - PL1**

MTBF for production line PL1:

$$\text{MTBF}_l = \frac{t_{ppr1}}{n_{pp1}} = \frac{5634}{45} = 125,2 \text{ h / year}$$

MTTR for production line PL1:

$$\text{MTTR}_1 = \frac{t_{aw1}}{n_{n1}} = \frac{1126,5}{53} = 21,25 \text{ h / year}$$

MTTR for furnace P1:

$$\text{MTTR}_{P1} = \frac{t_{awP1}}{n_{nP1}} = \frac{849}{39} = 21,77 \text{ h / year}$$

Similar calculations of MTBF and MTTR were made for production lines PL2 and PL3 and for furnaces P2 and P3.

Table 5 includes information obtained during the monitoring of the operation of these three furnaces.

Figure 9 presents the share of failures on individual furnace parts (for lines PL1 and PL2).
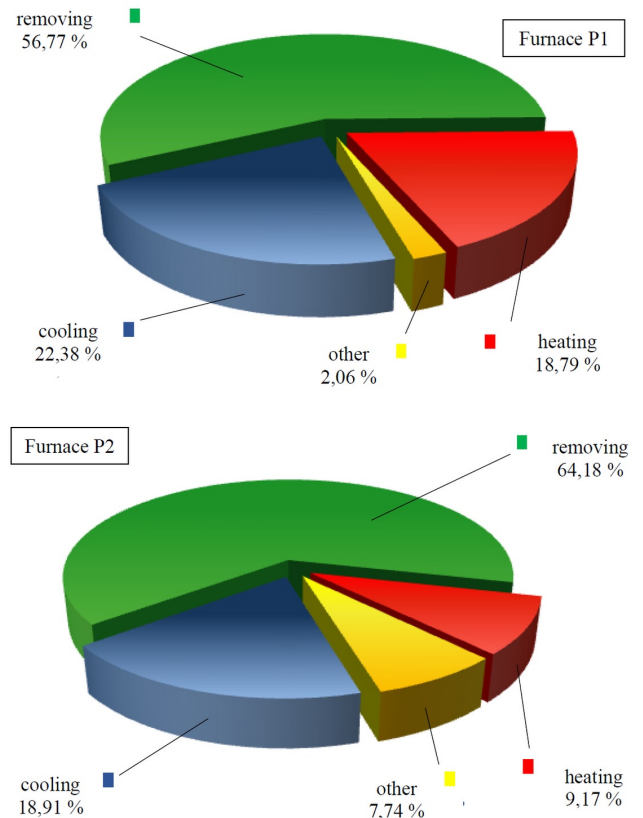


Fig. 9. *Percent share of time of individual failures for furnaces P1 and P2 (own study)*

Table 5. *Failure times of furnaces P1, P2, P3*

| | | | Line PL1 – Furnace P1 | Line PL2 – Furnace P2 | Line PL3 – Furnace P3 |
|---|---|---|---|---|---|
| Total failure time [h] | | | 1126,5 | 509,5 | 1331,5 |
| Furnace | | | **Furnace P1** | **Furnace P2** | **Furnace P3** |
| Components of the furnace repair time | cooling | [h] | 190 | 66 | 187 |
| | | [%] | 22,38 | 18,91 | 18,80 |
| | slagging | [h] | 482 | 224 | 629,5 |
| | | [%] | 56,71 | 64,18 | 63,30 |
| | heating | [h] | 159,5 | 32 | 156 |
| | | [%] | 18,79 | 9,17 | 15,68 |
| | other | [h] | 17,5 | 27 | 22 |
| | | [%] | 2,06 | 7,74 | 2,21 |
| | total | [h] | 849 | 349 | 994,5 |
| | | [%] | 75,37 | 68,50 | 74,69 |

The data from the one-year monitoring and the analyses allowed calculating the indices chosen for the system evaluation. The summary results of the studies of the production lines are presented in Table 6.

Diagrams (Fig.10, 11 and 12) present the sum may of indices calculated for three studied lines PL1, PL2 and PL3.

One of the conclusions is that the availability D as a parameter whose improvement should be a priority.

The failure times for individual lines differ – for lines LP1 and LP3 the times are significantly longer than for line LP2. The reason is the different technology. The line LP2 is Adapted only to the processing of zinciferous sludge, and the batch material for the remaining lines can also be the dust from electric steel-melting shops. Other physical

*Table 6.   Summary results for individual production lines*

|  | Production line PL1 | Production line PL2 | Production line PL3 |
|---|---|---|---|
| Failure time [h/year] | 1126,5 | 509,5 | 1331,5 |
| Failure time [%] | 12,86 | 5,82 | 15,20 |
| Planned downtime [h/year] | 1999,5 | 996 | 797 |
| Planned downtime [%] | 22,83 | 11,37 | 9,10 |
| Correct operation [h/year] | 5634 | 7254,5 | 6631,5 |
| Correct operation [%] | 64,31 | 82,81 | 75,70 |
| **OEE [%]** | **70,13** | **74,62** | **70,13** |
| **MTBF [h/year]** | **125,2** | **190,9** | **150,72** |
| **MTTR [h/year]** | **21,25** | **10,38** | **26,11** |
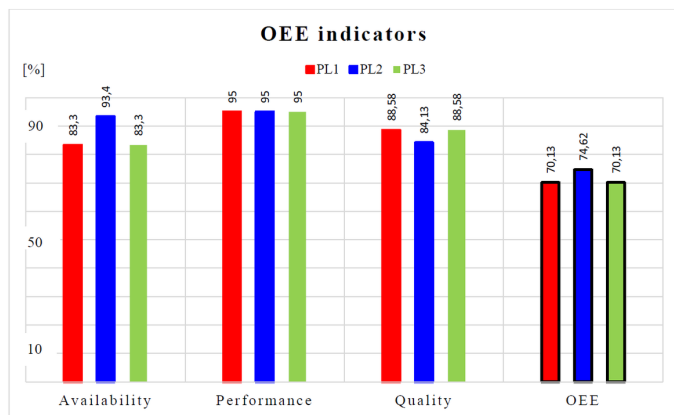


Fig. 10. Summary of OEE indices for studied lines (own study)

## 5.  Summary

The methodology proposed in the paper allowed a multicriterial analysis of the system operation after modernization. The analysis of the number and reasons of process lines downtimes allowed determining the limits of utilization of resources on the disposal of the company and determining the areas in which the improvements should made.



|  | PL1 | PL2 | PL3 |
|---|---|---|---|
| Availability | 83,3 | 93,4 | 83,3 |
| Performance | 95 | 95 | 95 |
| Quality | 88,58 | 84,13 | 88,58 |
| OEE | 70,13 | 74,62 | 70,13 |

Fig. 11. Component indices and OEE for studied lines (own study)
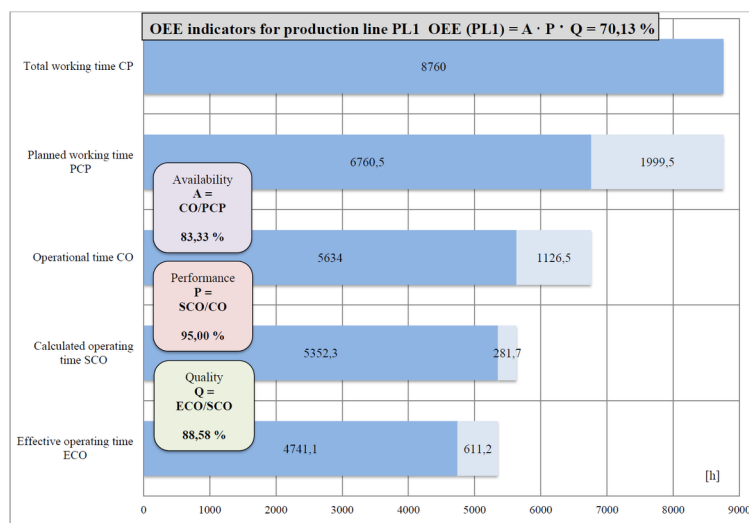


Fig. 12. Summary of OEE indices for production line PL1 (own study)

and chemical properties of the batch and somewhat different course of the process affect, inter alia, the growth rate of deposits inside the furnaces. The analysis proved that this is the most frequently occurring reason for the failures of these devices.

The assignment of the number and times of failures to individual subassemblies of modernized lines allowed indicating rotary furnaces as the most problematic components in terms of unplanned downtime. The failures of this equipment are responsible on the average for 72,83% of failure times on each line. The most time-consuming is the restoration of the equipment of equipment efficiency after the stoppage related to the occurrence of ring deposits. The average MTTR for furnaces is as high as 23,31 h/year. Consequently, technical solutions should be sought to limit the losses resulting from it. One of the suggestions is to use the Winchester industrial gun to speed up the deposit removal.

One of the effects of the technological process modernization was to obtain the European process effectiveness level. For systems producing products from production waste, OEE = 70% was assumed as a satisfactory OEE indicator. The measure of the achievement of this task was the OEE index for each studied production line.

The obtained results are satisfactory:

$OEE_{LP1} = 70,13\ \% > OEE_{min} = 70\ \%$,
$OEE_{LP2} = 74,62\ \% > OEE_{min} = 70\ \%$,
$OEE_{LP3} = 70,13\ \% > OEE_{min} = 70\ \%$.

At the same time, the limit values of average repair times and correct operation were adopted (expert knowledge):

MTTRmax = 30 h and MTBFmin = 125 h, which means that the indicators obtained for all three lines are within the adopted criteria (Tab. 6). Therefore, it can be concluded that the evaluation of the system operation after the modernization is acceptable.

The indices only slightly differ from the word level (75-80 %), which, taking into consideration the complexity of the analysed technological process, can be seen as a satisfactory result. This fact justifies the modernization and indicates that its effectiveness. At the same time, however, efforts should be made to improve the current state, particularly in terms of availability of production lines. the achieved availability level for lines LP1 i LP3 (83,3 %) requires a significant improvement.

## References

1. Albzeirat M.K., Hussain M.I., Ahmad R., Salahuddin A., Al-Saraireh F.M., Bin-Abdun N. Literature Review: Lean Manufacturing Assessment During the Time Period (2008-2017). International Journal of Engineering Management 2018; 2(2): 29-46. https://doi.org/10.11648/j.ijem.20180202.12

2. Antosz K., Pasko L., Gola A. The Use of Artificial Intelligence Methods to Assess the Effectiveness of Lean Maintenance Concept Implementation in Manufacturing Enterprises. Appl. Sci. 2020, 10: 7922. https://doi.org/10.3390/app10217922

3. Armstrong E. Productize: The Ultimate Guide to Turning Professional Services into Scalable Products. Vecteris: Cincinnati 2021.

4. Broek M., Teunter R., de Jonge B., Veldman J. Joint condition-based maintenance and condition-based production optimization. Reliability Engineering and System Safety 2021; 214: 107743, https://doi.org/10.1016/j.ress.2021.107743

5. Chen Ch., Wang C., Lu N., Jiang B., Xing Y. A data-driven predictive maintenance strategy based on accurate failure prognostics. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2021; 23 (2): 387-394, https://doi.org/10.17531/ein.2021.2.19

6. Cheng G.Q., Zhou B.H., Li L. Integrated production, quality control and condition-based maintenance for imperfect production systems. Reliability Engineering and System Safety 2018; 175: 251-264, https://doi.org/10.1016/j.ress.2018.03.025

7. Czerwinska K., Pacana A. Analysis of the implementation of the selected lean production method in the production company. Scientific Papers of Silesian University of Technology - Organization and Management Series 2019, 133: 43-54, https://doi.org/10.29119/1641-3466.2019.133.4

8. Evseenko S., Kupriyanov Y. Modernization of Production Planning Methodology in the Context of Virtualization and Increasing Multi-Agent Meta-Environment. Advances in Social Science, Education and Humanities Research 2020; 392: 84-87, https://doi.org/10.2991/assehr.k.200113.018

9. Gola A. Reliability analysis of reconfigurable manufacturing system structures using computer simulation methods. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (1): 90-102, https://doi.org/10.17531/ein.2019.1.11

10. Han X, Wang Z, Xie M. et al. Remaining useful life prediction and predictive maintenance strategies for multi-state manufacturing systems considering functional dependence. Reliability Engineering & System Safety 2021; 210: 107560, https://doi.org/10.1016/j.ress.2021.107560

11. Hashemi M., Asadi M., Zarezadeh S.. Optimal maintenance policies for coherent systems with multi-type components. Reliability Engineering and System Safety 2020; 195: 106674, https://doi.org/10.1016/j.ress.2019.106674

12. Jain A., Bhatti R., Singh H. OEE enhancement in SMEs through mobile maintenance: a TPM concept. International Journal of Quality & Reliability Management 2015; 32 (5): 503-516, https://doi.org/10.1108/IJQRM-05-2013-0088

13. Jain A., Bhatti R., Singh H. Total productive maintenance (TPM) implementation practice: A literature review and directions. International Journal of Lean Six Sigma 2014; 5 (3): 293-323, https://doi.org/10.1108/IJLSS-06-2013-0032

14. Koren Y., Gu X., Guo W. Reconfigurable manufacturing systems: Principles, design, and future trends. Frontiers of Mechanical Engineering 2018; 13(2): 121-136, https://doi.org/10.1007/s11465-018-0483-0

15. Krolczyk J., Legutko S., Szczepańska A. Value Stream Mapping as a tool for the optimization of production - case study. MATEC Web of Conferences 2017; 121: 02006, https://doi.org/10.1051/matecconf/201712102006

16. Levitin G, Xing L, Dai Y. Optimal operation and maintenance scheduling in m-out-n standby systems with reusable elements. Reliability Engineering & System Safety 2021; 211: 107582, https://doi.org/10.1016/j.ress.2021.107582

17. Levitin G., Finkelstein M., Li Y-F. Balancing mission success probability and risk of system loss by allocating redundancy in systems operating with a rescue option. Reliability Engineering and System Safety 2020; 195: 106694, https://doi.org/10.1016/j.ress.2019.106694

18. 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 Niezawodnosc - Maintenance and Reliability 2020; 22 (1): 170-178, https://doi.org/10.17531/ein.2020.1.20

19. Maganha I., Silva C., Ferreira L.M.D.F. Understanding reconfigurability of manfuacturing systems: An empirical analysis. Journal of Manufacturing Systems 2018; 48: 120-130, https://doi.org/10.1016/j.jmsy.2018.07.004

20. Michlowicz E.: Logistics engineering and Industry 4.0 and Digital Factory. Archives of Transport 2021; 57(1): 59-72, http://doi.org/10.5604/01.3001.0014.7484.

21. Modrak, V., Soltysova Z. Development of operational complexity measure for selection of optimal layout design alternative. Int. Journal Production Research. 2018, 56: 7280-7295. https://doi.org/10.1080/00207543.2018.1456696

22. Mostafa S.., Dumrak J., Soltan H. Lean maintenance roadmap. Procedia Manufacturing 2015; 2: 434 - 444, https://doi.org/10.1016/j.promfg.2015.07.076

23. Mouzani I., Bouami D. The Integration of Lean Manufacturing and Lean Maintenance to Improve Production Efficiency. International Journal of Mechanical and Production Engineering Research and Development 2019; 9(1): 593-604. https://doi.org/10.24247/ijmperdfeb201957

24. Narayanamurthy, G. and Gurumurthy, A. Systemic leanness: an index for facilitating continuous improvement of lean implementation. Journal of Manufacturing Technology Management 2016; 27 (8): 1014-1053. https://doi.org/10.1108/JMTM-04-2016-0047

25. Narayanamurthy, G. and Gurumurthy, A. Leanness assessment: a literature review. International Journal of Operations and Production Management 2016; 36 (10): 1115-1160. https://doi.org/10.1108/IJOPM-01-2015-0003

26. Nyhuis P., Wiendhal H-P. Fundamentals of production logistics. Theory, tools and applications. Springer - Verlag: Berlin Heidelberg 2009.

27. Pascal D. Lean Production Simplified: A Plain-Language Guide to the World's Most Powerful Production System. CRC Press Taylor & Francis Group: Boca Raton 2015.

28. Ruiz-Castro JE. A complex multi-state k-out-of-n: G system with preventive maintenance and loss of units. Reliability Engineering & System Safety 2020; 197: 106797, https://doi.org/10.1016/j.ress.2020.106797

29. Sangwa, N.R. and Sangwan, K.S. Development of an integrated performance measurement framework for lean organizations. Journal of Manufacturing Technology Management 2018; 29 (1): 41-84. https://doi.org/10.1108/JMTM-06-2017-0098

30. Sangwa, N.R. and Sangwan, K.S. Leanness assessment of organizational performance: a systematic literature review. Journal of Manufacturing Technology Management 2018; 29 (5): 768-788. https://doi.org/10.1108/JMTM-09-2017-0196

31. Soltysova Z., Modrak V., Nazarejova J. A Multi-Criteria Assessment of Manufacturing Cell Performance Using the AHP Method. Appl. Sci. 2022, 12: 854, http:// doi.org/10.3390/app12020854. https://doi.org/10.3390/app12020854

32. Syan C, Ramsoobag G. Maintenance applications of multi-criteria optimization: A review. Reliability Engineering & System Safety 2019; 190: 106520, https://doi.org/10.1016/j.ress.2019.106520

33. Tekez E.,Tasdeviren G. Measuring the influence values of lean criteria on leanness. Journal of Manufacturing Technology Management 2020; 31(7): 1391-1416, https://doi.org/10.1108/JMTM-09-2019-0321

34. Tiamaz Y., Souissi N. Classification of the lean implementation procedures for improving the business processes. 2018 International Conference on Intelligent Systems and Computer Vision (ISCV) 2018, pp. 1-6, https://doi.org/10.1109/ISACV.2018.8354019

35. Varkova N.Y. Modernisation of production as a factor of influence on economic stability of the industrial enterprise. SHS Web of Conferences 2017; 35: 01146, https://doi.org/10.1051/shsconf/20173501146

36. Werbińska-Wojciechowska S. Preventive Maintenance Models for Technical Systems. In: Technical System Maintenance: Delay-Time-Based Modelling. Cham: Springer International Publishing 2019, https://doi.org/10.1007/978-3-030-10788-8_2

# Tribological properties of the FeNi alloys electrodeposited with and without external magnetic field assistance

Indexed by:
Web of Science Group

Anna Maria Białostocka[a], Marcin Klekotka[b], Urszula Klekotka[c], Beata Kalska-Szostko[c]

[a] Bialystok University of Technology, Faculty of Electrical Engineering, Department of Electrotechnics, Power Electronics and Electrical Power Engineering, ul. Wiejska 45D, 15-351 Białystok, Poland
[b] Bialystok University of Technology, Faculty of Mechanical Engineering, Institute of Biomedical Engineering, ul. Wiejska 45C, 15-351 Białystok, Poland
[c] University of Bialystok, Faculty of Chemistry, Department of Physical Chemistry, ul. Ciołkowskiego 1K, 15-245 Białystok, Poland

## Highlights

- An effective method of the FeNi electrodeposition on two substrates is presented.

- The tribological properties of magneto-electrodeposited FeNi coatings were determined.

- The FeNi layers properties are influenced by an external magnetic field.

- The FeNi surface morphology is related to the Fe content and the type of the substrate.

## Abstract

The hereby work presents the tribological properties of the iron-nickel alloys and their dependence on the microstructure and thickness of the probes as well as the presence of an external magnetic field during the synthesis. Coatings were electroplated on the brass and copper metallic substrates using galvanostatic deposition in the same electrochemical bath condition (Fe and Ni sulfates) and the electric current density. The surface morphology of the films was observed by Scanning Electron Microscopy. The average composition of all FeNi coatings was measured using Energy Dispersive X-ray Spectroscopy. Tribo-mechanical properties such as microhardness, roughness, and friction coefficient were determined in the obtained structures. The morphology and tribologial properties of the FeNi coatings clearly depend on both the substrate (Cu, CuZn) itself and the presence of an external magnetic field (EMF) applied during the deposition process.

## Keywords

electrodeposition, magnetic fields, Cu and CuZn substrate, FeNi coatings, microhardness, roughness, friction, wear.

## 1. Introduction

There is a wide range of processes that can alter the morphology and properties of metals. This is one of the advantages of metals because processing itself may be the source of defects in the final product. Continuous monitoring leads to better materials synthesis methods, capable to achieve accurate control of the structure and atomic configuration of the crystals at the nanoscale [24, 31, 37]. The electrodeposition is successfully used during the production of the multi layers (ML) coatings because of its simplicity, low manufacturing cost, and versatility. Nowadays, this process is further developed to predict and tailor the functional properties as the result of a quantitative description of fundamental phenomena at the atomic scale. Electrodeposited metals and alloys characterize enhanced micromechanical properties in comparison to cast metals and alloys [9, 13, 32]. The variation of the parameters such as electrolyte composition, bath pH, temperature, agitation (the physiochemical conditions), current density, potential (the electric conditions), and magnetic field can alter the composition

and microstructure of the deposited layers. This, in turn, affects the tribological properties of the layers, which are extremely important for their subsequent use in many areas of industry and economy and is essential [5]. A good example is the increasing significant demands to control friction losses and decrease the wear of machine components. Process of the tribo-mechanical and physical properties tailoring, such as wear resistance, high-temperature corrosion protection, oxidation resistance, and lubrication properties of a metallic coating, allows forachieving many different application requirements [12, 13]. Pure metal (Ni or Fe) exhibits different specific properties than the alloy (FeNi) received after the manufacturing process [13]. Differences in their percentage composition result in potential changes in later applications. The FeNi coatings could be used for example as high-performance transformers cores, read/writer magnetic shield materials, magnetic actuators, composites molds/tooling, etc. [1, 2]. Mentioned ML coatings are investigated by many researchers and have attracted attention due to theirmechanical, electric, and magnetic properties [8,

13, 28, 32]. Alloying of the substrate surface is often implemented for improving the coating performance, especially to increase the efficiency of mechanical systems or reduce friction losses using advanced materials and surface technologies. Additionally could be applied low-quality substrate (nonmetallic or metallic material) and its cover may satisfy the user which reduces the cost of application. Direct current-coated materials have smaller grain sizes and less plastic deformation compared with pulse current. The environmental conditions and the material type directly influence the surface topography and thus the friction coefficient and the wear resistance of the obtained coatings [16]. The material of the substrate results in changesin the adhesive strength of the deposited films. This varies with the different film/substrate combinations in the result of the crystallographic coherency between them. The eutectic and the peritectic alloy systems are practically applied in the electronic industry. The adhesion mechanism in the electronic device has attracted much attention. It depends on the number of voids that are observed in the mentioned interface and causes a decrease in the adhesion. The high concentration of the hydrogen existing in the elemental depth profiles results in the exfoliation of the film. Different interfacial structures were considered by Okamoto, Wang, and Watanabe [19] as well as Nweze and Ekpunobi [18] or Wei at. al. In both articles,the substrate affects the crystal structure of the coatings and the crystallographic coherency itself. Additionally, the parameters of the electrodeposited layer are the function of the substrate thickness and its magnetic character [13, 14]. Gurrappa and Binder [11] showed an unbreakable link between specific free surface energy, adhesion energy, lattice orientation of the electrode surface, crystallographic lattice mismatch at the nucleus-substrate interface, and nanostructures of the electrodeposited coating. Its distribution depends on the nucleation and growth processes. The nucleation could be instantaneous or progressive. In the case of the first formation, the nuclei are increasing with time. During the progressive formation, nuclei gradually grow and overlap. The results of electrodeposition are strongly affected by the cathode surface and its characteristics. The growing near substrate layer filled up the holes and defects of the electrode surface. This yields the desired and oriented surface morphology on metallic or nonmetallic materials. In this case, surface engineering studies are very important to createa template (cathode surface) for specific applications [11, 32]. Some studies have shown a relationship between the refinement of the surface microstructure and the friction mechanical properties of the layers. Smoother surfaces give a lower friction coefficient [1, 2, 9]. Surface texturing i. e. generating a specific surface structure is one of the ways which can help with the reduction of the friction as well as surface coatings application or surface roughness improvement [25]. In the recent two decades, scientists developed the electrodeposition process beyond the current state of the art. Some of them (Fahidy, Aogaki) used the external magnetic field to extend the range of available properties in the context of magnetic, mechanical, or thermal applications of alloys and compounds. The others (Fritoceaux, Russo) changed the kind of substrate to check how it influences the properties of the coating/substrate system. All explorations have to give a response to the question of how the above act on the obtained coatings. Last three decades scientists reported that the applied external magnetic field (EMF, perpendicular, parallel) results in changes in the morphology and structure of the manufactured layers. Electrochemical reactions and magnetic field forces influence each other which causes the magnetohydrodynamic effect (MHD) creation. EMF induces the additional convection and reduces the diffusion layer (enhances mass transport). The conclusions of the research are as follows: i) parallel arrangement (II) of the field and electrode surface activates two-dimensional growth and smoother deposit; ii) perpendicular orientation (I_) a rougher growth respectively. The field configurations (II, I_) form the preferred crystal texture of the FeNi deposits [3, 36]. The use of electrochemical methods to modify the surface processes and materials modification itself are very important, especially when it comes to the needs of modern technologies. A lot of properties are exten-

sively investigated for this reason, including morphology, structural properties, hardness, corrosion, wear resistance, etc. Companies are intended in the development process of the nanostructured materials manufacturing which will play a significant role in their customer's market [11]. Nowadays, knowledge of the subject of magneto-electrodeposition plays a very important role in science and technology. An example could be the application in many industry branches like space research, military applications, security systems, high-density magnetic memories, navigation, medicine, etc. The aim of this paper was the experiments whose results allow the discussion about the morphological and crystallographic properties of the alloys deposited in the different physical conditions. What is new is looking for layers with properties that meet the needs of modern users. Currently, zinc can be found in the products of many industries and the economy. Among others, in medicine (bone repair material), in the automotive industry (anode in alkaline batteries). In many cases, Zn needs protection because it reacts with the surrounding environment (degradation, adhesion, corrosion resistance). For this reason, it is of interest to scientists from many research centers and still needs research to clarify many issues [27, 34].

## 2. Materials and methods

A direct current electrodeposition process was employed for the preparation of the FeNi layers. The platinum vertical plate (width-6 mm x height-5 mm x thickness-0,5 mm) was used as the anode, and electrochemically polished Cu and CuZn (Cu63Zn37 – manufacturer data) plates were used as the cathode (width-10 mm x height-20 mm x thickness-0,25 mm). The electrolyte for the preparation of the iron-nickel layer consisted of $FeSO_4 \cdot 7H_2O$ (7,5 g), $NiSO_4 \cdot 7H_2O$ (7,8 g),and $H_3BO_3$ (2 g) without the presence of additives. The whole process was performed galvanostatically using the potentiostat/galvanostat instrument (Matrix MPS-7163). The sets of two permanent magnets (width-75 mm x height-50 mm x thickness-10 mm)were placed in the especially designed laboratory stand around (parallel, perpendicular) the cathode surface. The magnetic field strength was measured with the gauge FH51 (Magnet-Physik) and ranged from 80 mT to 400 mT (with an accuracy of 2%) but near the electrode surface was distributed uniformly. The experiments were conducted at room temperature. Tribological analyses were carried out using a UMT TriboLabtribometer (Bruker, Billerica, MA, USA), with a ball-on-disc system under dry friction conditions and each of them took 300 s. The countersample was immobile 6 mm corrundum ball. The reciprocating motion of a sample with the amplitude of 500 μm was repeated 3 times, with the following parameters: Fn = 2 N, f = 10 Hz. The knowledge about the composition was obtained as the result of spot surface measurement made with an energy dispersive X-ray spectrometer integrated with the SEM. X-ray diffractometer with Mo Kα radiation ($\chi$= 0,713067 Å) provided information on the crystal structure characterization. The probes were cut and the thickness of the cross-sections was measured. The LEXT OLS 4000 Confocal Laser Scanning Microscope (CLSM, Olympus, Tokyo, Japan) with a 3D image feature was used to characterize the metallic coatings: thickness, volume and depth of wear tracks. The often used method of surface description reflects the two other parameters. The first of them is the skewness ($S_{sk}$) which is sensitive to occasional deep valleys or high hills. It measures the symmetry of the variation of a profile about its mean line and gives information about the number of hills or valleys on the surface. Zero skewness informed about symmetrical height distribution. Positive value testifies to the fact of the existence of fairly high spikes above a flatter average. A negative value describes filled valleys, and deep scratches in a smoother plateau. The second is the kurtosis ($S_{ku}$), which informs about the probability density sharpness of the profile. An increase in the $S_{sk}$ value results in an increasing trend in the value of the friction static coefficient. In the case of $S_{ku}$, we observe the opposite trend. Positive $S_{sk}$ shows surfaces with good adhesion resistance but negative leads to lower values [25]. Both mentioned variables should be

calculated to achieve a useful understanding of a composite surface [1, 2, 20, 25, 26, 29]. Roughness of the probes was measured with the use of the optical method (not as a classic linear profilometer) but it was shown as a section of the surface with marked roughness values. Microhardness was tested using the Vickers method with a load of 0,9807 N and repeated 10 times. The crystal sizes were calculated by the use of Powder Cell program which adapts experimental (powder diffraction patterns) and theoretical data automatically.

## 3. Results and Discussion

### 3.1. X-Ray Diffraction (XRD)

The crystalline structures of the alloys deposited on both substrates (Cu, CuZn) show a dependence in respect of the deposition time (Fig.1). After 900 s of the process, the FeNi coatings are dominated by the FeNi structure. Long-term deposition causes a mixture of FeNi and Fe phases with the primary occurrence of bcc-Fe which confirms the data presented in Fig.3 (showing the changes in the Fe content). The addition of zinc (37 %) to Cu, as it is in the brass substrate changes the crystal cells parameters by 2 percent (value of the crystal parameter). For Cu, $c$ is equal to 3,57 Å and for CuZn – 3,65 Å which are close to the theoretical values of $c$ = 3,581 Å in FeNi (JCPDS Card No. 47-1405) and $c$ = 3,647 Å in Fe (JCPDS Card No. 06-0696) respectively [21, 22]. The external magnetic field application influences the crystallites size. In both cases (Cu and CuZn) crystallites are larger when growing without EMF. In average respectively d=23,0 ± 1,0 nm in comparison of the EMF presence d=14,0 ± 1,0 nm (FeNi crystallites - Cu substrate and Fe – CuZn). Therefore EMF works as the grain refiner.

### 3.2. Scanning Electron Microscopy (SEM)

Fig.2 shows the morphology of the fabricated FeNi coatings which changes due to the presence and the orientation of the EMF obtained by scanning electron microscopy. The surface of the deposit becomes smooth and compact when no external magnetic field was applied. Resulted layers seem to grow following a progressive nucleation mechanism. The presence of the external magnetic field during the deposition process doesn't influence the nucleation mechanism itself but affects the subsequent growth of the objects. In this case (at the early stage of the deposition, 900 s), it leads to a preferential development of 3D forms into clusters on the 2D nucleation centers at fixed points (deterministic mode). Over the time (3600 s), the number of these places increases, and 3D growth centers are scattered over almost the entire surface at various stages of development and various sizes (stochastic mode). The coating on both substrates (Cu, CuZn) showed competitive growth of 2D and 3D forms.

The major effect of the magnetic field is known as the magnetohydrodynamic (MHD) effect which results in additional convection and therefore reduction of the diffusion layer thickness. This caused the enhanced mass transport to the existing deposit and increased
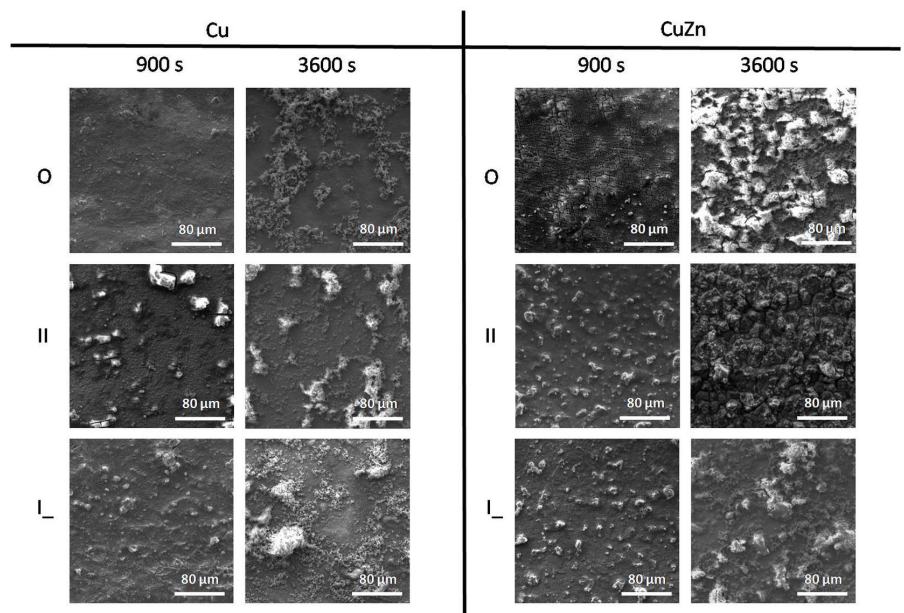


Fig.2. Set of SEM images of the deposited films with and without EMF: different time deposition (900, 3600 s); different substrate (Cu, CuZn); current density – 50 mA (cm²)⁻¹; without EMF (A), II EMF (B), I_ EMF (C)

the secondary nodules (second micro-MHD effect) [17]. This growth type appears as the result of instantaneous nucleation which characterizes a slow growth of the nuclei with a quite small number of active sites at the same moment [4]. Only the layer on CuZn substrate after II EMF application characterizes cracks and nodular shapes of the growing layer. This is directly connected with the appearance of high content of Ni (decrease of Fe content in time – Fig.3) and the increase of the hydrogen evolution rate [35]. This results in the grooves (among 3D nuclei) indented by MHD-flow and the hydrogen formation (reaction between Ni in alloy and Zn from the substrate) [17, 35].
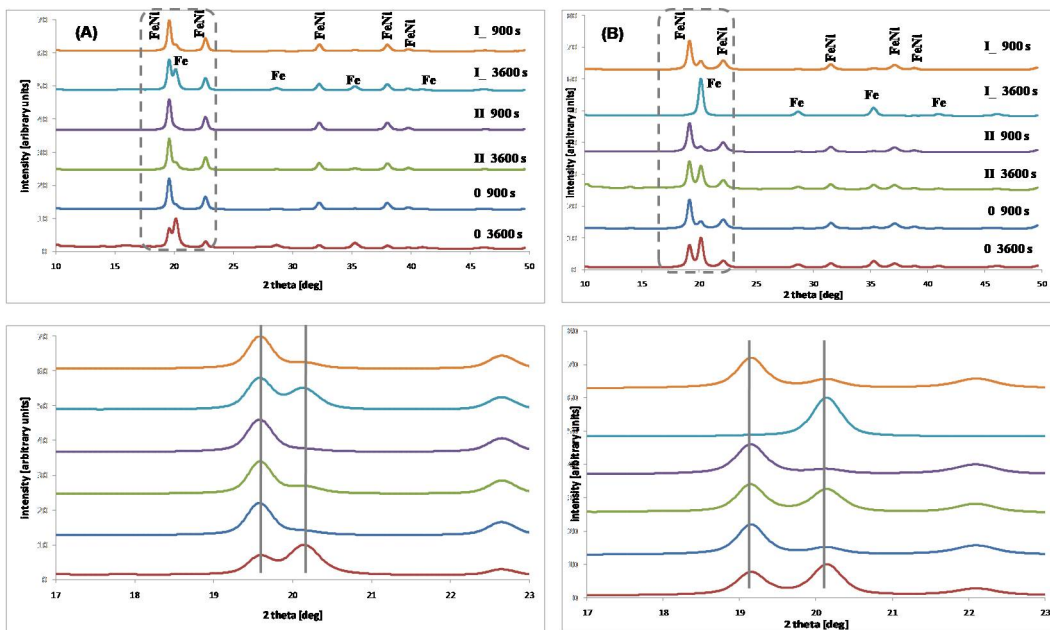


Fig.1. Selected XRD patterns are presented in panels concerning substrate surface: copper (A), brass (B), and time (900 s, 3600 s), respectively (0 – without EMF, II – parallel EMF, I_ - perpendicular EMF); current 50 mA (cm²)⁻¹
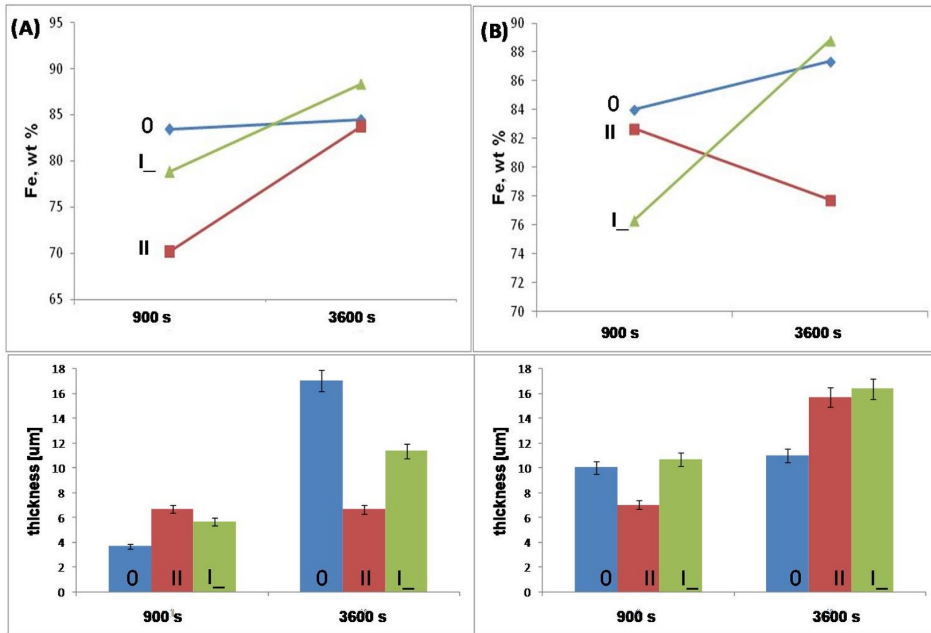
Fig. 3. Effect of Fe content (upper row) and FeNi film thickness (lower row) in time; (A) Cu substrate, (B) CuZn substrate; 0 – without EMF, II – parallel EMF, I_ - perpendicular EMF
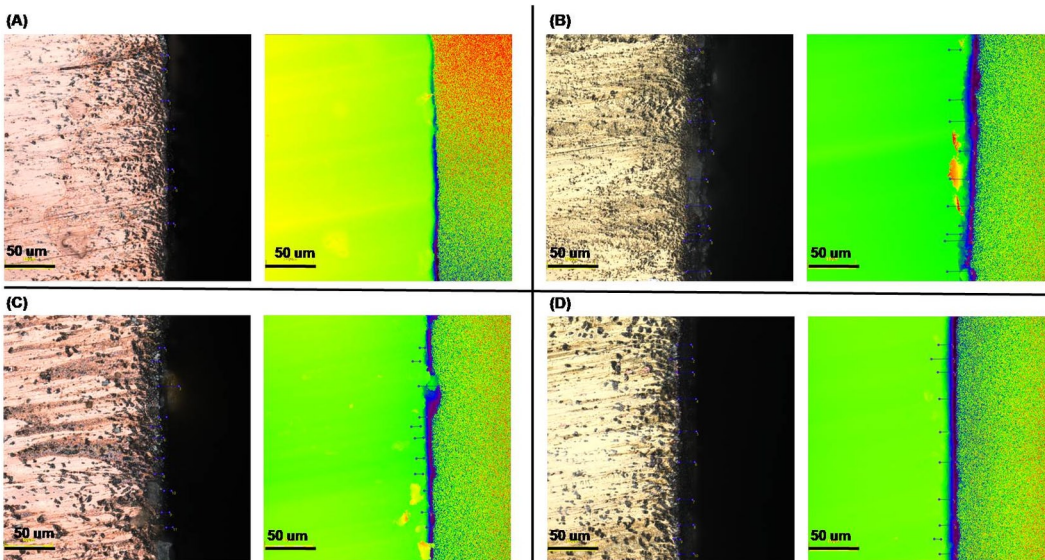


Fig. 4. Set of SEM images of the deposited films cross-sections without (A, B) and with II EMF (C, D): time deposition 900 s; different substrate – Cu (A, C), CuZn (B, D); current density – 50 mA (cm²)⁻¹

electrodeposition (Fig.3A), but in the case of CuZn substrate (Fig.3B), this situation appears after a long time of electrodeposition process duration. It is inherently related to the increase in roughness of the deposited FeNi coatings in the above-discussed cases (Fig.5 and 6) due to enhanced mass transport (MHD effect) of the refined grains.

### 3.3. Tribological tests

First, the roughness of the substrates was measured. As a result, the values of skewness and kurtosis were obtained. For Cu substrate: 0,11 and 4,55 and CuZn: -0,66 and 4,82 respectively.

Positive values of the $S_{sk}$ parameter inform about the predominance of peaks on the measured surfaces (Fig.7.). The surfaces have irregularities, the rough surface becomes spiky and a large number of hills are in contact. The higher values of positive skewness result in good adhesion force and resistance under a certain load which characterizes obtained FeNi coating. A lot of asperities increase the real area of contact [30]. This is especially visible in the cases of EMF absence (Cu substrate) and perpendicular EMF (CuZn substrate) for 900 s deposition time duration, when the values of $S_{sk}$ are the highest –11,08 and 11,88 (Fig.7). In the same cases, values of $S_{ku}>3$ (17,04 and 18,50) reveal the presence of larger grains in the profile of the surface - 20,00 nm (Cu substrate) and 17,00 nm (CuZn substrate) [7, 10, 36].

Due to the dynamic nature of changes in friction conditions, low reproducibility is very common in tribological tests. These changes can be caused by the lapping of kinematic pairs surfaces and generating wear products, often in a stochastic way. These products can be removed beyond the friction zone or remain inside and intensify the secondary wear. The high dynamic of changes in the tribological system has a significant impact on the resistance to motion of the kinematic pair, which analysis is usually used to determine the frictional characteristics of the material. Despite this, friction tests allow to observe some tendencies and dependencies. Results of the wear tests were summarized in Fig.8 and 9.

At high kurtosis values, the asperities become much peakier which drastically decreases the effect of the friction force. Therefore, the coefficient of friction increases slightly to reach the greatest value and decreases further (Fig.8 and 9) [30]. The roughness has the lowest values in both mentioned cases. There are consequences of the appearance of fragments of the damaged surfaces and debris rather than the sharp edges of the deposited FeNi layer [6, 15, 26]. In the case of Cu substrate, the surface characterizes the predominance of peaks but CuZn – of valleys. Roughness profiles of the substrates influence the layers growth to a small extent.

As the result of the electrodeposition process composition gradients in FeNi coatings occurred. The experiments clarified the dependence of the thickness and time for the elemental composition of the coatings. Fig.3A shows the upward trend of the iron amount in all cases (without and with the presence of the EMF) with the time increasing. It is defined as an anomalous co-deposition, where the iron (less noble metal) percentage in the deposit is higher than its presence in the electrolyte (Fe:Ni – 1:1), especially with the longer time of deposition (Brenner, Tabakovic) [1]. The EMF application (at the later stage of deposition) reduces the thickness of the layer on a copper substrate compared to a deposition without an external magnetic field during the longer time of process duration (Fig.4). This is the result of the MHD effect which was mentioned earlier. The case of deposition on CuZn substrate (Fig.3B) differs from mentioned above Cu substrate. The iron content decreases after a longer time of the II EMF application as reflected in the sediment morphology (Fig.2 – CuZn, 3600 s, II EMF). The presence of EMF enlarges the thickness of the deposit. In the case of Cu substrate, it is visible especially in the early stages of
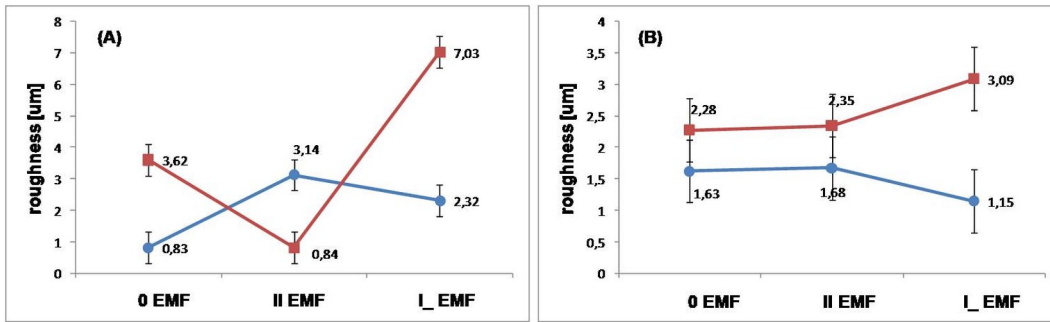
Fig. 5. Values of the roughness of tested surfaces: Cu substrate (A), CuZn substrate (B); time: 900 s – blue line (points), 3600 s – red line (square)
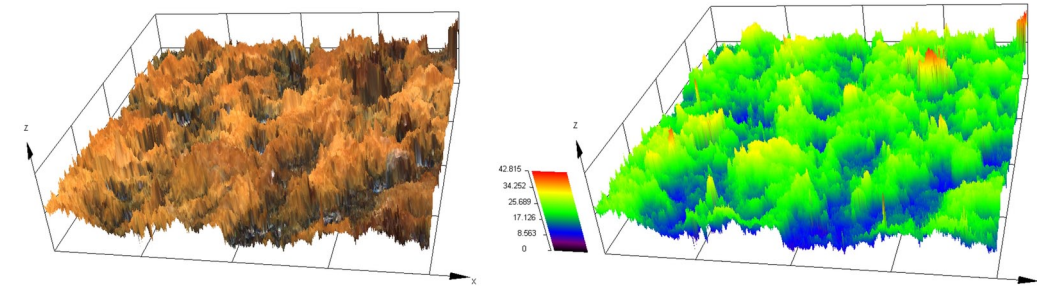


Fig. 6. Set of images of the surface roughness of the deposited films, Cu substrate, 3600 s, without EMF
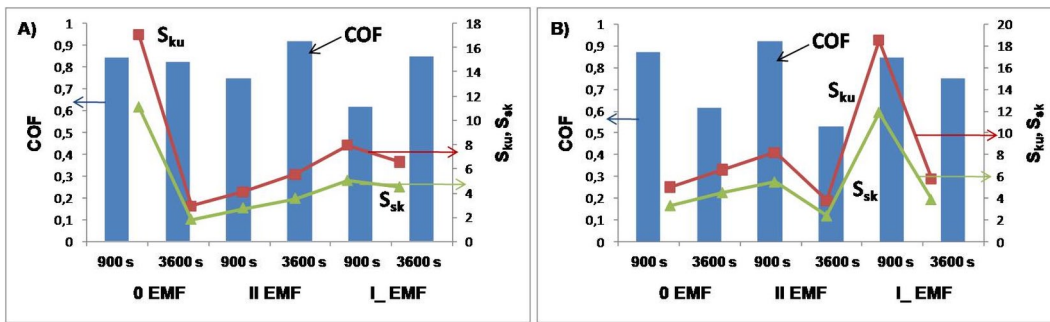


Fig. 7. Values of the friction coefficient after 300 s of the process (COF, blue box), the kurtosis - $S_{ku}$ (red, square), and the skewness- $S_{sk}$ (green, triangle) for coated samples of Cu substrate (A) and CuZn substrate (B)



Fig. 8. Results of tribological tests for Cu substrate for different times of layer deposition 900 s and 3600 s (0 – without EMF, II – parallel EMF, I_ - perpendicular EMF): coefficient of friction vs. time (A), CLSM images of friction marks (B), SEM-EDX analysis of wear tracks (C)

The resistance to motion analysis shows that they are highly dependent on the type of the substrate and the deposition time of the FeNi coating. An intensive rise and subsequent sharp decline in the value of the friction coefficient often occurs in the initial stage of wear ($S_{ku}$>5). It is related to abrasion of surface irregularities, and it can be observed especially for CuZn substrates with a 900 s layer deposition time. The stage of gradual lapping of the surface layer is associated with the lack of rapid changes in the value of the coefficient of friction. As the substrate was exposed, the resistance to the motion increased (Fig. 8 and 9). It particularly occurred for Cu substrates.

The results of the wear track morphology studies presented in Fig.10 showed that in most cases the depth of friction marks increased with the thickness of the layer. Although the thicker FeNi coatings were significantly worn, in most cases their breakthrough was not observed. The opposite tendency can be seen for the layers with a shorter deposition time, where the substrate was fully (Cu) or partially (CuZn) exposed due to friction. Thus, it seems that the coatings on CuZn substrates show higher wear resistance. It is related to an increased degree of adhesion of the layer to the substrate which is in connection with the highest values of skewness and kurtosis - Fig.7 [23, 26, 30]. After the layers deposition, the samples were not rinsed in the ultrasonic cleaner due to the possibility of the layer chipping. They were rinsed only with distilled water. During the tribological tests, the wear products (substrate particles - Cu, Zn, and their oxides) produced a mixture, so it may seem that the trace after friction is shallower (Fig.8, 9 and 10).

On the other hand, SEM images presented in Fig.11 show that thicker coatings may crack and chip. This generates wear products, which are placed in the friction zone. The brittleness of coatings formed at 3600 s deposition time is associated with their decreased microhardness (Fig.12). Due to the movement of the kinematic pair, wear products are crushed and shredded. Thermal energy from friction can lead to their oxidation and formation of oxides. A large amount of wear debris intensifies wear. The microscopic analysis shows that the dominant type of wear was three-body abrasion. Overtime, when the substrate was
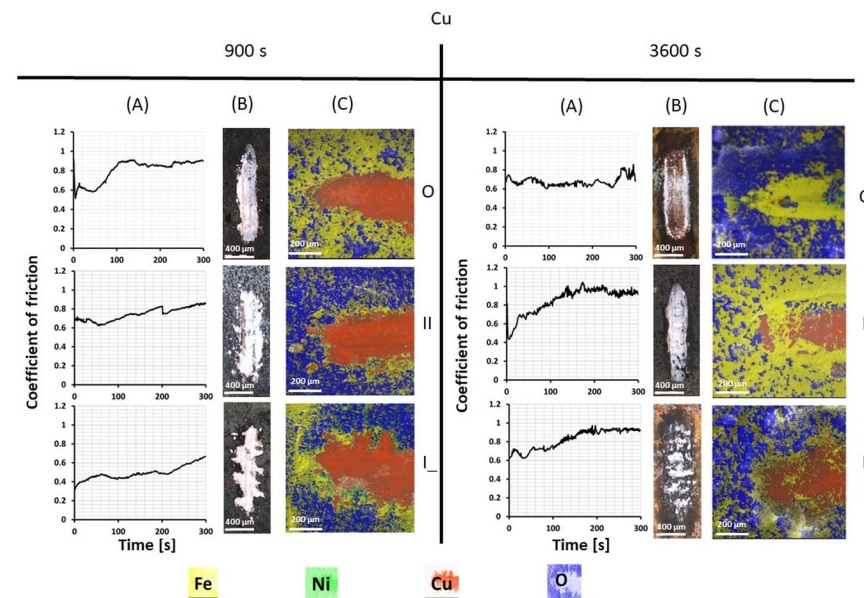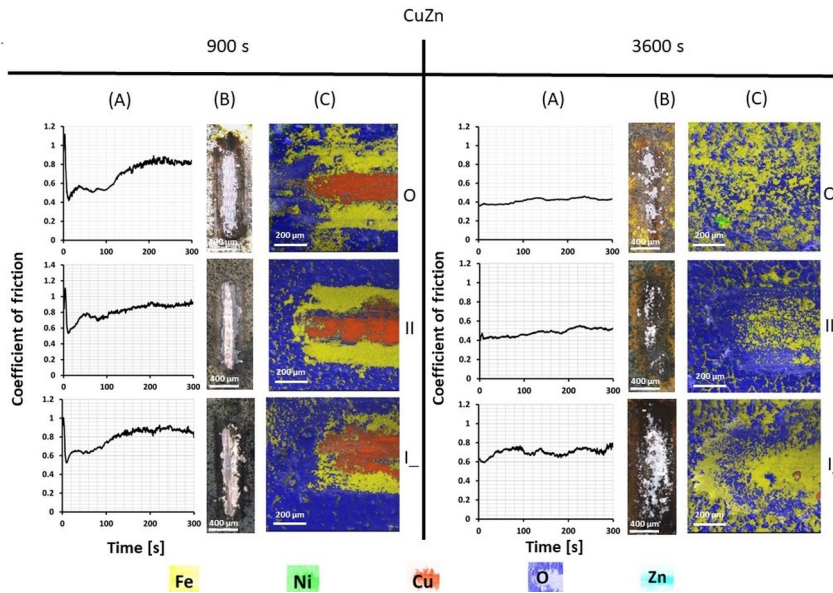
*Fig. 9. Results of tribological tests for CuZn substrate for different times of layer deposition 900 s and 3600 s (0 – without EMF, II – parallel EMF, I_ - perpendicular EMF): diagrams of coefficient of friction vs. time (A), CLSM images of friction marks (B), SEM-EDX analysis of wear tracks (C)*
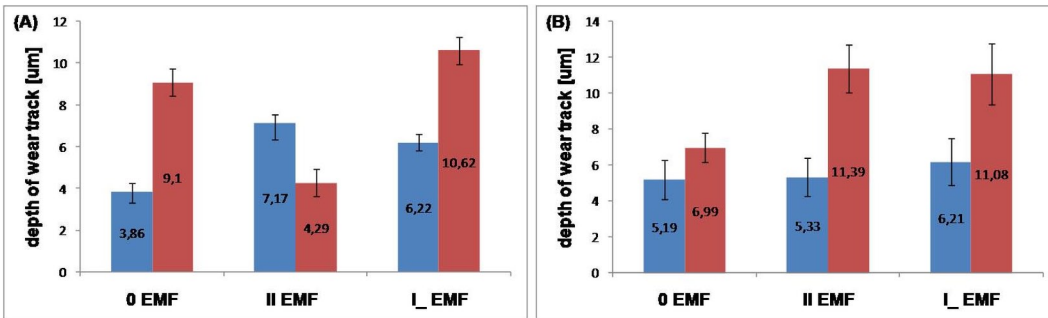


*Fig. 10. Depth of the friction marks for FeNi coating on: Cu substrate (A) and CuZn substrate (B); $F_N$=2 N, f=10 Hz; 900 s – left columns (blue), 3600 s – right columns (red)*



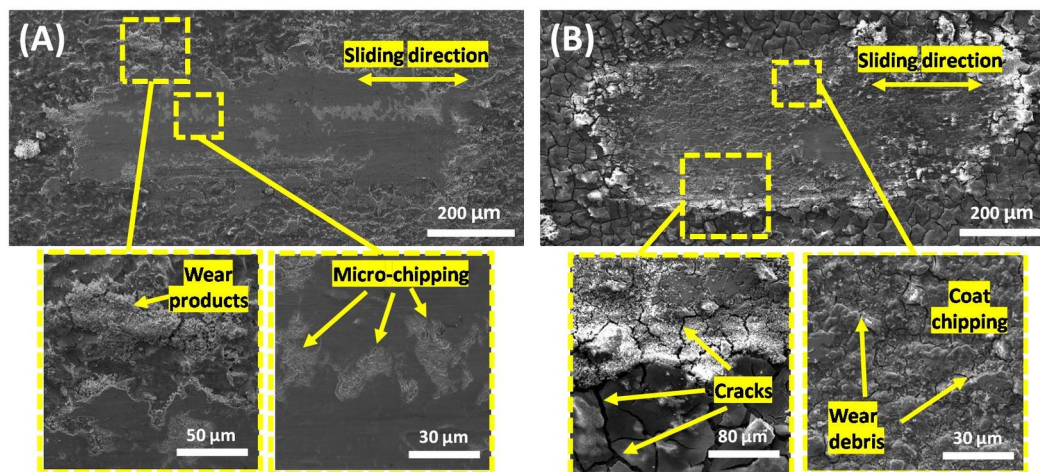*Fig. 11. Set of SEM photos of wear tracks for Cu substrate, 900 s of II EMF (A); CuZn substrate, 3600 s of II EMF (B)*

also XRD analysis (Fig.1) then the peaks typical for metallic Fe appear and are much more intensive.

In most of the considered cases, the intensity of friction marks and values of microhardness (Fig.12) increased with the deposition time except for II EMF on CuZn substrate. The aforementioned dependence is also influenced by the composition of the FeNi coating, as the result, the high content of the iron oxides causes deeper friction marks (Fig.8, 9, and 10). Higher values of the microhardness are connected with the Fe content increase which was shown in Fig.3. The opposite relationship of microhardness in the case of II EMF on CuZn substrate is in agreement with a decrease in Fe content. This is due to the observation that the thickness of the layer was increasing till the end of the experiment. X-ray analysis of the elemental composition of the obtained layers subjected to friction process confirms the increased wear resistance connected with raised Fe content in both scenarios with the absence and presence of the external magnetic field [34].

## 4. Conclusion

The experimental investigation suggests that the electrodeposition process can be modified by applying various orientations of the external magnetic field. The type of layer growth changes from progressive to instantaneous. This may be attributed to the effect caused by mass transfer induced in the electrolytic solution (MHD effect). Extending the duration of the process itself also affects the mode of growth. The early stages of deposition are assigned as a deterministic mode and later – a stochastic one. The surface morphology of the coatings is related to the Fe content and type of the substrate. Further, certain changes in the coefficient of friction are due to an increase in the kurtosis value. Higher wear resistance, as well as coatings adhesion to the substrate, are depending on the higher value of skewness. The increase in microhardness is attributed to the favorable iron electrodeposition which was supported by the XRD study [32]. These findings are of practical significance since they can aid engineers in selecting appropriate surface roughness/ textures based on their specific requirements [30].

Undoubtedly, the external magnetic field used during the deposition allows for the fragmentation of the FeNi coatings, thus obtaining better tribological properties. It is a starting point for further research on the wettability of such surfaces and the durability of layers in nanoelectronics.

exposed to the motion, the amount of generated wear particles decreased, and the surface was smoothed out [38].

Increasing the Fe content (Fig.3) while extending the time of deposition is correlated with higher values of the microhardness (Fig.12). Except in the case of the film deposition on the CuZn substrate in the presence of II EMF when the dependence is opposite, the Fe content and the value of the microhardness decrease. This scenario confirms
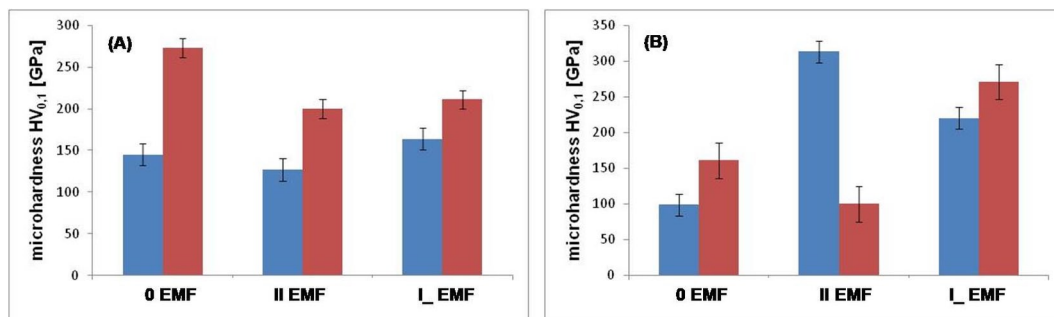
*Fig. 12. Values of microhardness of tested surfaces: Cu substrate (A), CuZn substrate (B); deposition time: 900 s – left columns (blue), 3600 s – right columns (red)*

## References

1. Białostocka A, Idzkowski A. The effect of Ground Changes and the Setting of External Magnetic Field on Electroplating FeNi Layers: Progress in Automation, Robotics and Measurement Techniques in Automation 2019:684-696.
2. Białostocka AM, Klekotka U, Kalska-Szostko B. The Influence of the Substrate and External Magnetic Field Orientation on FeNi Film Growth. Energies 2022; 15 (10): 1-12. https://dx.doi.org/10.3390/en15103520.
3. ChenL, Liu Z, Wang X, et al. Effects of Surface Roughness Parameters on Tribological Performance for Micro-textured Eutectic Aluminum–Silicon Alloy. Journal of Tribology 2020; 142 (2): 021702, http://dx.doi.org/101115/1.4044990.
4. DaltinAL, Benaissa M, Chopart JP. Nucleation and crystal growth in magnetoelectrodeposition. Materials Science and Engineering 2018; 424: 012022, http://dx.doi.org/10.1088/1757-899X/424/1/012022.
5. Dragos O, Chiriac H, Lupu N, et al. Anomalous Codeposition of fcc NiFe Nanowires with 5-55 % Fe and Their Morphology, Crystal Structure and Magnetic Properties. Journal of The Electrochemical Society 2016;163(3): D83-D94, http://dx.doi.org/10.1149/2.0771603jes.
6. Duboust N, Ghadbeigi H, Pinna C, et al. An optical method for measuring surface roughness of machined carbon fibre-reinforced plastic composites. Journal of Composite Materials 2017; 51(3): 289-302, http://dx.doi.org/10.1177/0021998316644849.
7. Dzierwa A, Gałda L, Tupaj M, et al. Investigation of wear resistance of selected materials after slide burnishing process. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020;22 (3): 432-439, http://dx.doi.org/10.17531/ein.2020.3.5.
8. Dziurka R, Madej M, Kopyściański M, et al. The influence of microstructure of medium carbon heat-treatable steel on its tribological properties. Key Engineering Materials 2015;641: 132-135, http://dx.doi.org/10.4028/www.scientific.net/KEM.641.132.
9. Góral A, Lityńska-Dobrzyńska L, Kot M. Effect of Surface Roughness and Structure Features on Tribological Properties of Electrodeposited Nanocrystalline Ni and Ni/Al$_2$O$_3$Coatings. Journal of Materials Engineering and Performance 2017; 26(5): 2118-2128, http://dx.doi.org/10.1007/s11665-017-2662-2.
10. Grzesik W, Niesłony P, Habrat W. Investigation of the tribological performance of AlTiN coated cutting tools in the machining of Ti6Al4V titanium alloy in terms of demanded tool life. Eksploatacja i Niezawodnosc – Maintenance and Relibility 2019; 21 (1): 153-158, http://dx.doi.org/10.17531/ein.2019.1.17.
11. Gurrappa I, Binder L. Electrodeposition of nanostructured coatings and their characterization – A review. Science and Technology of Advanced Materials 2008;9: 043001 (11 pp), http://dx.doi.org/10.1088/1468-6996/9/043001.
12. Gül H, Uysal M, Akbulut H, et al. Tribological Behavior of Copper/MWCNT Nanocomposites Produced by Pulse Electrodeposition. Acta Physica Polonica A 2014;125: 254-256, http://dx.doi.org/10.12693/APhysPolA.1 25.254.
13. Khazi I, Mescheder U. Micromechanical Properties of Anomalously Electrodeposited Nanocrystalline Nickel-Cobalt Alloys: A Review. Materials Research Express 2019, http://dx.doi.org/10.1088/2053-159/ab1bb0.
14. Kuru H, Kockar H, Alper M. Giant magnetoresistance (GMR) behavior of electrodeposited NiFe/Cu multilayers: Dependence of non-magnetic and magnetic layer thickness. Journal of Magnetism and Magnetic Materials 2017;444: 132-139, http://dx.doi.org/10.1016/j.jmmm.2017.08.019.
15. Macek W, Szala M, Trembacz J, et al. Effect of non-zero stress bending-torsion fatigue on fracture surface parameters of 34CrNiMo6 steel notched bars. Production Engineering Archives 2020; 26(4): 167-173, http://dx.doi.org/ 10.30657/pea.2020.26.30.
16. Mbugua NS, Kang M, Zhang Y, et al. Electrochemical Deposition of Ni, NiCo Alloy and NiCo-Ceramic Composite Coatings – A Critical Review. Materials 2020;13: 3475, http://dx.doi.org/10.3390/ma13163475.
17. Morimoto R, Miura M, Sugiyama A, et al. Long-Term Electrodeposition under a Uniform Parallel Magnetic Field. 1. Instability of Two-Dimensional Nucleation in an Electric Double Layer. The Journal of Physical Chemistry 2020; 124; 52: 11854-11869, http://dx.doi.org/10.1021/acs.jpcb.0c05903.
18. Nweze CI, Ekpunobi AJ. Electrodeposition of Zinc Selenide Films on Different Substrates and Its Characterization. International Journal of Scientific & Technology Research 2014; 3; 9.
19. Okamoto N, Wang F, Watanabe T. Adhesion of Electrodeposited Copper, Nickel and Silver Films on Copper, Nickel and Silver Substrates. Materials Transactions 2004; 45; 12: 3330-3333.
20. Palomar-Pardavé M, Scharifker BR, Arce EM, et al. Electrochimica Acta 2005; 50: 4736-4745, http://dx.doi.org/10.1016/j.electacta.2005.03.004.
21. Persson K. Materials Data on FeNi (SG:123) by Materials Project. 10.17188/1197364 (2016).

22. Persson K. Materials Data on FeNi$_3$ (SG:221) by Materials Project. 10.17188/1190197 (2015).
23. Rao VR, Bangera KV, Hegde ACh. Magnetically induced electrodeposition of Zn-Ni alloy coatings and their corrosion behaviors. Journal of Magnetism and Magnetic Materials 2013; 345: 48-54, http://dx.doi.org/10.1016/j.jmmm.2013.06.014.
24. Rezende GLT, Cesar VD, do Lago CBD, et al. A review of Corrosion Resistance Nanocomposite Coatings. Electrodeposition of Composite Materials; 147-185, http://dx.doi.org/10.5772/62048.
25. Sedláček M, Gregorčič P, Podgornik B. Use of the roughness parameters S$_{sk}$ and S$_{ku}$ to control friction – a method for designing surface texturing. Tribology Transactions 2016, http://dx.doi.org/10.1080/10402004.2016.1159358.
26. Sedláček M, Podgornik B, Vižintin J. Correlation between standard roughness parameters skewness and kurtosis and tribological behaviour of contact surfaces. Tribology International 2012; 48: 102-112, http://dx.doi.org/10.1016/j.triboint.2011.11.008.
27. Shuai C, et al. A peritectic phase refines the microstructure and enhances Zn implants. Journal of Materials Research and Technology 2020; 9(3): 2623-2634, https://dx.doi.org/10.1016/j.jmrt.2020.04.037.
28. Sriraman KR, Manimunda P, Chromik RR, et al. Effect of crystallographic orientation on the tribological behavior of electrodeposited Zn coatings. Communication 2016;6: 17360, http://dx.doi.org/10.1039/c5ra15490a.
29. Svahn F, Kassman-Rudolphi Å, Wallén E. The influence of surface roughness on friction and wear of machine element coatings. Wear 2003; 254: 1092-1098, http://dx.doi.org/10.1016/S0043-1648(03)00341-7.
30. Tayebi N, Polycarpou AA. Modeling the effect of skewness and kurtosis on the static friction coefficient of rough surfaces. Tribology International 2004; 37: 491-505, http://dx.doi.org/10.1016/j.triboint.2003.11.010.
31. Thomas BG. Metals Processing. Structure, Processing, and Properties of Engineering Materials. editor J. Adams, Wesley A., chapter 14.
32. Torabinejad V, Aliofkahazraei M, Rouhaghdam SA, et al. Tribological behavior of electrodeposited Ni-Fe multilayer coating. Tribological Transactions 2016, http://dx.doi.org/10.1080/10402004.2016.1230687.
33. Tudela I, Zhang Y, Pal M, et al. Ultrasound-assisted electrodeposition of nickel: Effect of ultrasonic power on the characteristics of thin coatings. Surface & Coatings Technology 2015; 264: 49-59, http://dx.doi.org/10.1016/j.surfcoat.2015.01.020.
34. Wang Z-B, Li W-Y, Shang S, et al. Performance degradation comparisons and failure mechanism of silver metal oxide contact materials in relays application by simulation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (1): 86-93, http://dx.doi.org/10.17531/ein.2020.1.10.
35. Wei X, et al. Impact of anode substrates on electrodeposited zinc over cycling in zinc-anode rechargeable alkaline batteries. Electrochimica Acta 2016; 212: 603-613, https://dx.doi.org/10.1016/j.electacta.2016.07.041.
36. Yang L-L, Chen ChCh, Yuan J, et al. Effect of applied magnetic field on the electroplating and magnetic properties of amorphous FeNiPGd thin film. Journal of Magnetism and Magnetic Materials 2020; 495: 165872, https://dx.doi.org/10.1016/j.jmmm.2019.165872.
37. Zangari G. Electrodeposition of Alloys and Compounds in the Era of Microelectronics and Energy Conversation Technology. Coatings 2015: 195-218, http://dx.doi.org/10.3390/coatings5020195.
38. Żurowski W. Structural factors contributing to increased wear resistance of steel friction couples. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2012; 14 (1): 19-24.

# Intelligent mobility: A model for assessing the safety of children traveling to school on a school bus with the use of intelligent bus stops

Indexed by:
Web of Science Group

Jakub Murawski [a], Emilian Szczepański [a], Ilona Jacyna-Gołda [b], Mariusz Izdebski [a], Dagmara Jankowska-Karpa [c]

[a] Warsaw University of Technology, Faculty of Transport, ul. Koszykowa 75, 00-662, Warsaw, Poland
[b] Warsaw University of Technology, Faculty of Mechanical and Industrial Engineering, ul. Narbutta 85, 02-524 Warsaw, Poland
[c] Motor Transport Institute, ul. Jagiellońska 80, 03-301 Warsaw, Poland

## Highlights

- An analysis of the literature in the area of children travel to/from school was made

- The research procedure for children's travel with the use of IoT was developed

- An original model for the assessment of the children's mobility system was developed

- The developed model was verified on the real example of a children's travel system

## Abstract

The aim of the article is to develop a model for assessing the safety of children's travel. Safety is the most important indicator describing the mobility system of children, even more important than the costs of operating it. Due to the dynamic development of intelligent solutions, it is possible to undertake additional activities supporting the improvement of children's safety when traveling to and from school. However, their implementation requires an adequate assessment of a children's mobility system. Currently, there are no solutions that could comprehensively support the decision-making process in this sphere. The article presents the issues of children's mobility, a literature review in this area, mathematical model for assessing school bus travel, and a computational example. The presented approach is an original solution allowing for evaluation of the existing systems and their development scenarios. In addition, it enables the comparison of children mobility systems of different complexity and scale.

## Keywords

mobility management, safety, transport of children, Internet of Things, intelligent mobility.

## 1. Introduction

Children's mobility is a complex issue and is widely discussed in the literature due to travel safety [23] and children's health [4]. The problem is significant since the number of trips of one child on the route home-school or school-home is approximately 380 times a year, and can contribute to a high traffic volume resulting from the large number of children traveling [24]. Hence, even small changes in the mobility of children can have a significant impact on traffic volumes, air pollution around schools [3] spatial planning [22], environmental impact [41], social conditions [21], economic issues [5], travel planning [16], and children's behaviour in traffic [38] etc.

The decision made by parents on means of transport (e.g. taking a child to and from school in a passenger car) can be impacted by insufficient level of road safety or limited availability of other forms of transport. However, it influences the increase in car traffic, which results in lowering this safety [33]. Therefore, an important issue is to control and reduce the risk in individual areas. This can be achieved by developing models that minimize the likelihood of an accident (e.g. [10]), but also by ensuring the roadworthiness of vehicles from a systemic perspective [12].

These issues indicate the need to develop a comprehensive, integrated children mobility management support system that will ensure the safety of children from the moment they leave home until they return from school [15]. Such a system should ensure the use of school buses and take into account the needs of all stakeholders and users of school transport, i.e. parents (guardians), school representatives, transport providers, municipal authorities and other entities, such as the police or road managers. This means that in order to ensure the proper implementation of school transport and to increase the safety of children traveling to and from school, it is essential to integrate the activities of all the above-mentioned stakeholders and to appropriately define the scope of their duties and tasks.

As shown by statistical data, the interest in the safety of children brings the desired results. Fig. 1 shows the number of road accidents involving pedestrians under the age of 14 in Poland in 2010-2019. As it can be seen, the number of accidents involving children is constantly decreasing, and the level of safety of vulnerable road users is

(*) Corresponding author.
E-mail addresses: J. Murawski (ORCID: 0000-0003-2902-3882), jakub.murawski@pw.edu.pl, E. Szczepański (ORCID: 0000-0003-2091-0231), emilian.szczepanski@pw.edu.pl, I. Jacyna-Gołda (ORCID: 0000-0001-8840-3127), ilona.golda@pw.edu.pl, M. Izdebski (ORCID: 0000-0002-9157-7870), mariusz.izdebski@pw.edu.pl, D. Jankowska-Karpa (ORCID: 0000-0001-5146-4882), dagmara.jankowska@its.waw.pl

increasing. A similar trend is visible in other European Union countries. However, this does not change the fact that the number of accidents involving children pedestrians is significant. For example, in Poland in 2019 there were 723 accidents involving young pedestrians under the age of 14, in which 15 children were killed, 220 were seriously and 504 slightly injured [42]. It is significant that according to the World Health Organization, road transport accidents are the most common cause of death among children over 5 years old [43].
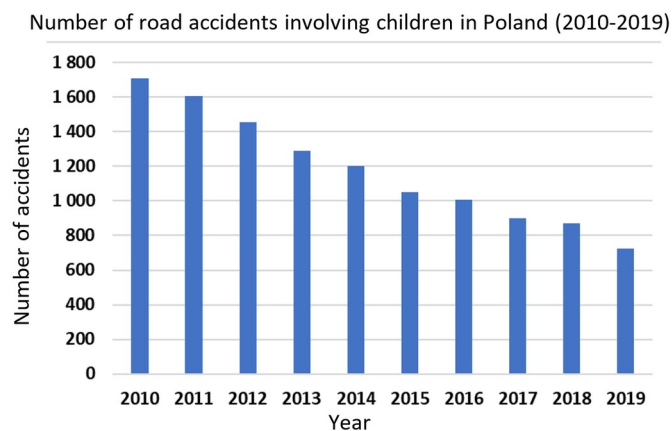
Number of road accidents involving children in Poland (2010-2019)



*Fig. 1. Number of road accidents with children pedestrians in Poland in 2010-2019*
*Source: own study based on Polish Road Safety Observatory data [42]*

Taking all of the above into account, according to authors of the article, the improvement of school travel safety is a particularly important issue and requires proper adaptation of the infrastructure and equipping the transport system with appropriate intelligent tools for monitoring children's travel to and from school.

The authors of the article presented a developed proprietary model for assessing school travel safety, considering the equipment with IoT tools. The research covers trips by organized school transport in non-urban areas, however it is possible to adapt the model to children travels made by public transport, as well as by other means of transport or on foot. The presented research is based on the effects of the Safeway2School project [44]. Extending the research would require mapping additional means of transport, taking into account their characteristics, as well as the assessment of integrated transfer nodes presented, for example, in the work of Jacyna et al. [11]. These considerations will be the subject of further studies.

The article is mapped out as follows. Section 2 presents the current state of knowledge broken down by areas related to the research topic. Section 3 presents the concept of the proposed method for using intelligent solutions in children's travel to school. Chapter 4 contains a decision-making model which studies the indicators of bus travel safety, based on the developed concept. The use of the developed concept of the proprietary safety model was verified on the real data of the selected part of the road network, as presented in section 5. The article ends with a summary containing the conclusions of the considerations and indicates recommendations for the stakeholders of the entire process of organizing children's travel to and from school.

## 2. The state of knowledge in the field of safe travel of children to and from school

### 2.1. The choice of transport means for children school travel

The decision on how to arrive to school reflects the highest possible level of safety of a child. This thesis is confirmed by research [39], which shows that for parents who bring their children to school by car, the decisive factors for such a choice were: ensuring the highest possible level of safety and the comfort of travel. Indeed, taking a child to school by car, thanks to the presence of a parent, provides the child with a high level of safety, but on the other hand, this mode of transport has many disadvantages. First of all, the use of a passenger car results in the reduction of physical activity for children, and thus may have an impact on weight gain [27] and lowering the health of children (e.g. increasing the risk of cardiovascular diseases) [4]. Apart from that, the use of a passenger car has a negative impact on the natural environment, is inconsistent with the idea of sustainable development [24] and ineffective from an economic point of view. In the literature, there are also studies indicating that car use has negative social effects, because children spend less time with their peers [21].

Therefore, it is extremely important to develop and implement solutions that improve the safety of children in transport. Thanks to that, it is possible to reduce selected risks associated with other forms of transport than using a passenger car to get to school, and thus the use of other types of transport that can not only ensure the safety of children, but also can have a positive impact on children's health, and not be harmful to the environment [2].

It is worth noting, however, that it is not possible to determine the best means of transport in advance, as this issue should be considered individually in each case. The choice of a given means of transport on a given route is associated with certain risks, which may depend, among others, on the length of the route from home to school, available road infrastructure (e.g. sidewalks and bicycle paths), accessible pedestrian crossings, traffic volume on streets around the school, speed of vehicles around the school, availability of school transport, location of bus stops, spatial development in the school surroundings and many other factors [37]. Moreover, as indicated by the authors [32], an undoubtedly important aspect in the choice of means of transport is the assessment of the accessibility of public transport.

### 2.2. Activities increasing children school travel safety

The organizational and technological solutions can be distinguished among the activities increasing the safety of children in transport,. The most popular forms of ensuring safety for children traveling to school are traffic calming zones on the streets around educational units, which can be classified as organizational solutions. In these zones, various measures are applied to ensure safety of all road users. Different tools are used for this purpose, with reduction of the speed limit being the most common one. The conducted research shows that noticeable effects in terms of increased safety are achieved by limiting the speed to at least 30 km/h [34]. Other solutions used in traffic calming zones around schools include additional horizontal and vertical signage, pedestrian refuge, narrowed lanes, restriction of entry for selected types of vehicles, or additional lightning. Zones of this type may be introduced during the morning and afternoon rush hours, or they may be valid throughout the whole day [2, 18].

An example of a more restrictive measure to ensure the safety of children in road traffic is project "Schulstraße" implemented in Vienna, which closes streets in the immediate vicinity of schools to car traffic for half an hour during the peak morning traffic. The aim of the project is to contribute to the improvement of road safety and to be an incentive to change the means of transport to a more environmentally friendly one, e.g. a bicycle or a scooter. The first effects of the described project are extremely positive, and its implementation is planned by other cities as well (e.g. Wrocław). The "Schulstraße" project is one of several activities implemented in Vienna, which are in line with the idea of Smart City. They can include, among others promoting pedestrian traffic, using facilities for electric vehicles or restricted parking zones [26].

Solutions aiming to improve road safety have been also introduced in Sweden, where in 1997 the parliament adopted the long-term strategy "Vision Zero". It assumes the development and shaping of the transport system in such a way that in the future no one is killed or seriously injured as a result of a road accident. Therefore, all transport solutions are designed in such a way as to ensure maximum safety

for all road users. As a result, the number of road accident victims is decreasing annually, and the country's road transport system is considered as one of the safest in the world (9 people under the age of 17 died as a result of road accidents in Sweden in 2019). Despite this, further innovations aimed at achieving the goal specified in the strategy are being implemented. Apart from Sweden, strategies based on "Vision Zero" have also been adopted by among others Denmark, Norway and the USA [1].

The subject of systemic guaranteeing the safety of children on the way to school was the subject of, inter alia, Safeway2School project, funded by the European Commission under the 7th Framework Program. As part of the project, innovative solutions in the field of school travel planning, warning systems for school bus drivers were developed, as well as the formulation of educational programs dedicated to all stakeholders of the school transport system. The developed solutions were tested in pilot locations in Sweden, Austria, Poland and Italy, which made it possible to evaluate their usefulness and effectiveness [44].

The implementation of experimental concepts in operating transport systems is difficult to apply due to the scale and their sensitivity. All modifications and improvements entail high costs and therefore must be preceded by detailed studies on the developed analytical or simulation models. The authors of the work [14] indicate such regularity by presenting the issues of the development of transport systems and decision-making problems related to the issues of the impact of transport on the environment. Impact and evaluation models for transport solutions need to be developed. Świderski et al. [35] presented a transport service evaluation model based on artificial neural networks, while Rudyk et al. [28] proposed an analytical model for assessing vehicle fleet management in terms of safety, which was used to feed the rationality of the simulation model. In case of school transport travels, there is a noticeable lack of a tool allowing for its assessment in terms of safety and a comprehensive approach to children's mobility problem.

## 2.3. Using IoT solutions for monitoring children travels to and from schools

Solutions based on the Internet of Things technology take an important place in the literature on road safety. In trying to define IoT, it can be said that it is a set of physical (real) objects, connected in such a way that they can communicate with each other (send and receive data) without human interference [36]. IoT has a huge potential to be used in many areas of life, including road traffic. As indicated in [6], IoT provides communication vehicle-to-vehicle (V2V) as well as vehicle-to-infrastructure (V2I). Thus, it is possible to ensure communication between cars, which can help avoid collisions [8]. The communication channels provided by the use of IoT technology are presented in Fig. 2. The authors of the work [20] emphasize the great importance of IoT in the development of logistics systems, including external transport, and the potential it has in terms of supervision over the correct implementation of processes and improvement of systems.

As mentioned, the applications of IoT technology in road transport are widely described in the literature. IoT technology can also ensure communication between the vehicle and traffic lights. In [19], a solution is described which, thanks to the use of data from motor vehicles and information on pedestrian traffic, allows functional optimization of road lights. Another possible application of IoT technology is to provide vehicle-to-pedestrian (V2P) communication. For example, [9] describes a technology that allows a vehicle to be warned about an approaching pedestrian and pedestrian (using a dedicated mobile application) about a risk from an oncoming vehicle. Taking into account the above considerations, it can be said that the IoT technology in the future will make it possible to limit or even exclude the role of humans in driving a car, which can contribute to a significant improvement in road safety [40].
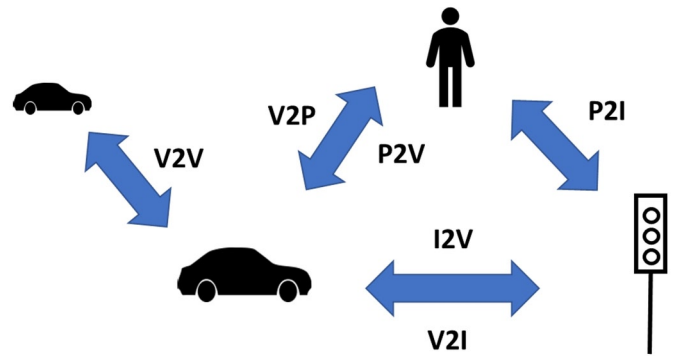


*Fig. 2. Communication channels provided by IoT technology*
*Source: own study*

In the context of the described subject, it is worth emphasizing that the IoT technology is also applicable in terms of improving the safety of children traveling to schools. Among the technological solutions that improve safety of children, there are various types of applications that enable tracking of school buses. As a rule, their operation is based on the data collection concerning, inter alia, school bus location. The data collected on the server is made available to selected stakeholders, e.g. parents, school representatives or transport organizers. In the past, solutions of this type were most often based on GPS technology (as described in [7, 29]) and, slightly less frequently, on passive RFID technology [30]. Commercial applications of this type include Track-SchoolBus app. It should be emphasized that the functioning of the mobility management support system must be based on a properly developed architecture, equipped with devices for collecting and transmitting information, but also on data warehouses allowing for data collection and analysis, as well as transferring information among transport process stakeholders. Such solutions are implemented, for example, in cargo transport, an example of which can be the EPLOS [13] system and its architecture. Similar solutions can successfully operate on a micro scale basis in terms of children mobility management. However, currently the literature lacks examples of a systemic approach covering such a large range of stakeholders who should be involved in the process of children's travel to school.

The use of IoT technology in the field of monitoring school buses opens new possibilities. For example, [25] presents a system that allows the student's absence monitoring, unscheduled stops, deviation from the route or exceeding the speed limit. On the other hand, the solution described in [31] allows for the identification of fuel leakage or control of the temperature inside the bus. Another idea for using IoT technology is for an intelligent bus stop, which was described in [17]. In this case, the bus stop is a closed, air-conditioned unit (stops of this type can be found, among others, in the Persian Gulf countries). Thanks to the use of IoT technology, it allows, among others optimal management of electricity by predicting the occupancy level at the stop, remote control of air conditioning and lighting, and measuring the level of air pollution around the stop. However, there are no solutions that are considered in this article concerning additional control, e.g. lighting of a stop and emitting warning signals for other road users.

Based on the analysis of current knowledge and literature review, it should be stated that it is necessary to conduct research aimed at developing a concept for the development of transport systems with regard to safety issues. In this article, the concept of Intelligent Bus Stops described in [17] was extended and it was planned to integrate IoT technology in the supervision of children's travel to school. Moreover, a proposal was made of an analytical model, based on the existing evaluation models, dedicated to the assessment of the safety of children's travel to school. This model has the potential to be used in simulation models. Considering the dynamic development of IoT technology, it can be expected in the near future that further transport solutions based on this technology will become available, which can

bring many benefits to society, also in the context of improving the safety of children on the way to school. Hence, a dedicated model, which is highly flexible and can be expanded with additional elements, will be an important tool in assessing the impact of implementing various concepts and solutions in the field of safety.

## 3. Procedure of the study method in the field of supporting children's travel to and from school with the use of IoT tools

Figure 3 shows a scheme of a mobility system using intelligent solutions (IoT). The basic element of such a system is an intelligent bus stop. It can be equipped with many different elements that improve the safety of passengers, including getting on and off the bus. Such a bus stop can be equipped with additional lighting activated on the basis of reading sensors and beacons or NFC transmitters when a child arrives there (and is equipped with a smartphone application or an additional transmitter). Also, while getting on and off the bus, additional lighting may be activated in the bus stop area, informing other road users about the necessity to become more alert. Allocating unique numbers to transmitters can be used for monitoring the journey of many road users, as well as for collection of data allowing for further improvements in the children's mobility system. Additional equipment may include cameras or photocells.
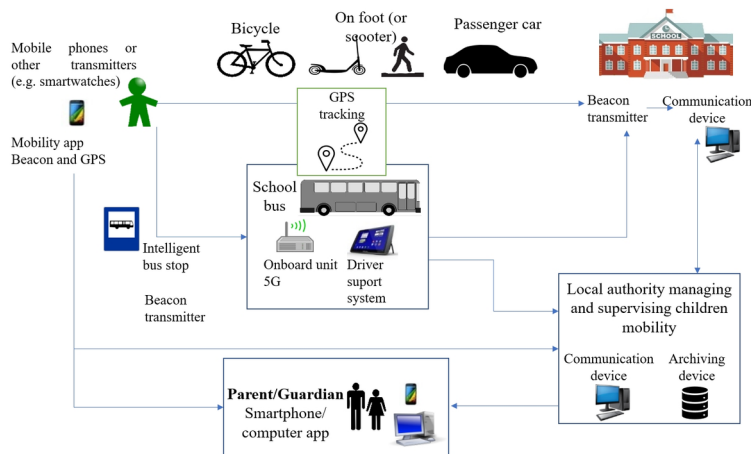


Fig. 3. *Information/data flow in children's mobility management system with the use of IoT*
*Source: own study based on [20]*

The travel process to and from school is similar for each mode of transport. This similarity is mainly expressed in the stages of the journey; however it is superficial. In reality, each way of getting to school or returning home has different characteristics due to elements such as time, comfort, safety and additional travel support equipment. Currently, in most cases this process is not supported by smart solutions. In case of children and adolescents, smartphones play a main role as they allow for communication with the guardians. There are no systemic solutions that integrate communication among children, school, children mobility management office and guardians. With the use of intelligent infrastructure provided by the Internet of Things, it is possible to implement solutions that increase the travel safety and, at the same time, the flow of information, and allow for an increase in the share of independent journeys made by children. Figures 4 and 5 show the different processes of children's travels to school.

As it should be noted, the simplest process and, at the same time, the one in which the share of additional equipment is the
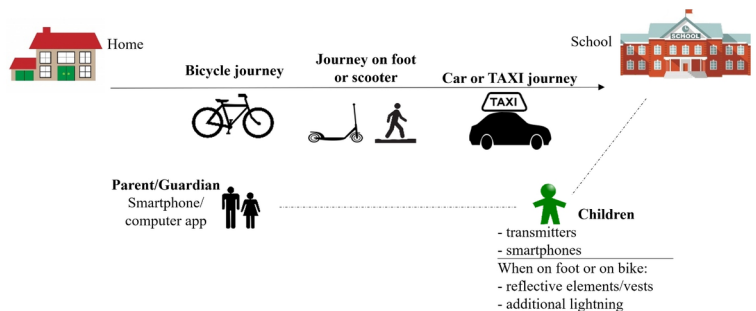


Fig. 4. *Process of school journey by bicycle/scooter/on foot/in car (without public transport)*
*Source: own study*

lowest, is a journey made by a passenger car. A child under the care of a parent/guardian does not need to be equipped with any additional systems. Assuming that the child is brought to school and taken home by the parent/guardian, it remains under the supervision of the parent all the time, thus the level of safety (from the parent's point of view) is quite high. Transmitters (with GPS tracking function) can assist in supervising the child's safety when traveling to school by taxi. Another, uncomplicated process, but with a lower level of safety, is commuting to school by bike or on foot (including a scooter). The child's safety depends to a large extent on the existing infrastructure (e.g. sidewalks, bicycle paths), but also on child's education and awareness levels or additional equipment (e.g. lighting, reflective materials). In this case, equipping the child with a transmitter and automatic notifications sent to school and guardians is desirable.

The most extensive process and at the same time the one in which an extensive system based on the Internet of Things can be used is the child's trip to school by a school bus. Due to long distances to schools, poorly developed infrastructure for pedestrians and cyclists, as well as the route to school leading through dangerous streets and complicated intersections, journeys by this means of transport are often organized by municipalities in the form of organized school transport. Also, parents often choose this mode of transport for their children, which makes it much easier for them to organize their daily activities without having to drive and pick up children from school, and thus the congestion around schools and on access roads remains lower. Therefore, school buses are a good choice for traveling to and from school. In case of this solution, it is possible to use many additional elements, e.g. infrastructural ones, including Intelligent Bus Stops, which can significantly increase the safety of
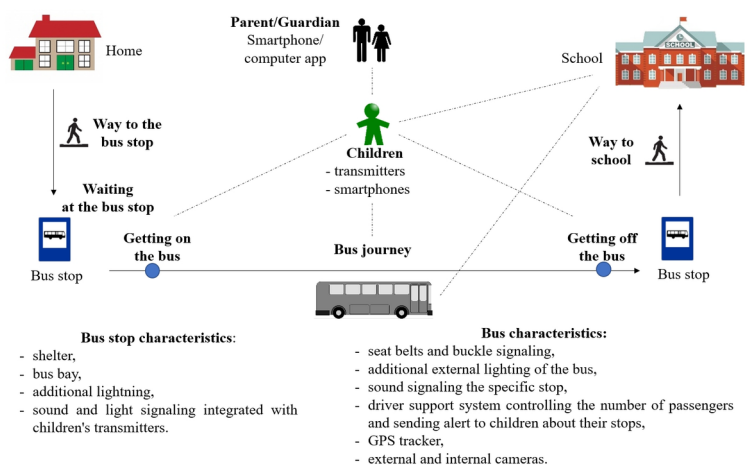


Fig. 5. *Process of school journey by school bus including the characteristics of additional equipment of individual elements of the system*
*Source: own study*

children traveling to and from school. Additionally, parents/guardians and the school representatives can be kept informed of the child's travel status. The use of beacon transmitters enables Bluetooth communication and confirmation of the location of the child, especially in case of weak or no GPS signal, as well as in the absence of a GPS receiver. An alternative to beacon transmitters can be NFC communication, but it is less popular than Bluetooth.

Nowadays, the safety of children traveling to school is one of the most important topics for students, parents, schools, and transport system managers. Due to the increase in the popularity of advanced electronic devices, various transmitters or smartphones, the implementation of solutions improving safety of children is possible with relatively low financial outlays. However, it is necessary to assess the impact of individual elements and its activities on the safety of the children. Therefore, the article proposes a formalized method of assessing children's travel to school by bus as a function of safety.

The study procedure in the field of supporting children's travel to and from school with the use of IoT tools is presented in Figure 6. Within this procedure, 7 stages were distinguished, which constitute a comprehensive approach to the implementation of improvements in the area of children's mobility. This method is universal, and it is possible to implement it in different means of transport, as well as in different areas.
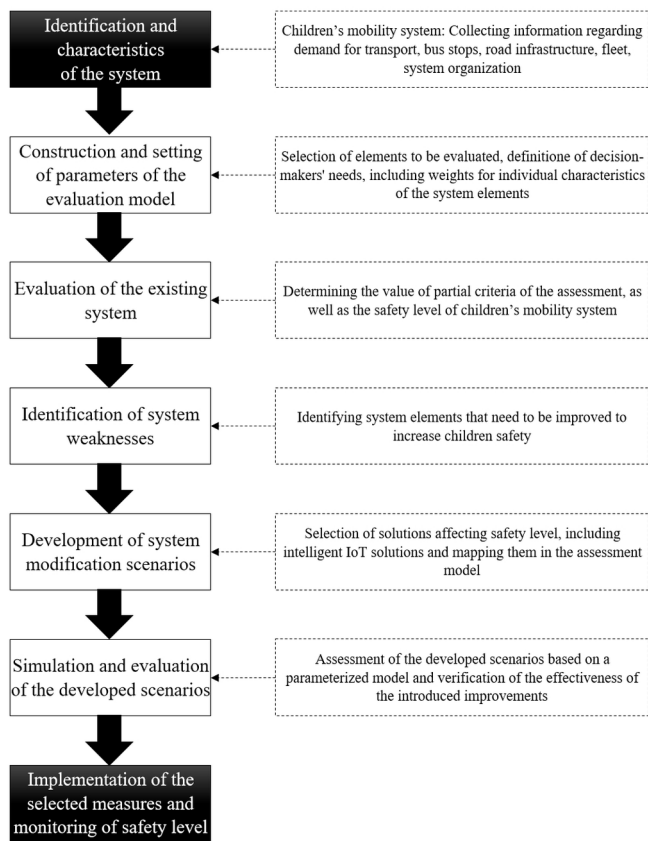


*Fig. 6. Stages and phases of school travel in children's mobility system*
*Source: own study*

The article examines one of the four main means of children transport to reach school, i.e. traveling by organized school bus, due to the high popularity of this mode and also the greatest impact of intelligent systems on this journey's safety. Access to school in the community is considered and it has been assumed there is no other method of traveling to school by any other form of public transport. Travel by public transport is possible in other locations where this kind of transport is organized, i.e. larger towns. The authors focused mainly on the areas where

accessibility of public transport is limited and the process of reaching the school is complicated.

## 3. The formalization of the mobility assessment model in the safety function

### 3.1. Model assumptions and system features

The assessment model for decision support of child mobility management will be presented in a systemic perspective, i.e., monitoring children's travel to school in a door-to-door perspective. The model includes an installation of the Smart Bus Stops on the route of school buses and equipping children with special transmitters (or telephones with a dedicated application).

A children's mobility system meets the demand for transporting primary school children from their place of residence to schools. The diagram of such a system, taking into account all stages of the journey, broken down into the phase to and from school, has been presented in Figure 7. Each phase may consist of different stages.

The phase in individual stages can be carried out by various transport means, while phases 2 and 3 are carried out only by the school bus. Phases 1 and 4 are carried out on foot, and phases 1.2 and 3.4 can be carried out by a passenger car (with a parent/guardian or taxi), bike, scooter or on foot. Due to the assumptions mentioned above, the authors considered phases related to bus travel, i.e. 1 and 2 as well as 3 and 4.

Phases 1 and 2 as well as 3 and 4 (the bus journey) are complex and have to include both used infrastructure and assistive devices. In addition, the length of a journey or the number of bus stops will also affect the assessment of a children's mobility system. There may be many travel routes to school in a system.

The described model provides a system assessment for Phase 1 - travel to school. The safety assessment criterion ($\overline{F_{B\_A}}$) was chosen for the mobility assessment. The total assessment of a safety level with the use of the bus will be expressed as follows:

$$\overline{F_{B\_A}} = \overline{F_{B1}} + \overline{F_{Bp}} + \overline{F_{B2}} + F_{B\gamma} + F_{B\theta} \tag{1}$$

It was assumed that a safety assessment will be the sum of 5 components, which mathematical notation is presented in section 4.2. The first three components are directly related to the implementation of stage 1, i.e. $\overline{F_{B1}}$ concerns the assessment of reaching the bus stop, $\overline{F_{Bp}}$ concerns the waiting time at the bus stop, $\overline{F_{B2}}$ concerns the bus journey assessment. The two additional components $F_{B\gamma}$ and $F_{B\theta}$ are determined by the decision-maker and can be added to the overall security level. The element $F_{BA\gamma}$ has an interpretation of the index correcting a level of safety due to knowledge gained by the students during the training and lessons related to road safety. Moreover, the element $F_{BA\theta}$ adjusts a level of safety with additional children's equipment, such as reflective vests, telematics devices or personal
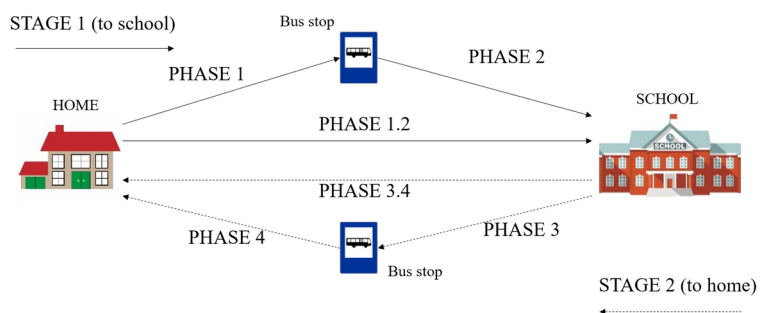


*Fig. 7. Stages and phases of travels to and from school in a children's mobility system*
*Source: own study*

lighting. They mainly affect children travelling by the school bus, but can also be used to travel to school by other means.

The safety assessment requires including many system features. To determine its level, a set $Q$ containing the weights of individual characteristics was defined. It was written as:

$$Q = \begin{cases} q_l(1), q_l(2.1), q_l(2.2), q_l(2.3), q_l(3), q_l(4) \\ q_p(1), q_p(2), q_p(3), q_p(4.1), q_p(4.2), q_p(4.3), \\ q_b(1), q_b(2), q_b(3), q_b(4), q_{st}(1), q_{st}(2), q_{st}(3) \end{cases} \quad (2)$$

The individual elements of the above-described set were divided into:
- determining the influence of the characteristics of a connection ($q_l$),
- determining the influence of the characteristics of a bus stop ($q_p$),
- determining the influence of journey and transport mean ($q_b$ and $q_{st}$).

Weights can be modified and changed depending on the needs and preferences of the model user. After calibration through appropriate adjustment of the weights, they should not be changed when different scenarios or systems are compared and assessed. Weights can be interpreted in various ways, depending on the needs of a model user. The paper assumes that weights will be treated as points allocated depending on the respective system characteristics. These points will be assigned from a set of integers. They can have positive or negative values.

The weights used to assess children's travel to the bus stop are presented in Table 1. They can also be used to build a more complex model that includes the assessment of journey to school by scooter, foot or bicycle. In the case of including in a model the journeys by a passenger vehicle, it would be necessary to introduce weights referring to, for example, the road category or the safety of the connection.

Another group of points allocated to a system relates to bus stops. It is essential due to the significance of the stage of waiting for the bus to arrive and getting children on and off the bus and the impact of these stages on the overall safety of children travelling to school. The group of weights related to these factors is defined in Table 2.

Another group of weights used for an assessment is the one related to the bus journey to school. This is the phase of the journey in which the infrastructure also plays an important role, however, equipping the vehicle with additional security systems can significantly increase the level of safety of children as passengers. The defined weights are presented in Table 3.

The system's safety assessment function consists of many elements related to all phases of the journey to school. Since stage 2 (returning

Table 1. The group of weights related to the assessment of connections between nodes

| Weight | Description |
|---|---|
| $q_l(1)$ | points (a weight) allocated for the presence of a sidewalk at a given connection |
| $q_l(2.1)$ | points allocated for the wide shoulder, it was assumed that there is no sidewalk in this case |
| $q_l(2.2)$ | points allocated for the narrow shoulder |
| $q_l(2.3)$ | points allocated for the other solutions or lack od additional improvements |
| $q_l(3)$ | points allocated for the lighting of a given connection |
| $q_l(4)$ | points allocated for the speed of vehicles on given connection, according to a principle the higher speed, the lower safety |

Table 2. The group of weights related to the assessment of bus stops

| Weight | Description |
|---|---|
| $q_p(1)$ | points (a weight) allocated for the presence of the bus stop turnout |
| $q_p(2)$ | points allocated for the presence of the bus stop shelter |
| $q_p(3)$ | points allocated for the presence of the bus stop lighting |
| $q_p(4.1)$ | points allocated for the presence of the intelligent lighting activated by the child's transmitter |
| $q_p(4.2)$ | points allocated for the presence of the road signs and the light signals, vertical and horizontal on the road |
| $q_p(4.3)$ | points allocated for the presence of other solutions improving safety |

Table 3. The group of weights related to the assessment of school buses and bus routes

| Weight | Description |
|---|---|
| **Points allocated for bus route** | |
| $q_b(1)$ | points (a weight) allocated for the number of stops on the bus stops during given route, the fewer stops, the higher the score |
| $q_b(2)$ | points allocated for the length of the bus route, the shorter route, the higher score |
| $q_b(3)$ | points allocated for the speed on given route |
| $q_b(4)$ | points allocated for the road lighting |
| Points allocated for bus equipment | |
| $q_{st}(1)$ | points (a weight) allocated for the equipping of the bus with the seat belts for passengers |
| $q_{st}(2)$ | points allocated for the additional lighting for the bus, which is activated at the bus stop |
| $q_{st}(3)$ | points allocated for the additional telematics equipment of the bus integrated with the system |

home) is analogous and the only difference is the waiting at the bus stop, its evaluation wasn't described. The road to school is critical in this case. The components of the safety evaluation function are the assessment of reaching the bus stop, the assessment of the bus stop and the assessment of the bus route weighted with the share of serviced notifications (the number of children getting on the bus). The considerations do not include access to the school from the bus stop, as it is assumed that it is a safe section (usually, the bus stop is at the school or the school area).

## 4.2. The assessment function of the bus journey

An assessment of the bus journey safety level requires the calculation of individual components of the function $\overline{F_{B\_A}}$. Its first element $\overline{F_{B1}}$ can be interpreted as the assessment of all phases of reaching the bus stop in a system for all bus routes:

$$\overline{F_{B1}} = \sum_{d \in D} \sum_{(w',w) \in L \wedge w \in \Psi(d)} f_{B1}(d,w',w)$$

It is the sum of the assessment of the arrival phases safety level of the $d$-th route - $f_{B1}(d,w',w)$, using the house-bus stop connection $(w',w)$ belonging to the set of the connections $L$. Considerations taking into account only bus stops which belongs to the set of bus stops visited during given bus route $\Psi(d)$. The assessment of reaching a bus stop is based on the characteristics of a connection between the home and a bus stop and the number of children using this connection. This assessment is made on each bus route ($d$) and at each bus stop ($w$) belonging to a set of bus stops visited during a given bus route $\Psi(d)$:

$$\forall d \in D \quad \forall w : w \in W \wedge r(w) = 2 \wedge w \in \Psi(d) \quad \forall w' : w' \in W \wedge r(w') = 1$$

$$f_{B1}(d,w',w) = \frac{x_a(d,w',w)}{x_{ap}(d,w)} \begin{bmatrix} [q_l(1)c_{lch}(w',w)] + \\ \begin{bmatrix} q_l(2.1)c_{lps}(w',w) + \\ q_l(2.2)c_{lpw}(w',w) + \\ q_l(2.3)(1-c_{lps}(w',w)-c_{lpw}(w',w)) \end{bmatrix} [1-c_{lch}(w',w)] + \\ q_l(3)c_{los}(w',w) + \\ q_l(4)\left[\frac{100 - V_{sr}(w',w)}{100} + \frac{V_{min}}{100}\right] \end{bmatrix}$$

$$\cdot \left( (1-\alpha_P(w',w)) \frac{l(w',w) - l_{Pgr}}{l(w',w)} \right)$$

The $r(w)$ parameter in the above function takes the value 1 when the node is interpreted as a home, and 2 when the node is interpreted as a bus stop. This function consists of three main elements (Figure 8 shows a graphic interpretation of the assessment of access to the bus stop). The first element is the share of the number of children using a connection to the $w$-th bus stop ($x_a(d,w',w)$), in relation to the overall amount children using this bus stop ($x_{ap}(d,w)$). The second element is a connection assessment, which depends on connection characteristics and the allocated weights of the individual characteristics 2.1)+$q_l$(2.2)+$q_l$(2.3), $q_l$(3), $q_l$(4)). The characteristic of a connection includes:

- $c_{lch}(w,w')$ – the percentage of the pavement in relation to a connection length,
- $c_{lps}(w,w')$ – the percentage of the wide shoulder in relation to a connection length,
- $c_{lpw}(w,w')$ – the percentage of the narrow shoulder in relation to a connection length,
- $c_{los}(w,w')$ – the percentage of the illuminated road in relation to a connection length,
- $V_{sr}(w,w')$ – the average speed of a connection, it is assumed that it cannot be higher than 100 kmph,
- $\alpha_P(w,w')$ – the indicator determining whether the level of safety is influenced by the length of the foot connection; it takes the

value 1 when the length of the link $l(w,w')$ is greater than $l_{Pgr}$ for foot journey,
- $V_{min}$ – speed of the cars on a connection, which does not affect safety,

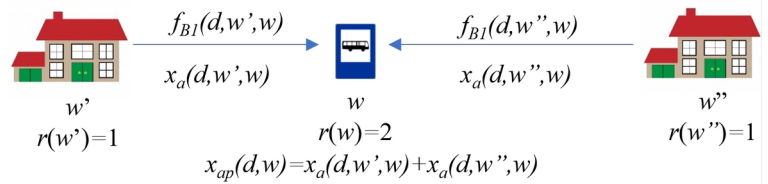

$$x_{ap}(d,w) = x_a(d,w',w) + x_a(d,w'',w)$$

Fig. 8. *Graphical interpretation of a safety assessment elements of access to a bus stop*
*Source: own study*

The assessment function of the access to the bus stop is weighted by the influence of the length of the route, if it is longer than the assumed $l_{Pgr}$ (the limit length of an access to a bus stop for safety reasons).

The second element of a safety assessment of the journey to school using the bus is waiting time at a bus stop. As in the case of assessment of access to a bus stop, this function can also be applied to all bus stops in the system and can be formulated as:

$$\overline{F_{Bp}} = \sum_{d \in D} \sum_{w \in \Psi(d)} f_{Bp}(d,w)$$

This element depends on the characteristics of a bus stop and the weights of its individual features. It is weighted by the number of children using a given bus stop ($x_{ap}(d,w)$) to the number of children transported on the entire bus route $x_{atr}(d)$.. It can be formulated as below

$$\forall d \in D \quad \forall w : w \in W \wedge r(w) = 2 \wedge w \in \Psi(d)$$

$$f_{Bp}(d,w) = \frac{x_{ap}(d,w)}{x_{atr}(d)} \begin{bmatrix} q_p(1)c_{pz}(w) + q_p(2)c_{pw}(w) + q_p(3)c_{pos}(w) + \\ q_p(4.1)c_{pt1}(w) + q_p(4.2)c_{pt2}(w) + q_p(4.3)c_{pt3}(w) \end{bmatrix}$$

In the above function, the following parameters characterizing a bus stop were adopted:
- $c_{pz}(w)$ – determining whether there is a bus stop turnout, takes the value 1 if there is a bus stop turnout or 0 if there is no bus stop turnout;
- $c_{pw}(w)$ – determining whether there is a bus stop shelter (1 – yes, 0 – no),
- $c_{pos}(w)$ – determining whether there is a bus stop lighting (1 – yes, 0 – no),
- $c_{pt1}(w)$ – determining whether a bus stop is equipped with additional lighting activated by a child's transmitter and a bus (1 – yes, 0 – no),
- $c_{pt2}(w)$ – determining whether a bus stop has additional light and / or sound marking activated by the child's transmitter (1 – yes, 0 – no),
- $c_{pt3}(w)$ – determining whether a bus stop has other additional ICT and IoT devices (1 – yes, 0 – no).

The third element is the trip to school by bus. The safety assessment of this element will be performed for an entire system as a sum over all routes in system. It is expressed as below:

$$\overline{F_{B2}} = \sum_{d \in D} F_{B2}(d)$$

The assessment of this element take into account getting on and off a bus and bus journey. The value of this element is weighted by the number of children traveling on the $d$-th route ($x_{atr}(d)$) in realtion to

all travelers using bus in a system ($x_{as}$). This function has been written as:

$$\forall d \in D \quad F_{B2}(d) = \frac{x_{atr}(d)}{x_{as}} \cdot \begin{bmatrix} q_b(1)\left(1 - \frac{\sum\limits_{w \in \Psi(d)} \text{sgn}(w)}{\sum\limits_{d \in D}\sum\limits_{w \in \Psi(d)} \text{sgn}(w)}\right) + q_b(2)\left(1 - \frac{\sum\limits_{(i,i') \in \Phi(d)} l(i,i')}{\sum\limits_{d \in D}\sum\limits_{(i,i') \in \Phi(d)} l(i,i')}\right) + \\ q_b(3) \cdot \frac{\sum\limits_{(i,i') \in \Phi(d)} V_{sr}(i,i')}{|\Phi(d)|} + q_b(4) \cdot [c_{los}(i,i')] + \\ + \sum\limits_{s \in S} y_a(d,s) \cdot \begin{bmatrix} q_{st}(1)c_{stpasy}(s) + \\ q_{st}(2)c_{stos}(s) + \\ q_{st}(3)c_{sttel}(s) \end{bmatrix} \end{bmatrix}$$

This function includes elements related to a bus route and the transport means itself. The bus route assessment includes the number of stops ($w$) visited on the $d$-th bus route; the length of the $d$-th bus route consisting of connections $l(i,i')$ belonging to the set of connections $\Phi(d)$ on the $d$-th bus route, the average speed on the connection $V_{sr}(w,w')$ and the share of the length of the illuminated road. On the other hand, the assessment related to the characteristics of the s-th vehicle (the parameter $y_a(d,s)$ - determines the assignment of the $s$-th vehicle to the $d$-th route) includes:

- $c_{stpasy}(s)$ – the presence of seat belts and necessity to fasten seat belts on a bus,
- $c_{stos}(s)$ – the additional lighting of a bus
- $c_{sttel}(s)$ – the level of integration of a bus with telematics systems.

## 5. Verification of the model based on the distribution of the traffic flow in VISUM

The practical application of the approach to the issues of managing the mobility of school children presented in the paper is shown on the example of a real transport system. The PTV VISUM was used to carry out the analyzes.

Verifying the correctness of the presented child mobility management system in school-age with the use of the Smart Bus Stops and radio transmitters has been carried out through a comparison of the system performance indicators with the current system without additional support. Two comparative variants were used to simulate the effectiveness of the school-age mobility management system using smart bus stops and radio transmitters.

In the first variant, both systems are compared, considering the maximum flow of children traveling to school. A time interval from 7:00 AM. to 8:00 AM has been assumed in the first variant. During this time, the highest traffic levels within the school area are generated. The second variant was simulated for the lowest traffic situations to check how both systems behave in two extreme road situations. In

a second variant, the assumed simulation period is from 11:00 AM to 12:00 PM.

Firstly, the transport network of the analyzed research area should be defined to verify the child mobility management system. The transport network, including point and line elements, is shown in Fig. 9 (point elements represent intersections, while the sections have an interpretation of connections between individual intersections). The analyzed area is the municipal community of Radzyń Podlaski, located in the Lubelskie Voivodeship in Poland.

The division of the area into traffic areas generating sources of traffic flows between the place of residence of children and the school is shown in Fig. 10 (region 28 marked red - location of the destination point – a school).
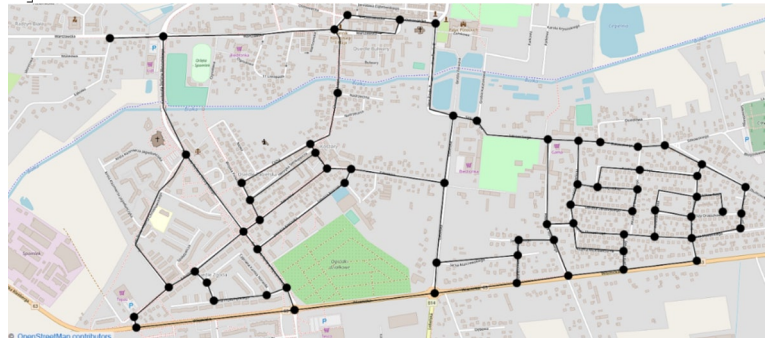


Fig. 9.  Transport network in the analyzed child mobility management system.
Source: own study based on PTV Visum
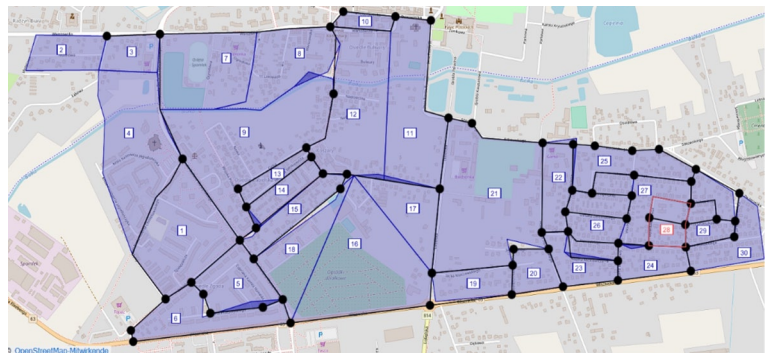


Fig. 10.  Communication areas in the territory of the child mobility management system
Source: own study based on PTV Visum

In order to determine the effectiveness of a child mobility management system for public transport (school bus), including Smart Bus Stops, the parameter of the distance between the starting bus stop and the child's place of residence was defined - Table 4. This parameter determines the total travel time of the child from leaving the home to arrival at school.

*Table. 4. Distance between home and bus stop*

| Area No. | Distance to bus stop [km] | Area No. | Distance to bus stop [km] | Area No. | Distance to bus stop [km] | Area No. | Distance to bus stop [km] | Area No. | Distance to bus stop [km] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,25 | 7 | 0,22 | 13 | 0,14 | 19 | 0,42 | 25 | 0,24 |
| 2 | 0,3 | 8 | 0,12 | 14 | 0,16 | 20 | 0,14 | 26 | 0,11 |
| 3 | 0,12 | 9 | 0,23 | 15 | 0,12 | 21 | 0,35 | 27 | 0,23 |
| 4 | 0,23 | 10 | 0,34 | 16 | 0,11 | 22 | 0,24 | 28 | 0,14 |
| 5 | 0,24 | 11 | 0,23 | 17 | 0,34 | 23 | 0,14 | 29 | 0,12 |
| 6 | 0,3 | 12 | 0,12 | 18 | 0,34 | 24 | 0,33 | 30 | 0.14 |

*Source: own study based on PTV Visum*

Fig. 11. *Location of Smart Bus Stops in the transport network.*
*Source: own study based on PTV Visum*



Fig. 12. *The flow of passengers in the morning hours in the Smart Bus Stops system*
*Source: own study based on PTV Visum*

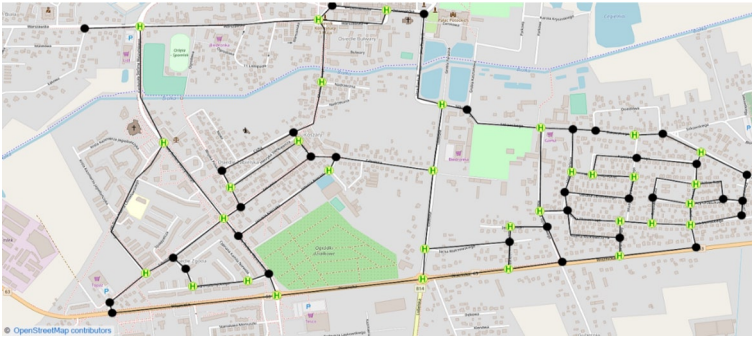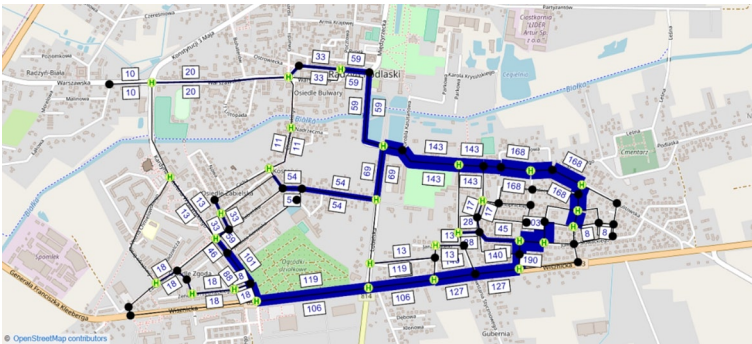Verification of a child mobility management system with the use of Smart Bus Stops and public transport was simulated for 29 intelligent bus stops - Fig. 11.

In order to operate all bus stops, four bus lines have been designated covering all Smart Bus Stops. The hourly volume of the passenger flow on individual sections of the line is presented in Fig. 12.

In the analyzed case study, the community of Radzyń Podlaski has 4 school buses and each bus is assigned to one line. The efficiency indicators in a child mobility management system including Smart Bus Stops are presented in Table 5.

In order to operate all bus stops in the morning, three bus lines covering all Smart Bus Stops have been designated. The hourly volume of the passenger flow on individual sections of the line is presented in Fig. 13. The efficiency indicators in a child mobility management system, including Smart Bus Stops, are shown in Table 6.

For the presented simulations of bus journeys, a safety assessment was carried out in a child mobility management support system. The assessment was carried out based on the developed mathematical model. The following assumptions have been made:

– only bus trips were considered, taking into account access to a bus, the modal split with passenger vehicles, pedestrian trips or bicycles was not taken into account, it will be topic of further research,

– the influence of speed was omitted due to the equal speed of 50 kmph in each section,

– the assessment related to the bus equipment was omitted because the same buses were adopted on each line,

Table. 5. *The effectiveness indicators of a child mobility management system including Smart Bus Stops for the morning hours*

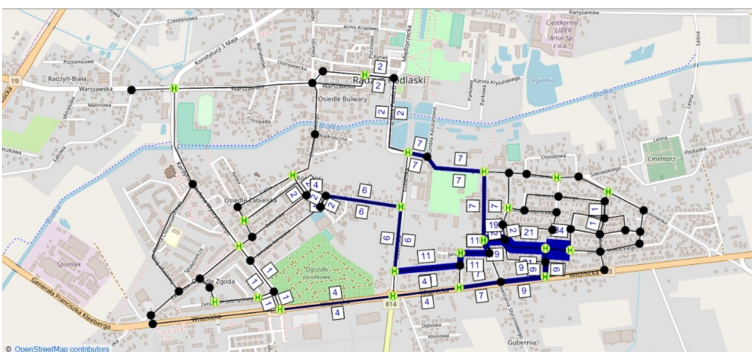| Line No. | Travel time [h] | Length of route [km] | Fuel consumption [l] | Total travel time [h] | Total fuel consumption [l] |
|---|---|---|---|---|---|
| 1 | 0,21 | 7 | 3,01 | | |
| 2 | 0,15 | 5 | 2,15 | 0,77 | 9,89 |
| 3 | 0,19 | 5 | 2,15 | | |
| 4 | 0,22 | 6 | 2,58 | | |

*Source: own study based on PTV Visum*



Fig. 13. *The flow of passengers in the forenoon hours in the Smart Bus Stops system*
*Source: own study based on PTV Visum*

– due to the nature of a transport network, it was assumed that there are either sidewalks or wide shoulders along with the connections (in the analysis of access to a bus stop).

For the assumptions mentioned above and weights presented in Table 7, three phases of the journey were assessed. The first phase is accessing a bus stop, the second phase is waiting at a bus stop, and the third phase is the bus journey. The results of an assessment are presented in Table 8. The obtained results are weighted by the number of students using a bus on their way to school to get the absolute value of an assessment suitable for comparing the adopted solutions at different times of the day. It should be noted that an assessment of the systems is very similar, which is also the result of the high quality of a transport network and bus stops. An assessment could give completely different results in communities or cities with worse road conditions and poorer additional infrastructure.

Table. 6. *The effectiveness indicators of a child mobility management system including Smart Bus Stops for the forenoon hours*

| Line No. | Travel time [h] | Length of route [km] | Fuel consumption [l] | Total travel time [h] | Total fuel consumption [l] |
|---|---|---|---|---|---|
| 5 | 0,11 | 4 | 1,72 | | |
| 6 | 0,05 | 3 | 1,29 | 0,25 | 4,3 |
| 7 | 0,09 | 3 | 1,29 | | |

*Source: own study based on PTV Visum*

Table. 7.  The effectiveness indicators of a child mobility management support system, including Smart Bus Stops

| Weight | Description | Points |
|---|---|---|
| $q_l(1)$ | points allocated for the presence of a sidewalk at a given connection | 5 |
| $q_l(2.1)$ | points allocated for the wide shoulder, it was assumed that there is no sidewalk in this case | 2 |
| $q_l(3)$ | points allocated for the lighting of a given connection | 3 |
| $q_p(1)$ | points allocated for the presence of the bus stop turnout | 5 |
| $q_p(2)$ | points allocated for the presence of the bus stop shelter | 2 |
| $q_p(3)$ | points allocated for the presence of the bus stop lighting | 3 |
| $q_p(4.1)$ | points allocated for the presence of the intelligent lighting activated by the child's transmitter | 3 |
| $q_p(4.2)$ | points allocated for the presence of the road signs and the light signals, vertical and horizontal on the road | 2 |
| $q_b(1)$ | points allocated for the number of stops on the bus stops during given route, the fewer stops, the higher the score | 4 |
| $q_b(3)$ | points allocated for the speed on given route | 4 |
| $q_b(4)$ | points allocated for the road lighting | 2 |

Source: own study

Table. 8.  The assessment of the safety of a children's mobility system using Smart Bus Stops for the morning and forenoon hours

| General data | | | | | | | |
|---|---|---|---|---|---|---|---|
| | In the morning | | | | In the forenoon | | |
| Demand | 385 | | | | 49 | | |
| Number of lines | 4 | | | | 3 | | |
| Line No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Line length [km] | 7 | 5 | 5 | 6 | 4 | 3 | 3 |
| Lighting | 0,89 | 0,76 | 1 | 0,98 | 0,89 | 0,98 | 0,96 |
| Number of children | 107 | 81 | 119 | 78 | 15 | 17 | 17 |
| Number of bus stops | 6 | 4 | 7 | 4 | 4 | 7 | 4 |
| **Safety assessment** | | | | | | | |
| $\overline{F_{B1}}$ - access to a bus stop (max. 10) | 5,51 | | | | 5,60 | | |
| $\overline{F_{Bp}}$ - waiting at the bus stop (max. 15) | 8,85* / 12,49** | | | | 8,86* / 12,82** | | |
| Journeys of each line | 2,06 | 1,66 | 2,42 | 1,65 | 2,18 | 2,21 | 2,61 |
| $\overline{F_{B2}}$ - summary trip assessment (max. 10) | 7,78 | | | | 6,99 | | |
| $\overline{F_{BA}}$ - summary assessment (max. 45) | 22,15* / 25,78** | | | | 21,45* / 25,41** | | |
| Percentage share (from the maximum possible points) | 49%* / 57%** | | | | 48%* / 56%** | | |

*without Smart Bus Stops / ** with Smart Bus Stops

Source: own study based on PTV Visum

As the developed example of pilot studies shows, in the assessed system, greater emphasis should be placed on increasing children's safety during access to a bus stop. On the other hand, an assessment indicated that the system obtained just over 50% of possible points (in the variant with Smart Bus Stops). However, it should be noted that it is impossible to get all possible points, and the maximum number of points represents the ideal situation. The presented research is at the conceptual level, and due to the costs and time of implementation of the architecture discussed in the paper, it was necessary to conduct simulation studies. The second factor forcing the use of simulation tests at this stage is the difficulty in implementation due to administrative barriers. The preliminary results allow stating that the proposed considerations are effective and may be the basis for further research, both in domestic and international conditions. The developed assessment model can support policymakers in managing child mobility.

## 6. Summary and conclusions

Children's safety plays a critical role in the area of children's mobility, especially from the parent's perspective. In times of many threats, not only related to road traffic, but also in criminal related situations (kidnapping, persecution, etc.), the importance of technology is unquestionable. However, it should be noted that the use of tracking technology may also limit the privacy rights and may also be a subject of cybercrime. The use of any supervision in online technology must therefore be preceded by detailed research and fully secured against unauthorized access. This does not however apply to the use of intelligent infrastructure, such as smart bus stops, which communicate with transmitters or children's smartphones. It should be clearly stated that the prevalence of advanced technologies has great potential in terms of improving the children safety level.

The research presented in the article was aimed at developing the concept of children mobility management system and providing a tool for assessing children's travel to school by school bus. The

verification presented in the article was to present the correctness and usefulness of the developed method. Due to the difficulty in obtaining detailed and reliable data essential to feed the method, it was necessary to carry out calculations for an isolated fragment of children's mobility system. An approach to children's mobility management system was presented and the use of the developed mathematical model applied to children's travel to school in a selected municipality. The use of the developed model for the safety assessment allows for the identification of the existing mobility system, but also for the assessment of other options for its modification and system operation under various conditions. In addition, the design of the model makes it possible to compare different systems of various complexity.

Further work in the practical aspect using the presented approach will require the acquisition of accurate data, collected over a longer period of time, for various municipalities and schools. This will allow for a proper assessment of the current state and the possibility of proposing concepts to improve the mobility system and the level of safety on a larger scale. In the research aspect, further work will consist in extending the model with additional ways of transporting children to school, as well as research in urban areas. Moreover, it will be possible to implement the analytical model in the simulation environment. This will allow for quick variations of different scenarios to further develop the mobility system. Such model can be made, for example, in the Flexsim or Anylogic environment.

## References

1. Belin M-Å, Tillgren P, Vedung E. Vision Zero – a road safety policy innovation. International Journal of Injury Control and Safety Promotion 2012; 19(2): 171–179, https://doi.org/10.1080/17457300.2011.635213.
2. Bina M, Confalonieri F, Abati D et al. Analysis of traffic upon school departure: Environment, behaviour, well-being and risk factors for road crashes. Journal of Transport & Health 2021; 22: 101119, https://doi.org/10.1016/j.jth.2021.101119.
3. Chaudhry S K, Elumalai S P. Active and passive transport choice behavior for school students and their exposure to different transportation modes. Recent Advances and Emerging Issues in Transport Research – An Editorial Note for the Selected Proceedings of WCTR 2019 Mumbai 2020; 48: 2916–2928, https://doi.org/10.1016/j.trpro.2020.08.191.
4. Davison K K, Werder J L, Lawson C T. Children's active commuting to school: current knowledge and future directions. Preventing chronic disease 2008; 5(3): A100.
5. Eguizábal S E, Berodia J L M, Portilla Á I, Ponce J B. Optimization model for school transportation design based on economic and social efficiency. Transport Policy 2018; 67: 93–101, https://doi.org/10.1016/j.tranpol.2018.01.015.
6. Elliott D, Keen W, Miao L. Recent advances in connected and automated vehicles. Journal of Traffic and Transportation Engineering (English Edition) 2019; 6(2): 109–131, https://doi.org/10.1016/j.jtte.2018.09.005.
7. Ghareeb M, Ghamlous A, Hamdan H et al. Smart bus: A tracking system for school buses. 2017 Sensors Networks Smart and Emerging Technologies (SENSET), 2017: 1–3, https://doi.org/10.1109/SENSET.2017.8125055.
8. Guerrero-ibanez J A, Zeadally S, Contreras-Castillo J. Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies. IEEE Wireless Communications 2015; 22(6): 122–128, https://doi.org/10.1109/MWC.2015.7368833.
9. Hussein A, García F, Armingol J M, Olaverri-Monreal C. P2V and V2P communication for Pedestrian warning on the basis of Autonomous Vehicles. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2016: 2034–2039, https://doi.org/10.1109/ITSC.2016.7795885.
10. Izdebski M, Jacyna-Gołda I, Gołda P. Minimisation of the probability of serious road accidents in the transport of dangerous goods. Reliability Engineering & System Safety 2022; 217: 108093, https://doi.org/10.1016/j.ress.2021.108093.
11. Jacyna M, Gołębiowski P, Szczepański E. City transport service model taking into account different means of transport. Proceedings of 19th International Scientific Conference Transport Means. Kaunas, Lithuania: Publishing House "Technologija, 2015: 160–168.
12. Jacyna M, Semenov I. Models of vehicle service system supply under information uncertainty. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22(4): 694–704, https://doi.org/10.17531/ein.2020.4.13.
13. Jacyna M, Wasiak M, Jachimowski R et al. The Concept of EPLOS Database of the Transport Infrastructure. In Kersys R (ed): Transport Means 2019. Sustainability: Research and Solutions. Proceedings of 23rd International Scientific Conference, Publishing House "Technologija": 2019: 1250–1255.
14. Jacyna M, Wasiak M, Lewczuk K et al. Decision problems in developing proecological transport system. Annual Set The Environment Protection 2018; 20: 1007–1025.
15. Jankowska D M. Integrated system for safe transportation of children to school - with the use of intelligent transport systems (ITS). Journal of KONES 2008; Vol. 15, No. 2: 137–144.
16. Jankowska-Karpa D, Wnuk A. Strategie planowania podróży, jako narzędzie poprawiające bezpieczeństwo dzieci w ruchu drogowym. Transport Samochodowy 2015; z. 2: 65–88.
17. Kamal M, Atif M, Mujahid H et al. IoT Based Smart City Bus Stops. Future Internet 2019, https://doi.org/10.3390/fi11110227.
18. Kattan L, Tay R, Acharjee S. Managing speed at school and playground zones. Accident Analysis & Prevention 2011; 43(5): 1887–1891, https://doi.org/10.1016/j.aap.2011.04.009.
19. Kumari A, Deepshikha, Satish B A. Automated traffic control for pedestrian safety. 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2017: 145–149, https://doi.org/10.1109/ICIMIA.2017.7975589.
20. Lewczuk K, Kłodawski M. Logistics information processing systems on the threshold of IoT. Zeszyty Naukowe. Transport / Politechnika Śląska 2020; 107: 85–94, https://doi.org/10.20858/sjsutst.2020.107.6.
21. Mandic S, Barra S L de la, Bengoechea E G et al. Personal, social and environmental correlates of active transport to school among adolescents in Otago, New Zealand. Journal of Science and Medicine in Sport 2015; 18(4): 432–437, https://doi.org/10.1016/j.jsams.2014.06.012.
22. Mårtensson F, Nordström M. Nordic child friendly urban planning reconsidered. In Bishop K, Corkery L (eds): Designing cities with children and young people, 1st edition. New York, Routledge: 2017: 36–46.
23. Mehdizadeh M, Nordfjaern T, Mamdoohi A R, Mohaymany A S. The role of parental risk judgements, transport safety attitudes, transport priorities and accident experiences on pupils' walking to school. Accident Analysis & Prevention 2017; 102: 60–71, https://doi.org/10.1016/j.aap.2017.02.020.
24. Nasrudin N, Nor A R M. Travelling to School: Transportation Selection by Parents and Awareness towards Sustainable Transportation. Procedia Environmental Sciences 2013; 17: 392–400, https://doi.org/10.1016/j.proenv.2013.02.052.

25. Raj J T, Sankar J. IoT based smart school bus monitoring and notification system. 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2017: 89–92, https://doi.org/10.1109/R10-HTC.2017.8288913.

26. Roblek V. The smart city of Vienna. In Anthopoulos L (ed): Smart City Emergence, Elsevier: 2019: 105–127, https://doi.org/10.1016/B978-0-12-816169-2.00005-5.

27. Rosenberg D E, Sallis J F, Conway T L et al. Active Transportation to School Over 2 Years in Relation to Weight Status and Physical Activity. Obesity 2006; 14(10): 1771–1776, https://doi.org/10.1038/oby.2006.204.

28. Rudyk T, Szczepański E, Jacyna M. Safety factor in the sustainable fleet management model. Archives of Transport 2019; 49(1): 103–114, https://doi.org/10.5604/01.3001.0013.2780.

29. Sadeghian P, Håkansson J, Zhao X. Review and evaluation of methods in transport mode detection based on GPS tracking data. Journal of Traffic and Transportation Engineering (English Edition) 2021; 8(4): 467–482, https://doi.org/10.1016/j.jtte.2021.04.004.

30. Shaaban K, Bekkali A, Hamida E B, Kadri A. Smart tracking system for school buses using passive RFID technology to enhance child safety. Journal of Traffic and Logistics Engineering; 1(2): 191–196, https://doi.org/10.12720/jtle.1.2.191-196.

31. Shinde P A, Mane Y B. Advanced vehicle monitoring and tracking system based on Raspberry Pi. 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), 2015: 1–6, https://doi.org/10.1109/ISCO.2015.7282250.

32. Soczówka P, Klos M, Zochowska R, Sobota A. An analysis of the influence of travel time on access time in public transport. Scientific Journal of Silesian University of Technology. Series Transport 2021; 111: 137–149, https://doi.org/10.20858/sjsutst.2021.111.12.

33. Stark J, Frühwirth J, Aschauer F. Exploring independent and active mobility in primary school children in Vienna. Journal of Transport Geography 2018; 68: 31–41, https://doi.org/10.1016/j.jtrangeo.2018.02.007.

34. Sun D, El-Basyouny K, Ibrahim S, Kim A M. Are school zones effective in reducing speeds and improving safety? Canadian Journal of Civil Engineering 2018; 45(12): 1084–1092, https://doi.org/10.1139/cjce-2018-0060.

35. Świderski A, Jóźwiak A, Jachimowski R. Operational quality measures of vehicles applied for the transport services evaluation using artificial neural networks. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(2): 292–299, https://doi.org/10.17531/ein.2018.2.16.

36. Thibaud M, Chi H, Zhou W, Piramuthu S. Internet of Things (IoT) in high-risk Environment, Health and Safety (EHS) industries: A comprehensive review. Decision Support Systems 2018; 108: 79–95, https://doi.org/10.1016/j.dss.2018.02.005.

37. Transportation Research Board. The Relative Risks of School Travel: A National Perspective and Guidance for Local Community Risk Assessment – Special Report 269. Washington, DC, The National Academies Press: 2002. doi:10.17226/10409, https://doi.org/10.17226/10409.

38. Wang H, Morgan C, Li D et al. Children's fear in traffic and its association with pedestrian decisions. Journal of Safety Research 2021; 76: 56–63, https://doi.org/10.1016/j.jsr.2020.11.010.

39. Westman J, Friman M, Olsson L E. What Drives Them to Drive?—Parents' Reasons for Choosing the Car to Take Their Children to School. Frontiers in Psychology 2017, https://doi.org/10.3389/fpsyg.2017.01970.

40. Yongjun Z, Xueli Z, Shuxian Z, shenghui G. Intelligent transportation system based on Internet of Things. World Automation Congress 2012, 2012: 1–3.

41. Travel and Environmental Implications of School Siting. U.S. Environmental Protection Agency: 2003.

42. Dzieci w wieku 0-14 lat w ruchu drogowym dla okresu od 2010 do 2019 roku. Polskie Obserwatorium Bezpieczeństwa Ruchu Drogowego: 2020.

43. Global Status Report on Road Safety 2018. Geneva, World Health Organization: 2018.

44. Integrated system for safe transportation of children to school(SAFEWAY2SCHOOL). [https://cordis.europa.eu/project/id/233967].
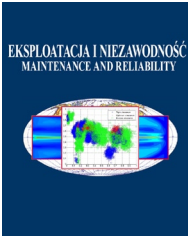
# Volumetric wear characteristics as a result of the tribological interaction between the soil with working parts cultivator's and plough's

Indexed by:
Web of Science Group

Oleksandr Vrublevskyi [a], Jerzy Napiórkowski [a], Klaudia Olejniczak [a], Jarosław Gonera [a]

[a] University of Warmia and Mazury in Olsztyn, Department of Vehicle and Machine Construction and Operation, ul. Oczapowskiego 11, 10-719 Olsztyn, Poland

## Highlights

- An original method for analysing the local wear of operating parts materials
- Identification of the surfaces subject to intensive wear under particular conditions
- Assessment of the wear process using the local volumetric wear coefficient
- Development of a method for scanner application and validation in tribology

## Abstract

This paper is concerned with the possibility of applying modern non-contact methods for assessing the wear as a result of tribological interaction between working bodies and the soil. An original method for wear testing using the test space discretization based on the 3D scanning technology was employed. A localized volumetric wear coefficient was proposed, allowing for wear analysis and improving the accuracy of the Holm-Archard model. The coefficient of local volumetric wear shows the influence of the nominal shape and the slip trajectory of the abrasive particle along the elementary surface on the intensity of wear. At local volumetric wear coefficient > 0.3, this factor determines the intensity of surface wear. Volumetric wear characteristics are the basis for prediction of wear consequences for different materials and techniques of reinforcement of working surfaces, subject to intensive wear in abrasive soil mass. The reliability of the study is confirmed by the comparison with the mass method for wear assessment and the results of the application of the proposed method for different conditions of abrasive wear of operating parts.

## Keywords

soil, abrasive wear, 3D scanning, operating parts, chisels, volumetric characteristics.

## Nomenklature

| | | | | |
|---|---|---|---|---|
| $A$ | cross sectional area, mm$^3$ | | $W$ | wear intensity, mm$^3$/s |
| $F$ | force, N | | $\alpha$ | angle of the wedge part, º |
| $H$ | hardness, N/mm$^2$ | | $\rho$ | Density, g/cm$^3$ |
| $K_v$ | local volumetric wear coefficient | | $\frac{\Delta V}{\Delta x}$ | wear rate, mm$^3$/mm |
| $L$ | friction distance, m | | | |
| $l$ | sample length, mm | | | Indexes |
| $m$ | mass, g | | $i$ | numer of volume |
| $S$ | sliding distance, mm | | $f$ | friction |
| $X, Y, Z$ | axis | | $N$ | normal |
| $V$ | volume, mm$^3$ | | $n$ | nominal |
| $v$ | speed, m/s | | $w$ | after test |

E-mail addresses:   O. Vrublevskyi (ORCID: 0000-0002-5871-6381): aleksander.wroblewski@uwm.edu.pl, J. Napiórkowski (ORCID: 0000-0003-2953-7402): jerzy.napiorkowski@uwm.edu.pl, K. Olejniczak: klaudia.olejniczak@uwm.edu.pl, J. Gonera (ORCID: 0000-0001-7758-2684): jaroslaw.gonera@uwm.edu.pl

## 1. Introduction

It is commonly believed, both in practical and scientific terms, that the selection of appropriate construction materials helps achieve adequate operational reliability of machines. At the same time, a significant aspect associated with the analysis, i.e. the effect of varying properties of the abrasive soils mass being processed on the operating part durability depending on the operating part type, appears to be ignored. This is why the increasingly modern material solutions for operating parts proposed nowadays still result in no expected increase in durability, corresponding to the economic expenditure incurred and lead to failures and thus to machinery downtime, resulting in user dissatisfaction [15].

The operating parts that process soil mass are characterised by the highest wear intensity of all machinery parts [14]. Tribological wear, which results from friction processes leads to physical, mechanical, and chemical interactions between the surface layers of tool and machinery moving parts [22]. The properties of the surface layer of the operating part subjected to friction and the way of interaction in the tribological pair of material - abrasive environment, have an effect on the intensity of this process. The intensity of a soil-processing cutting tool wear is determined by the abrasive soil mass properties, the operating part's characteristics, and the interaction mechanics occurring in the tribological pair [19, 20]. The rate of these changes is determined by randomly varying environmental conditions.

Steels have been, and in most cases will remain, the main construction material for operating parts to process a soil mass. On the one hand, this is due to the relationship between manufacturing costs and the durability of wear-resistant materials, while on the other, to the versatility of use, good weldability, ease of machining and their constantly improving mechanical properties [13, 17]. Based on the available literature, it can be concluded that the topic of operating parts' wear in the soil is still relevant, which is reflected in the large number of published scientific papers. In particular, they present the possibilities for increasing the durability of these parts by using materials with increased resistance to abrasive wear with specific chemical and mechanical properties. In addition to hard faced materials, alloy and micro-alloyed steels with the addition of niobium and boron are dominant among them [4, 7]. The available literature includes scientific papers on the modification of the chemical composition of boron steel, or on the application of specific thermo-plastic processing conditions to their manufacture in order to obtain the required functional properties [12].

Although the issue of intensive wear in an abrasive soil mass concerns many industrial sectors, the results most often presented in the literature are those of research in the field of agricultural technology. Yazici [24] conducted a study on the reduction in wear of ploughshares manufactured by hot stamping and hardfacing. According to his study results, hot stamping and hardfacing with C-Cr-based alloys can be recommended as an effective solution to reduce ploughshare wear. A paper by Bayhan [3] presents a study into increasing wear resistance by coating ploughshare chisels manufactured from low-alloy steels with three different hardfacing electrodes designated EH-600, EH-350, and EH-14Mn. The study noted statistically significant differences in wear intensity under laboratory and field conditions. Having taken costs into account, it was concluded that hardfacing with the EH-600 and EH-350 electrodes had a measurable economic effect. Horvat et al. [10] presented a comparison of the wear of mouldboards and two ploughshares manufactured from different materials and hardfaced by the SMAW process. The change in geometry and the weight loss were smaller for both hardfaced ploughshare types as compared to the original ploughshare. A study by Bialobrzeska and Kostencki [5] compared the results of ploughshare wear tests by the field method with those obtained by the laboratory method for selected low-alloy steels with the addition of boron. They conducted both field and laboratory tests to determine the highest and lowest resistance to abrasive wear. Napiórkowski et

al. [18] demonstrated that abrasive wear resistance of silicon carbide was determined by the particle size distribution in the soil being processed. Silicon carbide achieved the highest resistance in a light soil. The hardfaced layer exhibited higher abrasion resistance under different soil conditions. Silicon carbide on the nitride bond exhibited higher abrasion resistance than that of the analysed boron steels in all soil types.

A commonly applied method for assessing wear is the determination of weight loss and the change in linear dimensions of the operating part [5, 10, 24]. Due to the different weights of parts and, thus, their different friction surface area, weight wear should be used when assessing the wear of parts of an identical design. For working elements composed of heterogeneous materials, the use of the mass method is not reliable. The progressive digitisation of surface condition monitoring processes, and the continuous development of tribological research create a demand for the development of methods for verifying changes not only in the surfaces involved in friction processes but also in the operating part's geometry.

Vrublevskyi et al. [23] proposes a new approach analysis of changes in the shape of a working component of an agricultural machine. For this purpose, chisels of plowshares used under varied soil conditions were analyzed. Numerical models of worn-out working parts created using 3D scanning were the basis for developing.

Using wear models, it is possible to forecast the rate of material loss from a given element. The classic theory of wear assumes that the assessment of the wear process begins with the identification of the hardness of the material, the intensity of material removal, the given load and the probability of the material removing the wear particle at a given contact moment [1, 2]. There are three main trends in model development. The first is based on empirical equations, the second is a mechanical approach, and the last is based on material. However, no generalization has been achieved with regard to consumption at any level. Meng [16] discovered during his research that 100 parameters were used in total in over 180 models. Important in the work [16] is the conclusion that in most works that volumetric changes in the working body are the main characteristic of the change in the functional properties of the working body and the main characteristic of wear for all models.

The first successful attempts have also been made to use 3D scanning to assess the wear process. Cucinotta et al. [6] applied 3D scanning to identify the wear of the cutting surface of the ploughshares mounted on four bodies of a soil-processing plough. They also presented a method for the development of a digital model of worn surfaces. Their study identified differences in volume losses and changes in the profile of the cutting edge between ploughshares depending on the mounting location. The literature also includes studies that present changes in the volume and shape of operating parts used in manufacturing automotive parts. Hawruluk et al. [8] conducted a study that enabled the determination, using 3D scanning, of the wear of forging dies used for manufacturing motor valves. The forging dies were tested in cycles, after a specified number of valves were manufactured. The losses in production tool material were determined, and the quality and changes in the geometry of the forging die surface layer were directly monitored. The study [9] also addressed the issue of forging tool wear. Following the performance of 3D scanning, an analysis of the tool wear course was conducted. Changes in volume in selected areas of forgings were determined following the forging of an appropriate number of components. Periodic tests enabled the identification of wear over time in relation to the reference model. For the scanning, an integral scanning system with an accuracy of 0.035 mm was used in accordance with ASME B89.4.22. The area testing method proposed by the authors does not analyse changes in the wear volume. The identification of the extent of wear of machinery operating parts using 3D scanning enables the determination of the tool wear limit value beyond which the processing of the material is already affected by defects and deteriorates the product quality [25]. An analogous problem is encountered when processing soil, e.g. in ploughing processes,

yet the wear is caused by different factors. The wear of tool operating parts results in increased energy consumption, decreased productivity and deteriorated process quality. Ranusa et al. [21] used a 3D scanner to identify the wear of hip prostheses. The authors compared the wear of parts periodically, following the performance of specified work using nominal parts. The studies considered the number of samples, process duration and the friction distance, while it is difficult to find studies into changes in volume in relation to changes in the geometry. In the current state of knowledge, it is important to select the appropriate material for parts that will undergo the least wear and ensure increased durability. The literature review shows that the wear volume can be determined, e.g. using the Holm-Archard model – in which the wear volume is determined by an indirect method from linear measurements [17]. This method, involving the determination of volume based on the measurements of the mass, is of limited accuracy due to the use of materials with varying densities.

When conducting research, much attention is paid to linking changes in the geometry of an operating part with the way it gets worn. What is particularly important in the identification of geometry changes is the cutting edge shape which most often determines whether the part reaches its limit state. The assessment of changes in the operating part shape is possible through a three-dimensional approach to the problem. Obtaining volumetric changes in the geometry of an operating part provides the basis for developing a digital spatial model. The application of volumetric characteristics will enable the identification of the most complex changes in the shape of operating parts and provide the basis for the elimination of changes in the part weight from the assessment of wear.

The aim of the study is to develop volumetric characteristics of the operating part wear process in an abrasive soil mass. Solving this problem will allow changes in the wear area to be studied not on a macro- but on a micro-scale, and the factors affecting local wear to be identified.

The study results are provided in the following order: Section 2 presents the materials and the methodology for volumetric wear testing. Section 3 presents the study results, and compares the mass method with the volumetric method. The analysis of the obtained wear characteristics is presented in Section 4. In the Conclusions, the obtained results are discussed, and the direction of further research is indicated.

## 2. Methodology of the study

The development of the volumetric wear characteristics and the prediction of wear required several test stages in a specific sequence (Fig. 1). First, a 3D scan of the nominal elements was performed. Subsequently, these elements were subjected to tribological tests, which were actually operational testings. During the tests, the decisive parameter was the type of soil and the amount of work performed. After this stage of the research was completed, the surface microstructure was assessed and the worn parts were 3D scanned. Based on the comparison of the 3D surfaces of the worn and nominal elements, the total volume of the used material was determined, as well as the elementary volumes of the test objects after discretization of the model. In addition, a verification was carried out to identify differences in consumption between the volumetric and mass methods.

Based on the obtained data, new volumetric wear characteristics and a method of predicting volumetric wear of working elements were proposed.

### 2.1. Material

The volumetric characteristics obtained as a result of the use of 3D scanning technology proposed in the work can be used to assess the wear of elements with a complex shape and elements consisting of zones of different density. The samples selected for the tests constituted the physical model. The chisels, on the other hand, are the actual working element of the plow for tilling the soil.
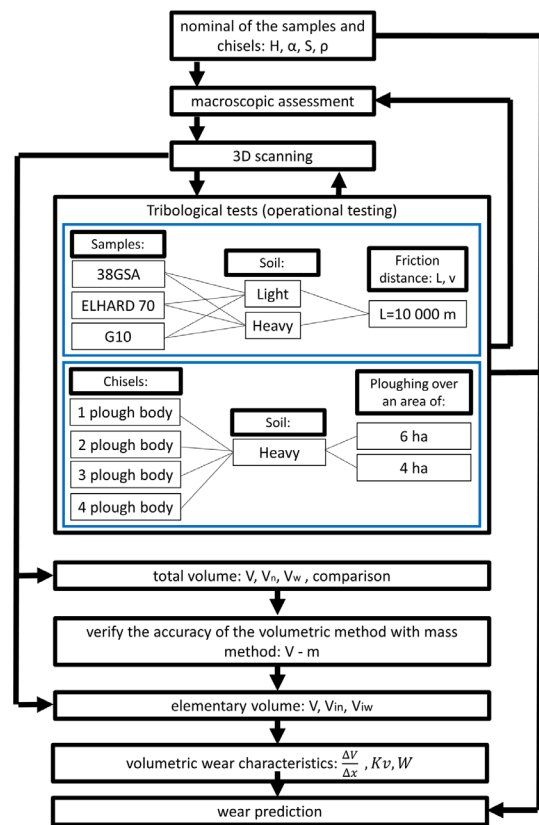


*Fig. 1. Diagram of tests performed at work*

In the first stage, the test material comprised samples made from 38GSA steel, padding weld EL HARD70, and sintered tungsten carbide G10 applied onto 38GSA steel. The nominal dimensions and shape of samples are provided in Fig. 2.

The measurements of test material hardness were conducted by the Vickers method under conditions compliant with standard PN-EN ISO 6507-1:1999. The measurements were conducted using a Zwick 32 hardness tester with a load of 9.807 N operating for 15 s. The chemical composition (Table 1) was analysed by the spectral method using a Leco GDS500A glow discharge emission analyser, with the following parameters applied: U = 1250 V, I = 45 mA, argon. For the light microscopy testing, a Zeiss Neophot 52 microscope coupled with a Visitron Systems digital camera was used. The scanning electron microscopy (SEM) testing was conducted using a JEOL JSM-5800 LV scanning microscope coupled with an Oxford LINK ISIS-300 X-ray mini-analyser. For the macroscopic assessment of the surfaces following friction tests, a KEYENCE VHX-6000 digital microscope was used.

The microstructure of the chisels was assessed using the Phanom XL scanning electron microscope. The microscope was compressed with an X-ray microanalyzer. Using the installed BSD, SEM and EDS detectors, it was possible to obtain information about the topography and chemical composition of the analyzed surface. Evaluation of the microstructure was carried out in the material contrast. Nital 5% was used to etch the chisels prior to microscopic observation. The chemical composition was assessed with the Thermo ARL Quantris spark spectrometer using the CCD technique.

The samples (Fig. 3) from low-alloy martensitic steel 38GSA were acquired directly from a steel plant. The steel was manufactured using hot rolling technology and subjected to normalisation under metallurgical conditions in order to break down the microstructure. The average hardness of the steel was 420 HV10. The microstructure comprised fine dispersed perlite with ferrite and martensite grains.
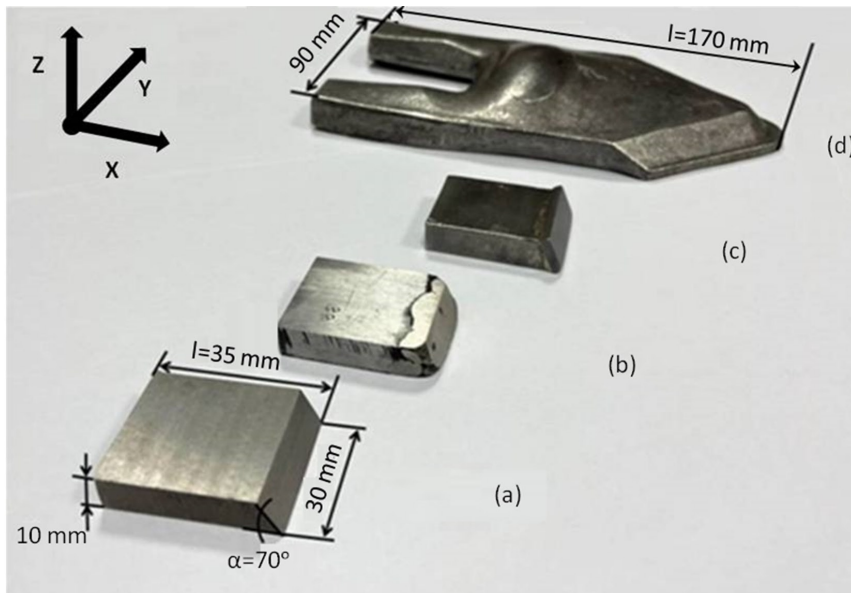
Fig. 2. The appearance of test materials: (a) steel 38GSA; (b) padding weld ELHARD70; (c) specimen with sintered carbide G10; (d) knock-on chisel

The Elhard 70 electrode (Fig. 4) was applied onto 38GSA steel. The padding weld had a ledeburite structure with large precipitates of primary chromium carbides and a few small-sized boron carbides. The padding weld hardness at the 5 mm thickness ranged from 674 HV10 (on the surface) to 871 HV10 (at a 3 mm depth).

The third material (Fig. 5) comprised samples made from 38GSA steel with sintered tungsten carbide G10 plates soldered on it. The structure comprised − α - initial tungsten carbide (WC) with a particle size ranging from 0.001 to 0.002, and α1 – a solid cobalt solution in the initial tungsten carbide (WC) that did not re-crystallise during sintering. The highest hardness measured on the carbide surface was 1423 HV10.

The operational testing was conducted on 80 mm knock-on chisels (Fig. 6). The system was developed to improve the plough's penetration ability, stabilise it, and maintain the pre-set cutting depth. The chisels were manufactured from boron steel with a martensitic structure and post-martensitic orientation, with very few carbide phase precipitates inside the martensite laths. The average hardness of the surface
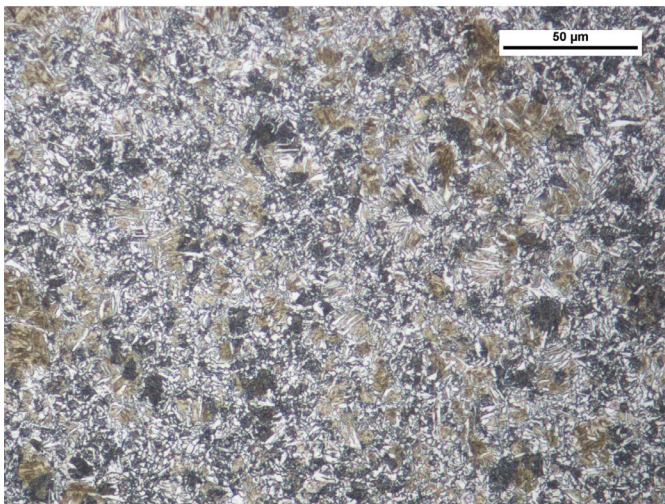


Fig. 3. Microstructure of fine-dispersion perlite with ferrite grains and martensite precipitates of 38GSA steel in the delivered condition; Magnification 500x, etched with 3% HNO3 (Mi1Fe), light microscopy



Fig. 5. Microstructure of the cemented carbide G10

layer was 560 HV. The nominal dimensions are presented in Fig. 1d, while their weight was 1000±5 g. Chisels of this type were mounted on the ploughshare using a dedicated holder.

## 2.2. Tribological tests

The operational testing of samples were conducted under the actual operational conditions in two soil types: heavy loam (designated as a heavy soil) and loamy sand (designated as a light soil). The light soil is an extremely abrasive soil, because to the high content of sand. The samples (Figs. 2a-c) were mounted on a 9-tined cultivator with spring tines, aggregated with an agricultural tractor (Fig. 7a). The samples were operated when the cultivator was operating at a depth of 0.12 m. Along the friction path equal to 5 km, the order of mounting the samples in the cultivator was changed. The average speed of the unit was 1.9 m/s, while the friction distance was 10 000 m. The soil moisture ranged from 12% for the light soil to 15% for the heavy clay, which corresponds to moist soil. The particle size distribution was tested by the laser diffraction method using a Mastersizer 2000 laser particle composition meter in accordance with standard ISO 13320 (Table 2).

The operational testing of chisel wear was conducted while ploughing at a depth of 0.25 m using a John Deere 6930 agricultural tractor with an ES 100 four-bodied reversible plough (Fig. 7b). The average tractor speed was 1.9 m/s, 8 chisels were tested. Two chisel sets were
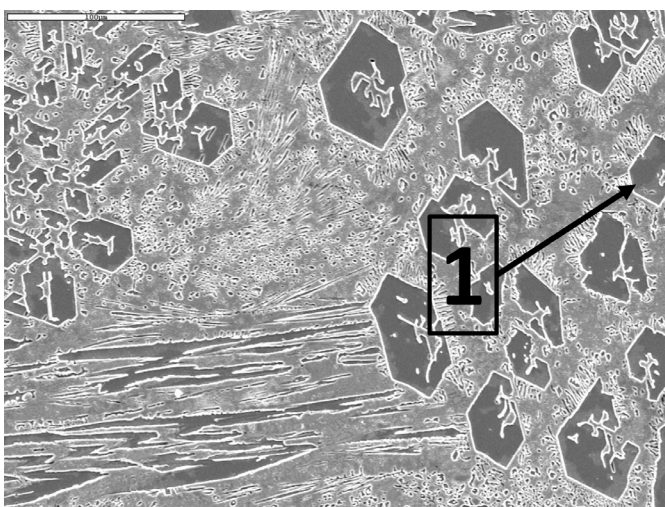


Fig. 4. Microstructure of the padded layer. Large, primary precipitates of chromium carbides (1) in the matrix of the mixture of alloy ferrite and carbides. Magnification 350x, etched with 3% HNO3 (Mi1Fe) and then electrolytically with chromic acid, SEM
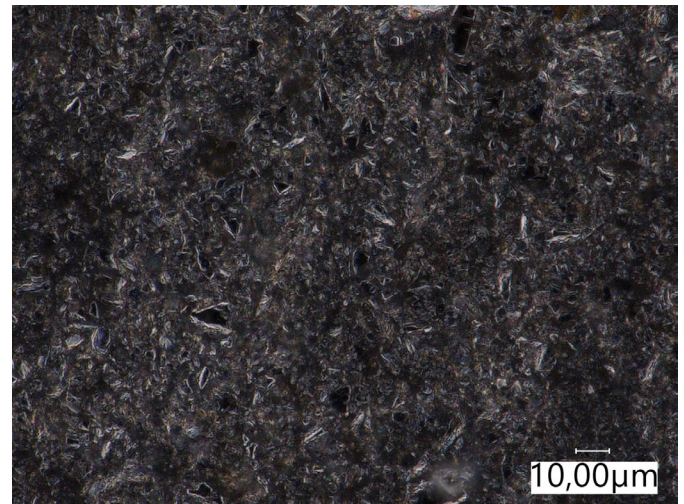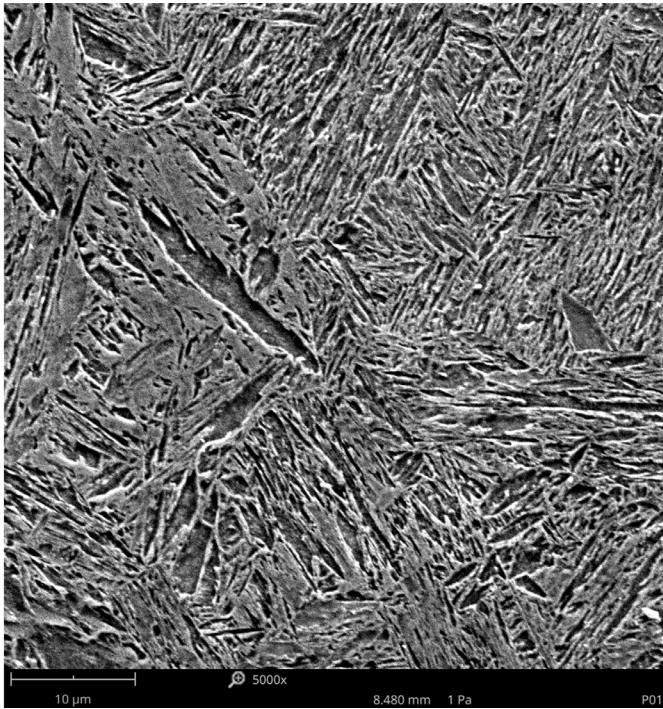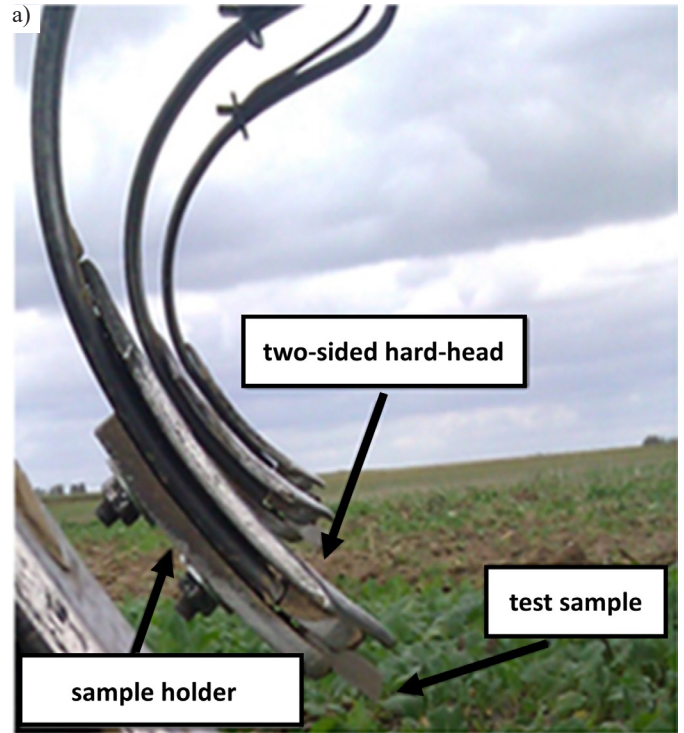
Fig. 6. Microstructure of boron steel in chisels



used for the operational testing. The first set carried out ploughing over an area of 6 ha, while the other over the area of 4 ha. The way the chisels were mounted is presented in Fig. 6b. The chisel designated with number 1 was mounted on the first body (closest to the tractor). The chisel wear tests were only conducted in heavy soil.

The weight of the test objects was measured using an AXIS B2000 balance with an accuracy of 0.1 g. For each material under specific conditions, tests were carried out with the use of five samples in order to increase the accuracy and determine the measurement error.

### 2.3. 3D scanning

Prior to the tribological testing of the analysed samples and chisels within an abrasive soil mass, the nominal geometry of the study subjects was scanned. To this end, an

Fig. 7. Method of mounting: (a) a specimen in the holder; (b) a chisel

Table 1. The chemical composition of the tested materials.

| Material | [% mass.] | | | | | | | | | Hardness [N/mm²] | Density [g/cm³] | Metallurgical structure |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C | Cr | Mn | S | Al | B | Mo | WC | Co | | | |
| Martensitic steel 38GSA | 0.35 | 1.17 | 1.07 | 1.17 | 0.022 | - | - | - | - | 420 | 7.80 | fine-dispersion perlite with ferrite grains and martensite precipitates |
| Padding EL-HARD70 | 5.0 | 38.0 | - | - | - | 3.5 | - | - | - | 871 | 6.2-6.9 | large, primary precipitates of chromium carbides in the matrix of the mixture of alloy ferrite and carbides |
| Tungsten carbide G10 | - | - | - | - | - | - | - | 94 | 6 | 1423 | 14.8 | initial tungsten carbide and a solid cobalt solution in the initial tungsten carbide that did not recrystallise during sintering |
| Chisels made of boron steel | 0.3 | 1.02 | 1.25 | 1.02 | 0.04 | 0.002 | 0.01 | - | - | 560 | 7.65 | boron steel with a martensitic structure and post-martensitic orientation, with very few carbide phase precipitates inside the martensite laths |

| Granulometric groups | Fraction diameter [mm] | Fraction content [%] | |
|---|---|---|---|
| SAND | 2.0 – 0.05 | 33.62 | 77.06 |
| SILT | 0.05 – 0.002 | 49.92 | 21.37 |
| CLAY | < 0.002 | 16.56 | 1.57 |
| Determination according to PTG 2008 | | **heavy soil** | **light soil** |

Atos Core optical 3D scanner with an accuracy of 0.02 mm, and the GOM Inspect® software were used. The scanner had a projector with a structural blue light and two cameras. The scanning (Fig. 8) of a single part required, depending on its size, between several (for the samples) and several dozen (for the chisels) individual scans to be taken, which were subsequently combined using the markers placed on the part being scanned. During the scanning, a point cloud was obtained from the surfaces of the analysed objects (triangulation). Subsequently, a 3D model was generated, from which a triangle network was prepared to obtain an STL model. The developed nominal specimen and chisel models were compared to corresponding models of worn operating parts within an abrasive soil's mass depending on the specimen type, soil conditions and the place where the chisel was mounted on a four-bodied reversible plough.



*Fig. 8. The process of 3D scanning of: a) specimen; b) chisel*

## 2.4. Volumetric characteristics

In order to obtain integral and differential volumetric characteristics of wear, the specimen was divided into elementary volumes. This division can be based on evenly distributed planes parallel to each other. The location of the plane in relation to the surface being worn can be determined by the specimen's longitudinal coordinate (Fig. 9).

The volumes $\Delta V_{in}$ (nominal) and $\Delta V_{iw}$ (after test), contained between the adjacent planes I and II, enable the determination of the volume $\Delta V_i$ of the worn material as $\Delta V_i = \Delta V_{in} - \Delta V_{iw}$. With a $\Delta x$ = 1 mm distance between the planes, the volumetric wear unit was mm³/mm.

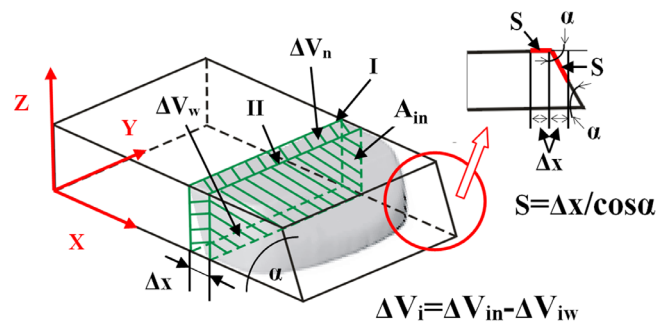Then, the total volume $V$ of the worn material was determined as:



*Fig. 9. The formation of elementary volumes on test specimen*

$$V = \sum_{i=0}^{l}(\Delta V_{in} - \Delta V_{iw}), \qquad (1)$$

where $l$ is the specimen length (Fig. 1).

With the volume of the worn material, it was possible to determine the local volumetric wear coefficient $K_v$, as a relationship between the $\Delta V_i/\Delta x$ ratio and the nominal cross sectional area $A_{i\,n}$ (Fig. 9) in the area under consideration:

$$K_{vi} = \frac{\Delta V_i}{\Delta x}\frac{1}{A_{i\,\overline{n}}}. \qquad (2)$$

The local wear coefficient defines the share of wear in the nominal elemental volume, taking into account the displacement path of the abrasive particle along the working surface. Equation (2) contains the $\Delta V_i/\Delta x$ ratio, which determines the volumetric wear rate. It is proposed that the volumetric wear rate should be defined as the ratio of the volume of the material used to the one-dimensional simplex, which determines the boundaries of the working surface under consideration. For a flat nominal surface, and assuming that the X-axis direction determines the movement of soil particles over the prismatic surface, the $\Delta x$ value can be defined as the sliding distance $S$ of a soil particle over the surface. Having considered the location of wear, the $\Delta V_i/\Delta x$ ratio has a definition similar to that used in the well-known Holm-Archard's expression: $\frac{V}{L} = K\frac{F_N}{H}$ [11]:

$$\frac{V}{L} = K\frac{F_N}{H} \qquad (3)$$

where: V - wear volume [mm3], load $F_N$ [N], friction distance $L$ [m], wear coefficient $K$ [-], and material hardness $H$ [N/m²].

Obviously, with a wedge-shaped work surface, the particle's sliding distance will be longer than $\Delta x$. In order to take into account of the trajectory of particle's movement over a wedge-shaped surface with

an α angle (Fig. 9), equation (2) should be presented in the following form:

$$K_{vi} = \frac{\Delta V_i}{S \cdot \cos\alpha} \cdot \frac{1}{A_{i\,\overline{n}}}, where \begin{cases} \alpha = 0^O \ without\ a\ wedge \\ \alpha < 90^0 \ wedge \end{cases} \qquad (4)$$

Then, for the elementary nominal volume from equations (3, 4), the normal load $F_N$ can be presented as follows:

$$F_{Ni} = H_i \cdot \frac{K_{vi} \cdot S \cdot \cos\alpha}{K \cdot L} \cdot A_{i\,\overline{n}}. \qquad (5)$$

Equation (5) enables the estimation of the load leading to a reduction in the worn volume of an operating part due to abrasive wear. The relationship takes into account the material properties (hardness), the surface geometry (volumetric wear coefficient and the cross sectional area of the operating part), and the surface quality (soil movement trajectory).

Wear models, e.g. equation 3, are often used to prediction the worn material volume. Therefore, the wear volume is the subject of direct measurement using 3D scanning technology.

## 3. Study results

After tribological tests, the surfaces of test materials were analysed in order to identify the ways they were worn under different soil conditions. When the materials were being worn in the light soil, the loosely bound abrasive particles, characterised by high freedom of movement, caused scratching and ridging on the friction surface (Fig. 10). As the fine fractions in the soil mass increased, the process of their penetration into the discontinuities of the surface layer occurred. Hence, there are numerous scratches and sand particle residues on the surface of the material being worn in the light soil. In the few cases when the possibility of movement was limited, micro-cutting took place, which resulted in chipping off the layer. It indicates that sand grains interact with the surface in a discrete manner. Only for the material from

sintered carbide G10, the surface is slightly scratched and the wear processes are hardly noticeable.

As the content of loam and dust particles in the soil increases, a different course of wear can be noted (Fig. 11). The sand fraction comprises solely quartz $SiO_2$. Dust and silt fractions contain mostly compounds of amorphous silica and silty minerals. The impact of silt and clay alone is negligible, but it intensifies when they are in combination with other fractions. When humid, these fractions act like an adhesive for quartz. Fatigue wear is evidenced by local surface tearing out resulting from multi-cycle wear, which includes elastic deformation, plastic deformations, the formation of micro-volume threshold strains characterised by a defective structure and the cutting of these micro-volumes. Numerous dents in the surface are filled with the abrasive mixture, owing to which $SiO_2$ grains have fewer degrees of freedom, therefore the share of sliding friction increases at the expense of rolling friction. The process in question is the wear caused by reinforced abrasive grains. The nature of wear changes, which results in its increased value.

When identifying the operating part wear values, the method described in subsections 2.3 and 2.4 was employed.

The cross sections were made along the X-axis (Fig. 12). The first cross section was located at a distance of 20 mm from the initial part of the chisel since no wear of the surface was noted up to a distance of 20 mm (the area where the chisel was mounted on the ploughshare). The second cross section was made at a distance of 69 mm. From the distance of 69 mm onwards, the cross sections were made at 5 mm intervals, as it was an area of accelerated wear of the chisel surfaces that varied depending on where the chisel was mounted on the four-bodied reversible plough. The greatest number of cross sections were made for chisels mounted on the fourth body due to the least wear.

In order to verify the accuracy of the 3D scanning method, measurements of both nominal and worn chisels or samples were conducted. The mass wear being measured (Table 3 and Table 4) was compared with the mass wear calculated based on the measurements of chisels or samples volumes obtained from scanning. The measurement results and the comparison are presented in Table 3 and Table 4, where: $m_M$ – chisels or samples weight loss, $V$ – the volume of chisels or samples
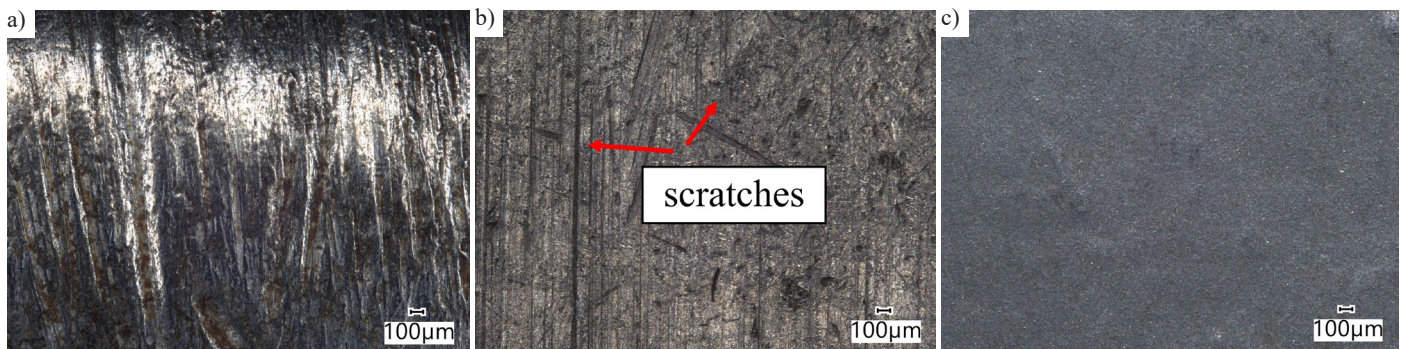


Fig. 10. The appearance of specimen surface after being worn in a light soil: (a) 38GSA steel; (b) ELHARD70; (c) sintered carbide G10



Fig. 11. The appearance of the specimen surface after being worn in a heavy soil: (a) 38GSA steel; (b) ELHARD70; (c) chisel
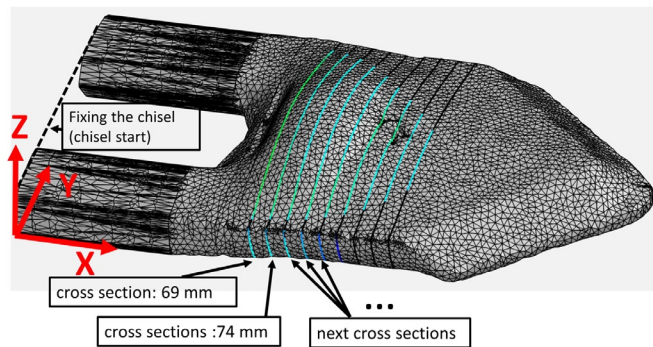
Fig. 12. Cross sections made along the X-axis

wear obtained from the 3D scanning process (eq. 1), $m_V$ – chisels or samples weight loss determined based on the scanning results.

Similarly, the verification of the proposed method was carried out by measuring the weights of nominal and worn samples (Table 3) as well as chisels (Table 4).

The differences between the measured weight loss and the weight determined by the volumetric method as a result of scanning (Δm) reach a maximum value of 1 g, which accounts for approx. 0.2% of the chisel's total weight. Such a negligible difference is indicative of the good accuracy of the applied 3D scanning method for the purpose of chisel volume determination.

An additional advantage of 3D scanning is the possibility for distinguishing the operating part's volume (Fig. 13), which enables detailed examination of local wear and the areas of operating part rein-

Table 3. Results of the specimen wear determination by the volumetric and mass methods.

| | heavy soil | | | light soil | | |
|---|---|---|---|---|---|---|
| | volumetric method (V) | | mass method (M) | volumetric method (V) | | mass method (M) |
| | V | $m_V$ | $m_M$ | V | $m_V$ | $m_M$ |
| | mm³ | g | g | mm³ | g | g |
| 38GSA | 3836.91 | 29.93 | 29.2 | 849.3 | 6.62 | 6.1 |
| ELHARD70 | 1404.31 | 10.72 | 11.6 | 616 | 4.59 | 5.6 |
| Tungsten carbide G10 | 805.28 | 6.52 | 6.7 | 286.117 | 2.31 | 2.6 |

Table 4. Results of the chisel wear determination by the volumetric and mass methods.

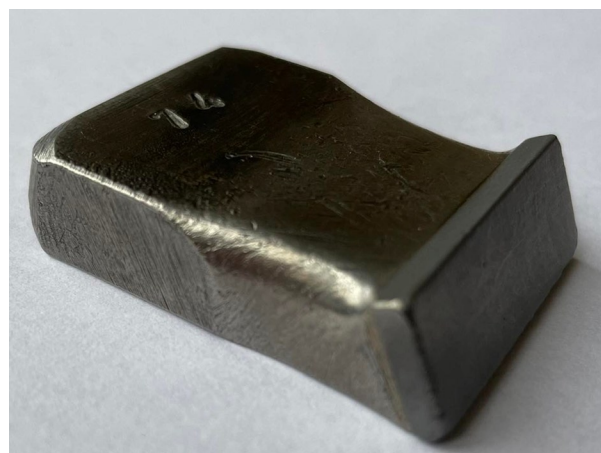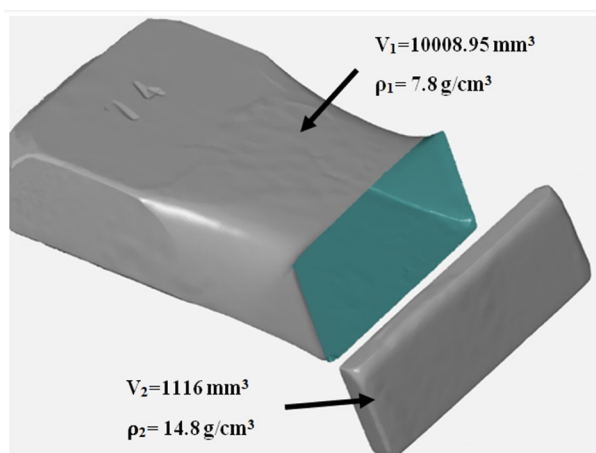| | | volumetric method (V) | | mass method (M) | |
|---|---|---|---|---|---|
| | | V | $m_V$ | $m_M$ | |
| Chisel No | | mm³ | g | g | Δm= $m_M$ – $m_V$ [g] |
| 6 ha | 1 | 91329.34 | 698.9 | 698.7 | -0.2 |
| | 2 | 82779.55 | 633.4 | 633.3 | -0.1 |
| | 3 | 77214.68 | 590.8 | 591.1 | 0.3 |
| | 4 | 65766.30 | 503.2 | 504.0 | 0.8 |
| 4 ha | 1 | 86736.96 | 663.7 | 663.8 | 0.1 |
| | 2 | 76278.64 | 583.7 | 584.0 | 0.3 |
| | 3 | 71001.02 | 543.3 | 543.4 | 0.1 |
| | 4 | 65158.89 | 498.6 | 499.6 | 1.0 |



Fig. 13. The division of volumes for the specimen with sintered carbide G10: (a) 3D scan; (b) the actual specimen with sintered carbide G10

forcement (with the least wear). Local reinforcement is a priority in increasing the operating parts' resistance to abrasive wear. Following the scanning, the volume of G10 carbide with an almost doubled density can be distinguished. Not only the material but also the shape of the operating part can determine the volume diversity.

## 4. Result analysis and discussion

On the example of a sample made of 38GSA steel and used in heavy soil, the accuracy of $K_{vi}$ determination in the area of intensive wear was +/- 0.07 (Fig. 14), while in the area with lower wear intensity, the standard deviation was +/- 0.025. On the basis of five samples, the average value of the measurements was obtained, presented in Figures 15 and 16. For the remaining materials, the consumption was also analyzed on the basis of measurements made for five samples of a given type and in two soil conditions. The average values of wear and standard deviation were lower than for the described sample.

In order to determine the influence of the cross section distance from the beginning of the sample (sample length = 0 mm) on change in the $K_{vi}$ coefficient, the analysis of variance was used. For each of the tested samples, the null hypothesis about the lack of differences between the coefficient values and the alternative hypothesis about the existence of significant differences in the coefficient changes was adopted.

ised by a transition from a prismatic wedge surface to a parallelogram surface. It is at the transition point that the change in wear intensity along the specimen length can be observed. The characteristics of the specimen with a padding weld exhibit similar properties (Fig. 15).



Fig. 15. Characteristics of local volumetric wear of samples (Figs. 1a-c)

A slightly different nature of the wear location is observed for samples reinforced on the front side of the operating surface with sintered carbide G10 (Fig. 16). For this solution, the wear of the specimen's front part is significantly reduced. The intensive wear of 38GSA steel on the prismatic surface results in a local increase in the linear trajectory $S$ of the abrasive particle $S > \Delta x$. At $K_{vi} > 0.3$, this factor determines the intensity of surface wear.

The description of the wear process results using the volumes enables a more detailed study into not only selected elementary volumes but also the surfaces forming these volumes. For example, the front and side surfaces can be selected. In this case, the shape changes the most on the side surfaces. Possible main reason for this is the way the nominal force $F_N$ and the friction force $F_f$ operate, and the particles' movement trajectory. On the front surface of the soil-attacking specimen, the friction coefficient value drops due to the inequality $F_N > F_f$, where: $F_f$ is friction force. For example, for $\Delta V_{24w}$, the total ratio is 41.125 mm³. On the side surfaces, the wear is 26.125 mm³, which accounts for 63.52% of the $\Delta V_{24w}$ value.

The characteristics presented in Fig. 17 can be used to determine the friction path $L$, at which the total wear of the elementary part of the sample $\Delta V_i$ occurs. For this purpose, after the predetermined friction path ($L = 10000$ m) at the time of scanning the samples at the speed ($v = 1.9$ m /s) of the sample displacement during the tribological test, the wear intensity for the volume $\Delta V_i$ can be determined. The wear intensity will be determined as $W = \Delta V_i \cdot v/L$. The volumetric wear intensity determines the ratio of the elementary volume of the material used to the unit of work time performed when cutting the soil. Consequently, at point C (Fig. 15) the wear intensity W will be maximum and will amount to $4.5 \cdot 10^{-5}$ mm³/s. In the section ABC wear intensity $W = 1.55 \cdot 10^{-4}$ mm³/s, which enables the determination of the friction distance $L$ up to the complete consumption of the samples. The total wear of the first elementary section of the sample (closest to the working edge) will take place after the friction path $L$ of 364.67 m has been traversed. However, the total wear of the specimen wedge surface will take place after the friction path of 4360.88 m has been traversed.
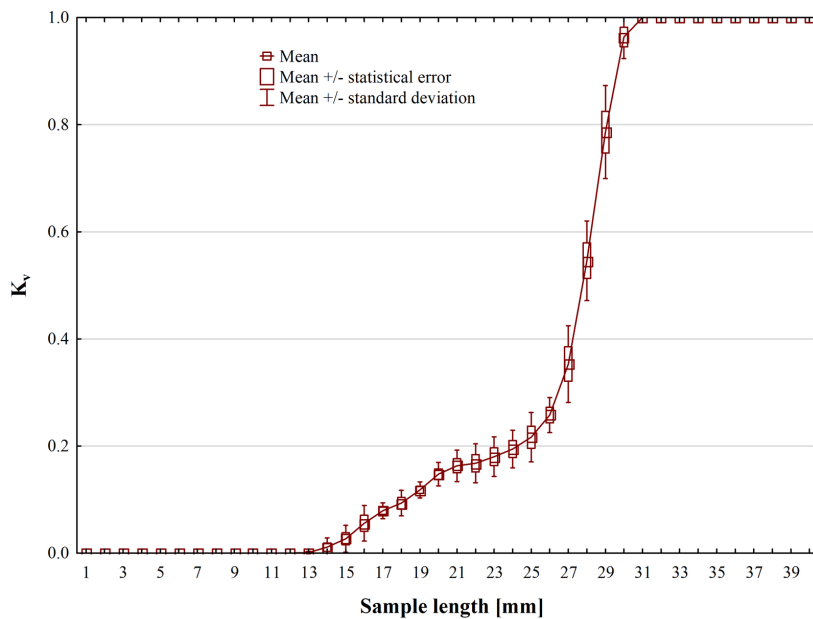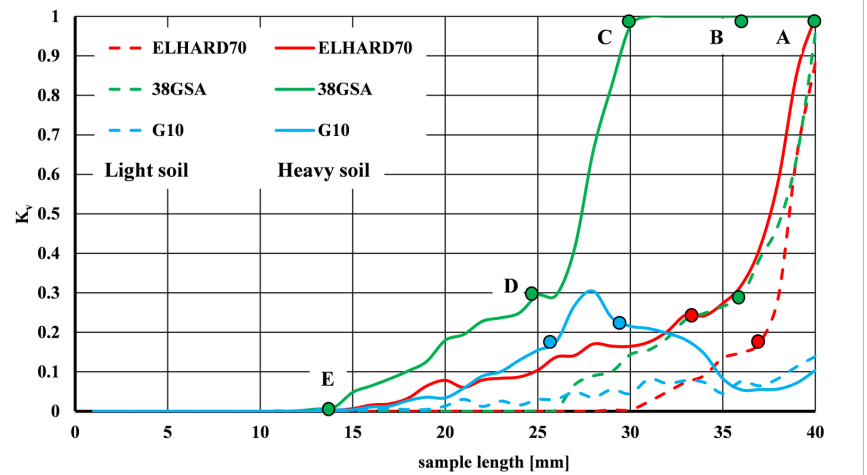


Fig. 14. Characteristics of local volumetric wear of 38GSA steel samples (heavy soil)

On the basis of the collected data, it can be observed that between the individual cross sections there were significant differences in the value of the $K_{vi}$ coefficient in the area of intense wear (sample length over 26 mm).

Characteristics $K_{vi}$ vary depending on the materials and the technology of producing the samples shown in Fig. 2. If a specimen is of a regular shape, the change in wear volume is linear. A change in the shape is defined as a deviation from the linearity of an increase/decrease in the local wear. This can be explained by the change in the direction of the nominal and friction force, and the time of contact between the abrasive particles and the surface that shapes the elementary volume $\Delta V_{iw}$. For example, samples from 38GSA steel are character-
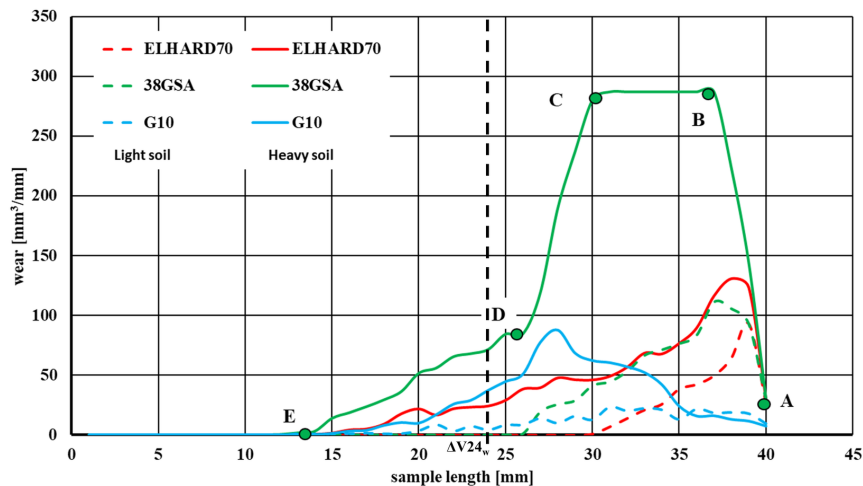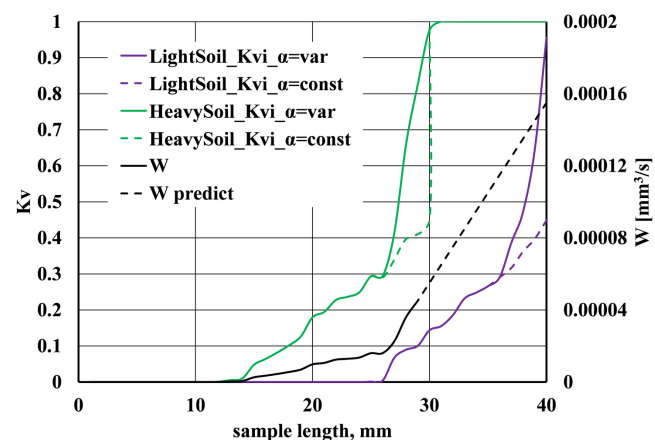
Fig. 16. Local volumes of the worn material



Fig. 17. The relationship between the volumetric wear coefficient $K_v$ and the abrasive grain sliding trajectory

It was also presented how the proposed $K_{vi}$ coefficient can be used to analyze the dependence of wear in relation to the shape of the surface. If the actual sliding trajectory of soil particles along the surface is neglected, assuming $a=const=0°$, local wear will be characterized in Fig. 17 by curves marked with broken lines. The characteristics will then be linear. In fact, $α = var$ (in Fig. 17 this corresponds to the curves drawn with solid lines). The previously noted breakthrough

in the characteristics therefore indicates a change/increase in the slip path along the wedge surface $α<90°$. For samples from 38GSA material tested in heavy soil, at the length x= 25.5 mm, there is a transition to the working surface of the wedge. This explains the sudden increase in $K_{vi}$.

Table 5 summarises the examples of volumes of each chisel, determined between the initial part of the chisel and the 84 mm cross section (the last section which included all the chisels) and between the 84 mm cross section and the final end of a particular chisel. As for the chisel which performed the ploughing over an area of 6 ha, a negligible volume value was noted virtually above the 84 mm cross section. The greatest volumes were noted for the chisels mounted on the fourth body, with these volumes being more than five times smaller for the chisel, which performed work over an area of 6 ha than those for a new chisel.

Based on the courses of the curves depicting the chisel wear, summarised in Fig. 18, it can be concluded that the wear intensity is strictly determined by the mounting holder location and the amount of work performed. What can be seen is the different wear course for chisels mounted on the first plough body.
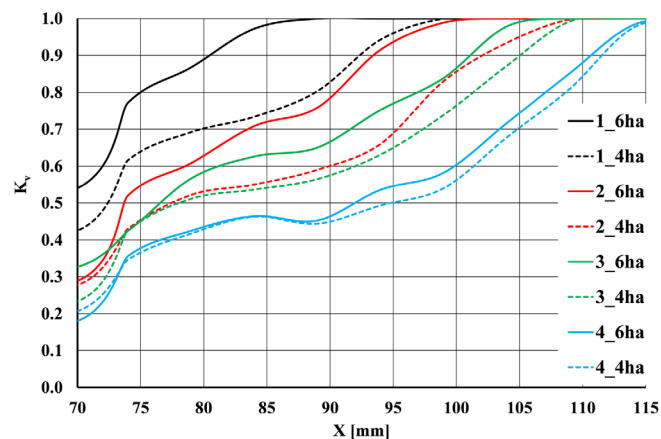


Fig. 18. Characteristics of local volumetric wear of chisels (Fig. 1).

As regards the chisels mounted on the first body which performed work over an area of 6 ha, total wear was found at a distance of 89 mm from the clamping edge. In the analysed area (above the distance

Table 5. Results of the chisel wear determination by the volumetric and mass methods

| Chisel No | Entire chisel volume | Volume below the 84 mm cross section | Volume above the 84 mm cross section |
|---|---|---|---|
| | V [mm³] | $\sum_{0}^{84} V_{iw}$ [mm³] | $\sum_{85}^{l} V_{iw}$ [mm³] |
| New | 132064 | 66189.6 | 65874.8 |
| 1 – 6 ha | 41235 | 41104.5 | 130.593 |
| 1 – 4 ha | 45427.4 | 42946.3 | 2481.13 |
| 2 – 6 ha | 50784.8 | 47785.5 | 2999.3 |
| 2 – 4 ha | 53985.8 | 46331.8 | 7653.95 |
| 3 – 6 ha | 56049.7 | 50546.5 | 5503.26 |
| 3 – 4 ha | 60663.4 | 52726.2 | 7937.23 |
| 4 – 6 ha | 66698.1 | 55188.2 | 11509.9 |
| 4 – 4 ha | 67105.5 | 54970.7 | 12134.8 |

X = 70 mm), the chisel wear exceeded 50% of the volume of a nominal operating part, and from the distance X = 75 mm, even 80% of the volume. The chisel also mounted on the first plough body, which performed work equivalent to 4 ha, exhibited less wear intensity, which is associated with the smaller amount of work performed. A characteristic of the chisels mounted on the first body was the steep wear increase (section C-B), which is due to the wear of the back part of the chisel. The chisel mounted on the second plough body (6 ha) exhibited a very similar course of wear. Despite having performed 50% more work, the wear was slightly lower. Having compared the wear of chisels mounted on the first and second plough body, it can be concluded that for the latter, their wear is approx. 20% less. There was significantly less wear for the chisels mounted on the third and fourth plough body. At a distance of 70 mm from the clamping edge, the chisel mounted on the third body was worn by over 30% after performing the work over 6 ha, and by approx. 25% after performing the work over 4 ha. The difference in wear between chisels performing different amounts of work contributes to an increase in wear by only a maximum of 5%. However, despite the different nature of the wear of individual chisels, an analysis of the wear on each chisel helps to distinguish the same characteristic areas of wear. For the purposes of analysis, wear intensity can be regarded as the slope of a curve. For each chisel, a section with the highest wear intensity value is clearly visible. This is a section located near the cutting edge, section A–B (marked in the figure for chisel 4). The intensity value is similar for almost every chisel. In the B-C section for chisel 4, a range can be distinguished in which the intensity of wear was the lowest. This was particularly true for chisels 3 and 4. On the other hand, within the C–D area, a moderate wear area with a higher intensity was distinguished. The D–E area was also characterised by a high value of wear intensity. However, in this area, the value was due to the characteristic shape of chisels in this area (excess material in new parts on the operating surface of the chisel). There is also a transition zone for the part's surfaces that are less susceptible to wear, and these are the sections running north to east. As the duration of wear processes increases, the transition zone decreases, which is evident in chisels 1 and 2.

## 5. Conclusions

The paper presents the use of modern, non-contact methods of assessing the wear of working elements processing soil abrasive mass. To test the wear, an original method was used that uses the discretization of the tested space based on the 3D scanning technology and the examination of the surface condition with a microscope.

The three-dimensional surface defect models developed in the article, obtained on the basis of 3D scanning with an accuracy of 0.02 mm, made it possible to describe the changes in the volumetric wear characteristics of the soil processing elements. New wear characteristics have been proposed that will enable the transition from generalized to local change values. The obtained data on volumetric consumption was presented in the form of the local volumetric consumption coefficient, which is the ratio of the volume of the used element to the nominal volume. The coefficient of local volumetric wear shows the influence of the nominal shape and the slip trajectory of the abrasive particle along the elementary surface on the intensity of wear. The proposed coefficient allows to extend the so far analyzes of the wear of working elements and to supplement the Holm-Archard model commonly used in such issues.

Using the volumetric wear data, it is possible to determine after which time / friction path the total wear of the workpiece occurs. This will facilitate the prediction of the actual working element life and the production of more wear-resistant soil cutting elements.

The credibility of the obtained dependencies is confirmed by the agreement as to the value of mass wear, as well as the results of the application of the proposed method to various friction conditions and materials of working elements of agricultural machines.

On the basis of the scanning results obtained, it can be concluded that the profile of the working element in the case of the tested samples changed significantly. Wear in the abrasive mass with the presented characteristics causes that the cutting edges undergo an intensive change in the thickness of the blade. It is possible to identify zones of moderate and heavy wear.

The developed method of identifying the wear process may be useful for tests related to the assessment of the wear of working elements and the assessment of wear resistance (in terms of geometry, material, conditions of use, etc.), as well as in the search for solutions to reduce wear, focusing on the most critical parts of the working element.

Depending on the type of material and type of soil, it was found different nature of wear. When the materials were being worn in the light soil, the loosely bound abrasive particles, characterised by high freedom of movement, caused scratching and ridging on the friction surface. As the content of loam and dust particles in the soil increases, a different course of wear can be noted. For the material from sintered carbide G10, the surface is slightly scratched and the wear processes are hardly noticeable.

The direction of further research is aimed at finding the relationships between the shape of the surface being worn and the local friction coefficient. Together with the proposed characteristics of local volumetric wear, it will enable further development of modeling of the wear process of soil cutting elements.

## References

1. Archard J F, Hirst W. The Wear of Metals under Unlubricated Conditions. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 1956; 236 (1206): 397–410, https://doi.org/10.1098/rspa.1956.0144.
2. Archard J F. Contact and Rubbing of Flat Surfaces. Journal of Applied Physics 1953; 24 (8): 981–988, https://doi.org/10.1063/1.1721448.
3. Bayhan Y. Reduction of wear via hardfacing of chisel ploughshare, Tribology International 2006; 39(6): 570–574, https://doi.org/10.1016/j.triboint.2005.06.005.
4. Bhakat A K, Mishra A K, Mishra N S. Characterization of wear and metallurgical properties for development of agricultural grade steel suitable in specific soil conditions. Wear 2007; 263 (1–6 ): 228–233, https://doi.org/10.1016/j.wear.2006.12.006.
5. Białobrzeska B, Kostencki P. Abrasive wear characteristics of selected low-alloy boron steels as measured in both field experiments and laboratory tests. Wear 2015; 328–329: 149–159, https://doi.org/10.1016/j.wear.2015.02.003.
6. Cucinotta F, Scappaticci L, Sfravara F, Morelli F, Mariani F, Varani M, Mattetti M. On the morphology of the abrasive wear on ploughshares by means of 3D scanning. Biosystems Engineering 2019; 179: 117–125, https://doi.org/10.1016/j.biosystemseng.2019.01.006.
7. Er U, Par B. Wear of plowshare components in SAE 950C steel surface hardened by powder boriding. Wear 2006; 261(3): 251–255, https://doi.org/10.1016/j.wear.2005.10.003.
8. Hawryluk M, Ziemba J, Zwierzchowski M, Janik M. Analysis of a forging die wear by 3D reverse scanning combined with SEM and hardness tests. Wear 2021; 476: 203749, https://doi.org/10.1016/j.wear.2021.203749.
9. Hawryluk M, Ziemba J. Application of the 3D reverse scanning method in the analysis of tool wear and forging defects. Measurement 2018; 128: 204–213, https://doi.org/10.1016/j.measurement.2018.06.037.
10. Horvat Z, Filipovic D, Kosutic S, Emert R. Reduction of mouldboard plough share wear by a combination technique of hardfacing. Tribology

International 2008; 41(8): 778–782, https://doi.org/10.1016/j.triboint.2008.01.008.

11.  Jacobson S, Hogmark S. Tribologi – Friktion, smörjning och nötning. Liber Utbildning AB, Uppsala, Sweden, 1996, ISBN: 9163415321

12.  Konat Ł, Jasiński R, Białobrzeska B, Szczepański Ł. Analysis of the static and dynamic properties of wear-resistant Hardox 600 steel in the context of its application in working elements. Materials Science–Poland 2021; 39(1), https://doi.org/10.2478/msp-2021-0007.

13.  Konat Ł, Napiórkowski J, Kołakowski K, Resistance to wear as a function of the microstructure and selected mechanical properties of microalloyed steel with boron. Tribologia 2016; 268 (4):101–114, https://doi.org/10.5604/01.3001.0010.6986.

14.  Kostencki P, T. Stawicki T. Durability and wear geometry of subsoiler shanks provided with sintered carbide plates. Tribology International 2016; 104: 19–35, https://doi.org/10.1016/j.triboint.2016.08.020.

15.  Mattetti M, Molari G, Sereni E. Damage evaluation of driving events for agricultural tractors. Computers and Electronics in Agriculture 2017; 135: 328–337, https://doi.org/10.1016/j.compag.2017.01.018.

16.  Meng H C, Ludema K C. Wear models and predictive equations: their form and content. Wear 1995; 181–183 (Part 2): 443–457, https://doi.org/10.1016/0043-1648(95)90158-2.

17.  Napiórkowski J, Lemecha M, Ł. Konat Ł. Forecasting the Wear of Operating Parts in an Abrasive Soil Mass Using the Holm-Archard Model. Materials 2019; 12: 2180, https://doi.org/10.3390/ma12132180.

18.  Napiórkowski J, Olejniczak K, Konat Ł. Wear Properties of Nitride-Bonded Silicon Carbide under the Action of an Abrasive Soil Mass. Materials 2021; 14(8): 2043, https://doi.org/10.3390/ma14082043.

19.  Owsiak Z. Wear of symmetrical wedge-shaped tillage tools. Soil & Tillage Research 1997; 43(3): 295–308, https://doi.org/10.1016/S0167-1987(97)00020-2.

20.  Quirke S, Scheffler O, Allen C. An evaluation of the wear behaviour of metallic materials subjected to soil abrasion. Soil & Tillage Research 1988; 11(1): 27–42, https://doi.org/10.1016/0167-1987(88)90029-3.

21.  Ranuša M, Gallo J, Vrbka M, Hobza M, Paloušek D, Křupka I, Hartl M. Wear Analysis of Extracted Polyethylene Acetabular Cups Using a 3D Optical Scanner. Tribology Transactions 2017; 60 (3): 437–447, https://doi.org/10.1080/10402004.2016.1176286.

22.  Stradomski Z. Mikrostruktura w zagadnieniach zużywania staliw trudnościeralnych. Wydawnictwo Politechniki Częstochowskiej, Częstochowa, Poland, 2010, ISBN: 978-83-7193-468-1.

23.  Vrublevskyi O, Napiórkowski J, Gonera J, Tarasiuk W. Numerical wear models for knock-on chisels in real operating conditions. Journal of Tribology 2022; 144 (9), https://doi.org/10.1115/1.4054020.

24.  Yazici A. Investigation of the reduction of mouldboard ploughshare wear through hot stamping and hardfacing processes. Turkish Journal of Agriculture and Forestry 2011: 461–468, https://doi.org/10.3906/tar-1105-29.

25.  Ziemba J, Hawryluk M, Rychlik M. Application of 3D Scanning as an Indirect Method to Analyze and Eliminate Errors on the Manufactured Yoke-Type Forgings Forged in SMED Device on Modernized Crank Press. Materials 2021; 14 (1): 137, https://doi.org/10.3390/ma14010137.
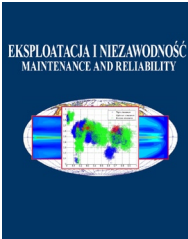
# Analysis and prediction of leak detection in the low-pressure heat treatment of metal equipment

**Indexed by:**
Web of Science Group

Sławomir Kłos[a], Justyna Patalas-Maliszewska[a], Michal Bazel[b]

[a]University of Zielona Góra, Institute of Mechanical Engineering, ul. prof. Z. Szafrana 4, 65-516 Zielona Góra, Polska
[b]Seco/Warwick S.A. , ul. Sobieskiego 8, 66-200 Świebodzin, Poland

## Highlights

- The infiltration method is not effective in predicting furnace leaks.

- The prediction of furnace leaks is possible with the help of a Lambda probe.

- Artificial neural networks are useful in predicting furnace leaks.

- The predictive furnace leak detection model can be used on-line.

## Abstract

The low-pressure heat treatment of metals enables the continuous improvement of the mechanical and plastic properties of products, such as hardness, abrasion resistance, etc. A significant problem related to the operation of vacuum furnaces for heat treatment is that they become unsealed during operation, resulting from the degradation of seals or the thermal expansion of the construction materials. Therefore, research was undertaken to develop a prediction model for detecting leaks in vacuum furnaces, the use of which will reduce the risk of degradation in the charge being processed. Unique experimental studies were carried out to detect leakages in a vacuum pit furnace, simulated using the ENV 116 reference slot. As a consequence, a prediction model for the detection of leaks in vacuum furnaces- which are used in the heat treatment of metals- was designed, using an artificial neural network. (93% for MLP 15-10-1) was developed. The model was implemented in a predictive maintenance system, in a real production company, as an element in the monitoring of the operation of vacuum furnaces.

## Keywords

vacuum heat treatment, vacuum furnace leaks, predictive maintenance, artificial neural networks.

## 1. Introduction

Metal heat treatment processes are a very important element of the manufacturing processes of products in industries such as the energy, automotive, aviation and mining industries etc. The process of hardening, carburising and/or nitriding is often the last process in the production cycle and allows the reproducible properties of the processed details, such as the surface hardness, core hardness or thickness of the diffusion layer to be achieved. The key to obtaining the assumed properties of the processed details is to maintain identical conditions for the implementation of heat treatment processes in the process chamber of the furnace, such as operating pressure, heating time, temperature, amount of technical gas etc. Even a slight change of parameters inside the vacuum furnace, during the heat treatment process, may result in irreversible damage to the charge and/or to the furnace. Both in the energy and mining industries, due to large dimensions and high production costs, the value of the charge to the furnace may exceed the value of hundreds of thousands of euros. The relevance of the problem may be proved by the scope of application of the details processed, for example in control systems, gears, heat exchangers in aeroplanes and cars, drive shafts for wind turbines, gears with large modules, elements of drill bits and cutters in mining machines, drive transmission systems, dies, surgical instruments etc.

Low-pressure carbonisation processes are usually carried out within temperatures ranging from 950°C to 1020°C at a pressure of approximately 5x100 mbar. In the tests conducted, the measured pressure values are given in millibars, where 1mbar = 100 Pa due to the use of this unit of measurement in economic practice in the field of vacuum technology. In low-pressure carburising processes, it is necessary to maintain the stability of the process, which guarantees obtaining a precise and repeatable thickness of the carburised surface layer within the specified tolerance range. Possible leaks may result in a change in the parameters of the heat treatment process as well as damage to the charge, as a result of failing to meet the technological requirements expected.

Predictive maintenance is an important component of the Industry 4.0 Concept, especially when using production resources where failures do not occur suddenly and the risk of their occurrence increases over time. An example of such a progressive failure may be a leak in a vacuum furnace. Due to changes in the temperature of the con-

struction of the furnace chamber and the wear of some elements, such as the seal in the furnace cover during operation, the furnace may become unsealed. Since modern vacuum furnaces are equipped with highly efficient. cascade pumping systems consisting of a mechanical pump and a Roots' pump, pressure build-up in the vacuum furnace due to leaks is often virtually unnoticeable. However, as a result of leaks, the composition of the atmosphere inside the furnace chamber changes, on account of a higher oxygen content, which may lead to damage to the charge or failure to achieve the assumed parameters in the carburising process and, as a result, not achieving all technological requirements.

The implementation of predictive maintenance methods into industrial practice is one of the key assumptions of the Industry 4.0 Concept [13][15]. The progressive automation and digitisation of production processes requires the implementation of systems that will enable states of emergency to be diagnosed in advance. This approach makes it possible to:

- reduce losses associated with damaged parts as a result of failures,
- reduce unplanned downtime,
- reduce the maintenance costs incurred, regarding specialised, technical personnel, such as automation engineers and mechanics.

Systems for monitoring the parameters of complex technical systems are particularly important where these are situated in remote regions of the world and where access to qualified engineering staff is limited. Currently, data on monitoring the operational parameters of devices is stored in databases of ERP or MES systems. The Industry 4.0 Concept assumes real-time monitoring and recording of machine operating parameters and stores them in the form of big data sets in the Cloud. Multi-criteria data analysis with the use of artificial intelligence methods will enable the construction of efficient algorithms that will allow failure conditions to be predicted and thus, effectively prevented due to the inspection and repair of production resources. Research related to the development of the structure of a predictive maintenance system, based on the monitoring of real-time data within the reference model of industry architecture (RAMI 4.0) and DSR (Design Science Research) to reduce costs and operations, were led by Sahba et al. [24]. Predictive maintenance models, based on mathematical programming and deep learning allow the technical condition of individual elements of the production system to be predicted [12]. Machine learning methods based on machine performance analysis and monitoring of production environment variables are used in many of the studies carried out, in order to predict the emergency conditions of production systems [6][5][17][27][29]. A very important aspect of the implementation of predictive maintenance into industrial practice, is the analysis of big data sets, based on computationally efficient algorithms [31]. Another approach involves the construction of mathematical models to calculate maintenance rates for any schedule of time up to failure [20]. An important issue regarding maintenance management, based on prediction methods, is an approach based on multi-criteria decision making, integrated with the traditional method of analysing failure mode, effects, and criticality (FMECA) [1].

Low-pressure carburising is a long-lasting and energy-consuming heat treatment process, especially considering its requirement to obtain thick carburising layers, i.e., over 2.00 mm. One of the important parameters that influence the stability and repeatability of the process is the low oxygen content in the vacuum furnace chamber. The most common cause of the deterioration of this parameter is a leak in the furnace, which results in air entering the furnace chamber. The issue of leakage, in devices operating in high vacuum conditions is an important area of research and the subject of scientific publications [4] [9]. There are many methods for testing the leakage of vacuum devices, such as the pressure increase or decrease test, the leak test with gas-sensitive vacuum gauges, the test by immersion in various formulations or spraying with foam, the Krypton 85 test, the high fre-

quency vacuum test, test with chemical reactions and dye penetration [21]. An important method for testing the tightness of technical objects is the tracer gas method. The most popular tracing gases are helium, ammonia, hydrogen, nitrogen and semi-precious gases. A commonly used method of leak detection is the helium method, so-called [13] [28]. Helium is very well suited to the leak detection of vacuum devices because it is a non-toxic, non-flammable gas and does not form explosive mixtures with other gases. It is also neutral to the environment and does not enter into chemical reactions with other substances. The use of helium is not limited by the range of temperatures or pressures as it is a very temperature stable gas, which facilitates the testing of objects in environments with extremely high or low temperatures. Helium is the gas- right after hydrogen- with the smallest unit particle which enables penetration of micro-fractures in materials, therefore helium tests are very accurate. The relatively low price of helium compared to other noble gases should also be noted. The method itself is relatively simple and consists in connecting a so-called helium detector to the object being studied, while the operator dispenses a small portion of helium to any possible leak. The test object must reach a high, negative pressure. The method requires a detector with high accuracy, viz., the detection of leaks at the level of 5x10-12 mbar l/s. When using a helium detector, it is also possible to approximate the location of the leak. Another method for detecting leaks in vacuum furnaces is the so-called infiltration method, which consists in heating the furnace to a certain temperature at a specific negative pressure - usually as low a pressure as possible and then, as it slowly cools, calculate the value of the infiltration, i.e., the decrease in the value of the vacuum level, within a specific timeframe. The infiltration testing process allows leaks to be detected, however, it is long-lasting (several hours to several dozen hours) and does not allow the leak to be located. In general, the leak test should be preceded by the identification of possible leaks using the helium method.

Based on an analysis of the results of the literature on the subject, the infiltration method was selected for detecting leaks in vacuum furnaces with experimental tests being carried out to analyse leakages in pit furnaces for specific settings of the ENV 116 standard slot.

The use of various tools such as the Markov Process, Bays Networks, artificial neural networks or simulation methods, based on the Monte Carlo method for predictive maintenance purposes, has been the subject of publications by many authors [7][25]. Predictive maintenance methods use the Markov Process, Bays Networks, artificial neural networks and simulation methods based, for example, on the Monte Carlo method.

Many publications cover the use of machine learning methods for predictive maintenance, based on an analysis of machine performance and the variables of the production environment [5][17][27]. Part of the research concerns the acquisition and storage of data in the Cloud and the construction of a system platform for predictive maintenance [26]. Artificial neural networks are used, among other things, in order to analyse and classify data, both current and forecast, with the help of in-depth learning techniques and will enable images to be both recognised and reproduced. In the maintenance department, models are expected for detecting and forecasting future failures, in real time. [3] Based on an analysis of the literature, the most popular data-driven decision methods include the Support Vector Machine (SVM), Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Random Forest, K-Nearest Neighbours and the Hidden Markov Model [22]. Currently, many methods, techniques and procedures, using intelligent production systems for maintenance workers, are based on deep learning techniques [17][32], however, there are still good examples of the use of artificial neural networks in maintenance, in the monitoring of tool wear, in the diagnosis of vibration in machining systems, in the thermal analysis of machines, in the analysis of other malfunctions affecting production, as well as in the diagnostics of finished products [9].

In the literature on the subject analysed, no model was found that could be used to predict the detection of leaks in devices used in the low-pressure heat treatment of metals. Therefore, research work was undertaken to design an effective leak detection model in vacuum furnaces, based on data obtained from the study of a working pit furnace.

The article presents the stages of a research experiment concerning the analysis of leaks with the use of a standard slot for various pressure and temperature parameters. Based on an analysis of the results of the experimental studies, a model for detecting leaks predictively, using artificial neural networks, was designed.

## 2. Research methods

### 2.1. Experimental work- testing furnace tightness- based on an analysis of the infiltration

The infiltration test is an effective, but relatively time-consuming method for obtaining knowledge about leakage in vacuum furnaces. The procedure for testing the tightness of the pit furnace, using the infiltration method, is shown in Figure 1.
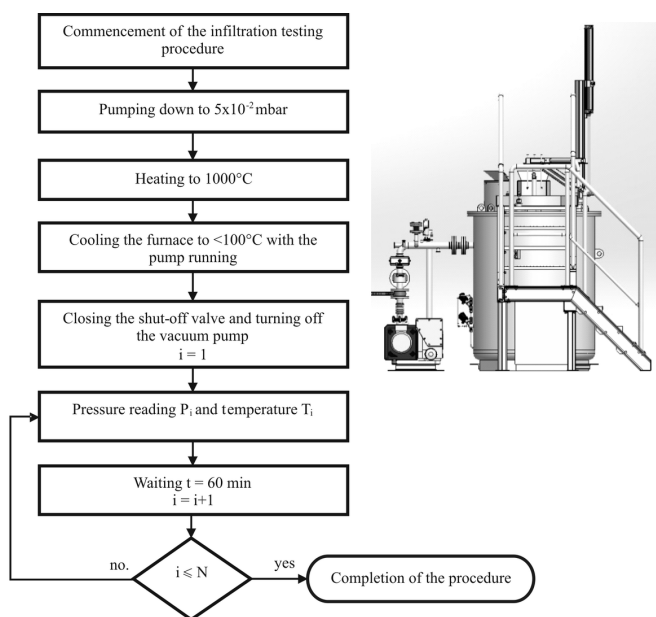


*Fig. 1. The procedure for testing a leaking pit furnace, using the infiltration method*

The longest stage in the infiltration test procedure here presented, is the cooling down of the furnace to a temperature of 50°C. Depending on the design of the furnace, that is, regarding the thickness of the insulation *and the size of the heating chamber etc.,* the furnace cooling down process may take from several hours to several dozen hours. The control calculations for the infiltration test are performed, based on formula (1) shown below:

$$N = 1000 \cdot V \cdot \frac{P_e \cdot \left( 0,5 \cdot \left( 1 + \frac{273 + T_b}{273 + T_e} \right) - P_b \right)}{60 \cdot t} \quad (1)$$

where

- $N$ - (leak rate) [mbar* l/s]
- $V$ – volume of the furnace heating chamber [l]
- $P_b$ – initial pressure [mbar],
- $P_e$ – final pressure [mbar],
- $T_b$ – initial temperature [°C],
- $T_e$ – final temperature [°C],
- $t$ – time [min].

The criterion for the tightness of the device, *vis-à-vis* the infiltration test, was set at 5,0 10-3 [mbar·l/s].

Table 1 shows examples of the results of the pit furnace infiltration test wherein a leak was detected.

The results presented in Table 1 differ many times from the criterion adopted, therefore the device tested shows increased leakage. After a review of the structure of the device and replacement of the valves, the infiltration test was carried out again. The test was performed with a slightly higher final temperature. The results are presented in Table 2.

As can be seen from the data presented in Table 2, the minimum indicator, adopted and established for the infiltration test, was exceeded over 100 times, thus indicating a leaking device. After re-analysis of tightness with the use of a helium detector and the sealing of the structural elements having been replaced, the infiltration test was carried out again; this gave satisfactory results. Table 3 presents the results of an infiltration test for a furnace which meets the criteria for leakages.

According to the analysis of the infiltration tests carried out, the method is not only effective, but is also long-lasting and may require several repetitions in the case of unsatisfactory results. In production conditions, it is also quite expensive, because it requires shutting down the heat treatment device for several hours. These devices are heavily loaded with orders in most enterprises and constitute bottlenecks in the manufacturing process. It is very important that, before starting an infiltration test, a helium test is carried out, in order to eliminate possible leakage in the device. Due to the specific features of the infiltration test, i.e., its long duration, furnace shutdown and high energy costs, this method cannot be used for a maintenance predictive system. Therefore, research work was undertaken to design a pit furnace leakage prediction model, using artificial neural networks.

### 2.2. Data collection and analysis

The Lambda probe is a sensor commonly used in the automotive industry to analyse the oxygen content in exhaust gas. The voltage generated by the sensor is lower than the higher oxygen content in exhaust gas, while the small amount of oxygen ions in the exhaust gas generates high voltage. To test for leakages in a pit furnace, the standard EVN116 slot (gas dosing valve) was used, which enables

*Table 1. Examples of the results of the first infiltration test for a leaky vacuum furnace.*

| V | t | $P_b$ | $P_e$ | $T_b$ | $T_e$ | N |
|---|---|---|---|---|---|---|
| [l] | [s] | [mbar] | [mbar] | [°C] | [°C] | 10-3 [mbar·l/s] |
| 1308 | 3600 | 1.40 | 4.06 | 42 | 42 | 967.83 |
| 1308 | 7200 | 1.40 | 6.87 | 42 | 41 | 995.00 |
| 1308 | 10800 | 1.40 | 9.79 | 42 | 41 | 1018.07 |
| 1308 | 14400 | 1.40 | 12.79 | 42 | 40 | 1037.80 |
| 1308 | 18000 | 1.40 | 25.12 | 42 | 39 | 1082.99 |
| 1308 | 21600 | 1.40 | 49.51 | 42 | 36 | 1102,87 |
| 1308 | 25200 | 1.40 | 74.52 | 42 | 34 | 1121.14 |

Table 2.  Examples of the results of the second infiltration test for a leaky vacuum furnace

| V | t | $P_b$ | $P_e$ | $T_b$ | $T_e$ | N |
|---|---|---|---|---|---|---|
| [l] | [s] | [mbar] | [mbar] | [ºC] | [ºC] | $10^{-3}$ [mbar·l/s] |
| 1308 | 3600 | 1.46 | 2.85 | 76 | 75 | 504.47 |
| 1308 | 7200 | 1.46 | 4.35 | 75 | 74 | 525.88 |
| 1308 | 10800 | 1.46 | 5.90 | 74 | 72 | 538.53 |
| 1308 | 14400 | 1.46 | 7.42 | 72 | 71 | 542.02 |
| 1308 | 18000 | 1.46 | 8.91 | 71 | 70 | 542.51 |
| 1308 | 21600 | 1.46 | 10.36 | 70 | 69 | 539.76 |
| 1308 | 25200 | 1.46 | 11.80 | 69 | 68 | 537.59 |
| 1308 | 28800 | 1.46 | 13.17 | 68 | 67 | 532.93 |
| 1308 | 57600 | 1.46 | 23.01 | 67 | 58 | 495.93 |
| 1308 | 86400 | 1.46 | 31.70 | 58 | 52 | 462.59 |

Table 3.  The results of the infiltration test for a furnace with a tight vacuum

| V | t | $P_b$ | $P_e$ | $T_b$ | $T_e$ | N |
|---|---|---|---|---|---|---|
| [l] | [s] | [mbar] | [mbar] | [ºC] | [ºC] | $10^{-3}$ [mbar·l/s] |
| 1308 | 3600 | 0.238696 | 0.244000 | 43 | 43 | 2.00 |
| 1308 | 7200 | 0.238696 | 0.249304 | 43 | 42 | 1.98 |
| 1308 | 10800 | 0.238696 | 0.255091 | 43 | 42 | 2.04 |
| 1308 | 14400 | 0.238696 | 0.259913 | 43 | 42 | 1.98 |
| 1308 | 18000 | 0.238696 | 0.265700 | 43 | 41 | 2.02 |
| 1308 | 21600 | 0.238696 | 0.271004 | 43 | 41 | 2.01 |
| 1308 | 25200 | 0.238696 | 0.276308 | 43 | 41 | 2.01 |
| 1308 | 28800 | 0.238696 | 0.282095 | 43 | 40 | 2.03 |
| 1308 | 57600 | 0.238696 | 0.287399 | 43 | 40 | 2.03 |
| 1308 | 86400 | 0.238696 | 0.292704 | 43 | 40 | 2.02 |

the leak rate to be determined manually. Figure 2 shows the test site (Figure 2).
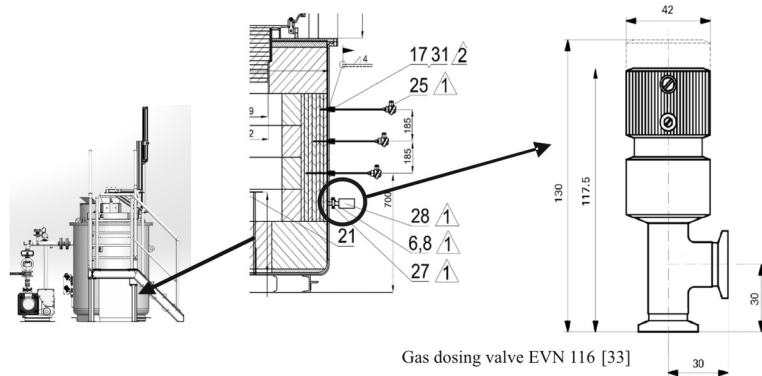


Gas dosing valve EVN 116 [33]

Fig. 2. Test stand for testing vacuum furnace leakage

As can be seen from the characteristics presented, setting the slot to 100 results in a horizontal leak at a level of $0.5 \cdot 10^{-3}$ mbar·l/s, but when the slot is set to 200, the leak is at a level of $0.8 \cdot 10^{-3}$ mbar * l/s. The test programme included an analysis of the characteristics of changes in the Lambda probe indications, depending on the size of the reference gap for different temperatures in the heating of the pit furnace. The following activities were planned for the purpose of conducting research experiments:

• setting a specific leakage for the reference slot, without opening the slot,
   • setting the maximum heating temperature $T_{max}$,
   • creating a vacuum in the furnace at the $P_{min}$ level,
   • heating the furnace to temperature $T_{max}$ with the pump running,
   • keeping $T_{max}$ for 15 min with the pump running,
   • closing the pump shut-off valve and turning the pump off,
   • opening reference slot $t_1$,
   • closing reference slot $t_2$,
   • turning the furnace off.

   The tests were carried out at different heating temperatures: 500°C (loading the charge), 800°C (near temperature for hardening) and 1000°C (vacuum carburisation temperature). Figure 3 shows the temperature diagrams and the pressure and Lambda probe readings during heating to a temperature of 500°C, with the vacuum pump working, in order to determine the reference characteristics of the Lambda probe for a tight furnace.

   The furnace was started at 7:20 a.m. at a temperature of 28°C; the target temperature of 500°C was reached after 60 minutes. The heating ramp was approximately 8°C/min. The vacuum in the furnace chamber was $7 \cdot 10^{-3}$ mbar. As can be seen from the diagram of the Lambda probe readings, taken during pumping out the furnace, the amount of oxygen ions in the furnace atmosphere decreased, which resulted in an increase in the voltage to approximately 880 mV.

   Experiments at the test site (Figure 2) were then carried out. They simulated furnace leaks, with the standard slots set to 100, 150, 200 and 250, respectively, according to the characteristics of the EVN 116
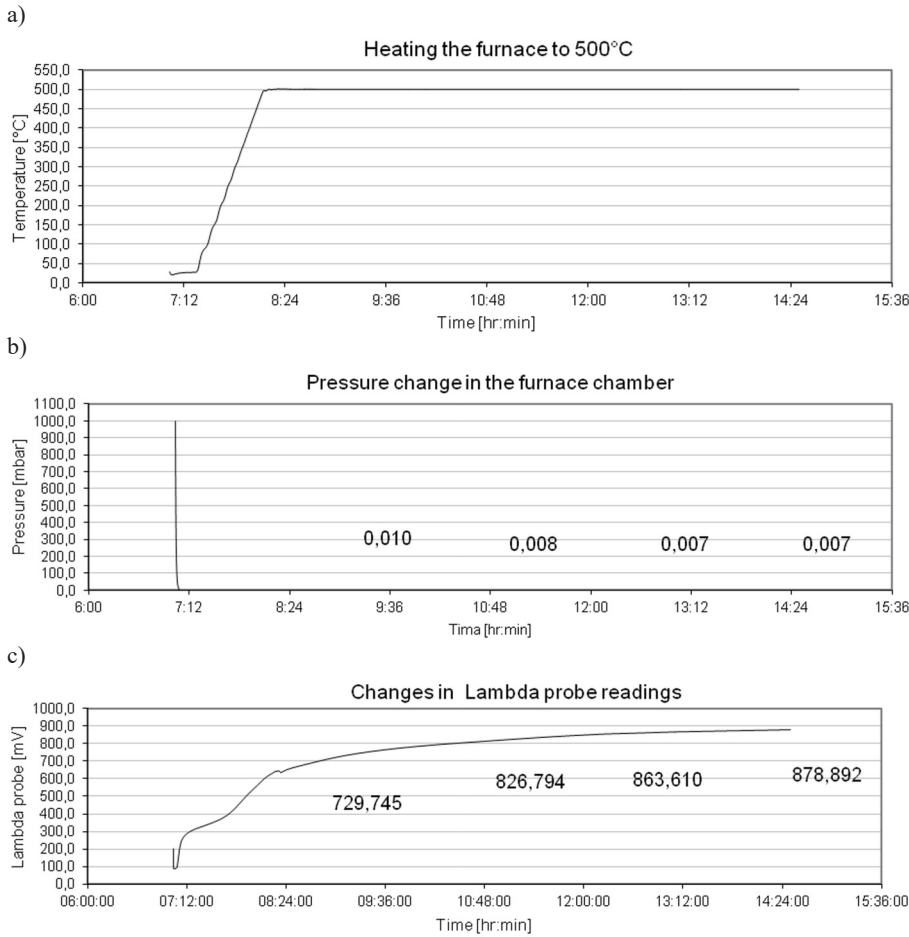
a)



b)



c)



*Fig. 3. Changes in a) temperature in the furnace chamber, b) pressure, c) voltage of the Lambda probe for a tight furnace*

standard slot in which the numerical codes of the manual setting of the standard slot correspond to the specified leak (the so-called "digital display", Figure 4), in the leakage range from approximately $10^{-3}$ do $10^{-2}$ mbar·l/s.
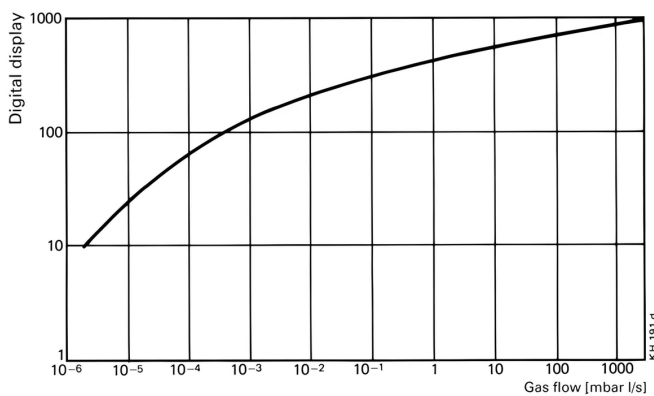


*Figure 4. Characteristics of EVN 116 standard slot [33]*

Changes in the furnace pressure for the individual settings of the reference slot and simulated voltage, as well as for the Lambda sensor readings are shown in Figure 5. As can be seen from the diagram presented, differences in the Lambda sensor readings for different sizes of the device leak are already visible 30 minutes after the opening of the standard slot. Table 4 shows the values of pressure and indications of the Lambda probe, which were recorded at equal intervals every 30 minutes, after opening the reference slot. Based on the data presented in Table 4, it is possible to diagnose any leakage in the device after 30 minutes and determine its size.

Analysis of the data obtained as a result of the conducted experimental tests (Table 4) shows that for increasingly larger leaks in the ENV 116 slot, the Lambda probe generates a correspondingly lower voltage.

The data received in the form of 1600 measurements concerning parameter values: y - Lambda probe voltage, x1 - temperature in the Furnace chamber, x2 - working pressure in the furnace chamber (values in bar), x3 - housing temperature, x4 – water flow in the cooling system, x5 - insulation temperature 1, x6 - insulation temperature 2, x7 – water flow in the cooling system of current passages, x8 - insulation temperature 3, x9 - water temperature at the inlet of the cooling system, x10 - gas pressure of the pneumatic system, X11 - temperature of the current bushing, X12 - heater current 1, x13 - heater current 2, X14 - heater current 3, x15 - partial pressure in the furnace chamber (values in millibars) constituted the basis for building a leak detection prediction model in vacuum furnaces using artificial neural networks.

## 3. Test results

The basic element of artificial neural networks are neurons, each of which is an autonomous processing unit. Each neuron carries out its own simple calculations and the structure, consisting of a large number of neurons, facilitates the multiplication potential of these calculations. The task of the neurons is to operate on the input data and present the results computed by the function activation. The neuron also describes a bias, which is an element that models

*Table 4. Lambda probe indications [mV] for various ENV 116 slot leaks*

|  | digital display | | | |
|---|---|---|---|---|
| time | 100 | 150 | 200 | 250 |
| 00:00 | 835 | 826 | 818 | 813 |
| 00:30 | 840 | 831 | 822 | 809 |
| 01:00 | 844 | 835 | 824 | 806 |
| 01:30 | 849 | 839 | 826 | 802 |
| 02:00 | 852 | 842 | 828 | 799 |
| 02:30 | 855 | 842 | 828 | 795 |

the threshold above which the neuron sends an impulse, that is, an adjustment of the activation threshold value. The artificial neural network thus designed has parameters (weights) that must be assigned initial values. The role of the activation function is to determine the degree of excitation of the neuron, on the basis of the values reaching it. Based on the function used, the output value of the neurons is calculated.

The model was built using an artificial neural network due to its utility in both reactive and preventive maintenance as well as predictive maintenance. The model was designed on the basis of real data obtained from the operation of a vacuum furnace. An artificial neural network was chosen because using it facilitates, among other things, data classification and identification, the forecasting of wear in machine elements [9], the forecasting of the abrasive wear of cutting tool blades and the monitoring of machines and devices in action. The ANN type of unidirectional MLP multilayer neural network was selected (Multi-layered Perceptron) [23]. The network consists of neu-
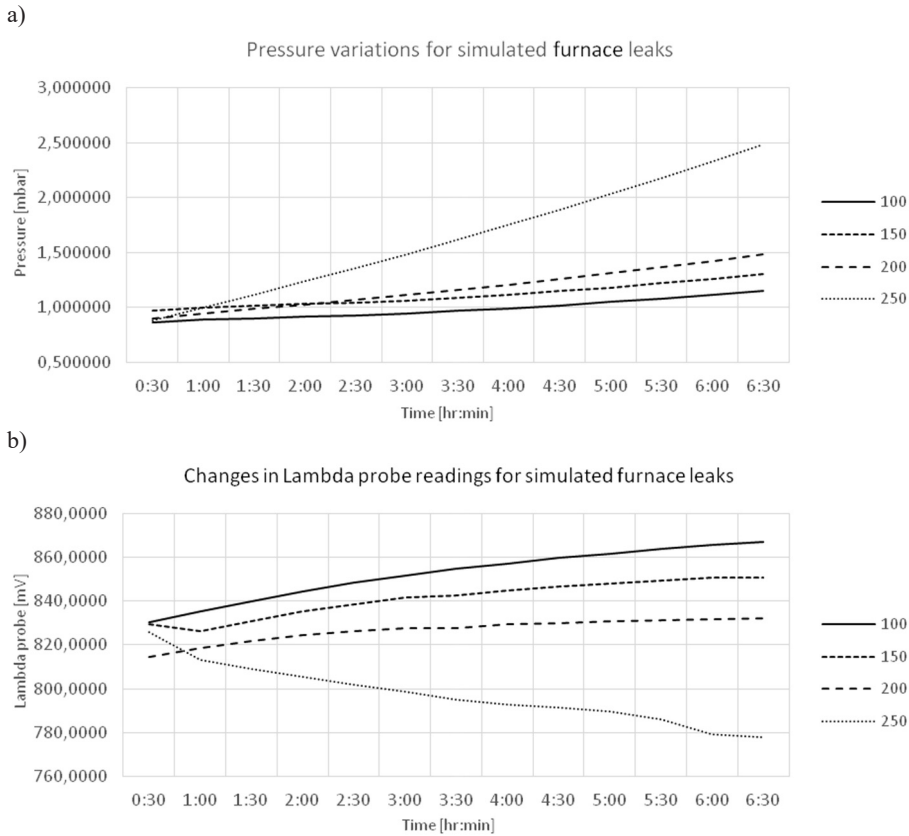
a)

Pressure variations for simulated **furnace** leaks



b)

Changes in Lambda probe readings for simulated furnace leaks



*Fig. 5. Changes in the readings of the pressure sensor and the Lambda sensor for different levels*

rons arranged in layers. Each of the neurons computes the weighted sum of its inputs; the excitation level thus determined becomes an argument for the transition function (activation function) that computes the output value of the neuron. For a unidirectional multilayer network, determining the appropriate number of hidden layers and the number of neurons in individual layers is not a simple issue. [2].

In order to build the model, an artificial neural network with a logistic activation function, a sigmoid unipolar function (formula 2) was used, due to the form of data on the Lambda probe indication for a tight furnace (Figure 3c).

$$(x) = \frac{1}{1 + e^{-x}} \qquad (2)$$

where: $x$ – is the input value of the activation function, $e$ - Euler's number. The activation function is especially useful in artificial neural networks with back propagation. The function maps the interval $(-\infty, \infty)$ to $(0,1)$ [19].

The weighting factors were determined in the process of training neurons through supervised learning. To generate a neural network Statistica, ver. 13.3 was used. The neural network model used in the study is shown in Figure 6.
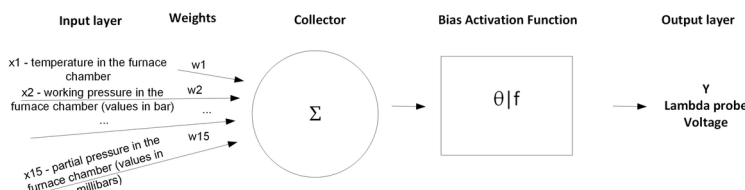


*Fig. 6. MLP network structure for leak detection prediction in vacuum furnaces.*

where:

$w_1 ... w_{15}$ - weights
Y -Lambda probe measurement value

Table 5 compares the MLP network, with the activation logistics function, in terms of the network quality achieved, for training, testing and validation, respectively and the error function. The best model was the MLP network with the structure 15–10–1, where 15-10-1 refers to the number of inputs (15), the number of neurons in the hidden layer (10) and the number of output networks (1).

The leak detection prediction model in vacuum furnaces was then verified with the use of ANN MLP 15-10-1 for the data presented in Appendix no. 1. Tests were carried out while the vacuum furnace was working and the actual results were compared with the forecast obtained (Table 6).

In Figure 7 a prediction model for the detection of leaks, in devices for the heat treatment of metals, has been presented.

## 4. Discussion

The aim of the study was to formulate a prediction model for detecting leaks in vacuum furnaces using artificial neural networks. The research required experiments to be conducted on a vacuum furnace in actual operation, followed by work related to the acquisition and analysis of data and the use of an appropriate ANN structure. The research was conducted in the R&D Department of SECO/WARWICK S.A. using a pit vacuum furnace. Table 7 presents the main characteristics indicating the originality of the research results obtained and the contribution made to the research on leakage in vacuum furnaces. In particular, the approaches to the detection of furnace leakages were described, taking into account: (1) the methods used, (2) effectiveness, (3) verification in business practice. According to the present authors' knowledge, there is no approach, in existing studies, integrating the results of experimental work on the detection of leaks in a pit furnace using the ENV 116 reference slot with the use of artificial neural networks.

Technical devices that work in high vacuum conditions and are subjected to large temperature changes, become unsealed after a certain period of time. This can be due to various reasons, such as seal wear, damaged valves, leaks due to expansion and contraction of the metal parts of the furnace etc. Leaks can occur suddenly, as a result of a fault, or gradually, as a result of the normal operation of equipment. It is possible to prevent gradual leakage of the heat treatment device by continuously monitoring selected parameters. As currently used, pumping systems for obtaining a high vacuum in furnace chambers are very efficient, therefore the occurrence of even a large leak may not be registered by the pressure sensors. Due to the economics of heat treatment processes and for the avoidance of damage to the charge, it is important that the prediction of potential leakage occurs before actual treatment is required, such as carburising or quenching, so that the charge is not degraded. This can be achieved by using artificial intelligence methods. The study demonstrated the usefulness of an artificial neural network and its effectiveness in supporting the prognosis of furnace leakage (93% for MLP 15-10-1). The model designed for the detection of leaks in vacuum furnaces (Figure 6) can be integrated in systems supporting the operation of the maintenance department. Such adaptation can even be done online, if all the features / input values taken into account can be changed and controlled online.

Table 5. Artificial neural networks for leak detection in a vacuum furnace

| Network name | Quality (learning) | Quality (testing) | Quality (validation) | Error (learning) | Error (testing) | Error (validation) | Learning algorithm | Error function |
|---|---|---|---|---|---|---|---|---|
| MLP 15-7-1 | 0.980799 | 0.751166 | 0.828486 | 177.310 | 158.8133 | 86.2470 | BFGS 40 | SOS |
| MLP 15-10-1 | 0.983770 | 0.751710 | 0.933961 | 184.058 | 245.6714 | 174.4568 | BFGS 42 | SOS |
| MLP 15-12-1 | 0.563121 | 0.462774 | 0.514705 | 3191.967 | 399.8066 | 433.7868 | BFGS 4 | SOS |
| MLP 15-14-1 | 0.966649 | 0.571415 | 0.575219 | 308.369 | 256.8963 | 249.1537 | BFGS 48 | SOS |
| MLP 15-13-1 | 0.973078 | 0.568790 | 0.681975 | 249.102 | 304.1500 | 156.2498 | BFGS 41 | SOS |

Table 6. Comparison of actual values with the values forecast for leakages in a vacuum furnace

| Time | Lambda probe voltage - actual value [mV] | Lambda probe voltage – forecast value [mV] | Error |
|---|---|---|---|
| 15:45:00 | 727,2063 | 793,5139 | -66,3076 |
| 16:00:00 | 743,9453 | 795,6114 | -51,6661 |
| 16:15:00 | 755,042 | 793,7966 | -38,7546 |
| 16:30:00 | 763,1294 | 792,5906 | -29,4612 |
| 16:45:00 | 769,8062 | 793,7554 | -23,9492 |
| 17:00:00 | 775,2605 | 792,3767 | -17,1162 |
| 17:15:00 | 779,8684 | 794,0572 | -14,1888 |
| 17:30:00 | 783,9121 | 794,9163 | -11,0042 |



Fig. 7. Prediction model for detecting leaks in vacuum furnaces

The article proposes a model for predictively detecting leaks in vacuum furnaces with the use of artificial neural networks. As a result of the experimental tests conducted, consisting in simulating the leakage, the Lambda sensor readings were determined for various settings of the standard slot size. The prediction model formulated for the detection of leaks in vacuum furnaces, as used in the heat treatment of metals using an artificial neural network (93% for MLP 15-10-1) will allow leakage in the heat treatment equipment of metals- *already in the heating phase*- to be detected before the start of the carbonisation or hardening processes, thus protecting the load against damage.

As part of further research and development work on the predictive maintenance of metal heat treatment devices, work will be carried out on detecting the conditions preceding burnout of the furnace heating elements, resulting from an increase in the level of carbon deposits on the heating elements and current passages and from low-pressure carburising processes.

## 5. Summary and conclusions

The low-pressure heat treatment of metals is most often carried out in the last phase of a production cycle. Inadequate heat treatment parameters, such as too much oxygen in the furnace chamber or too high a pressure, caused by leaks in the furnace, may lead to production shortages or the production of products with lower strength and reliability. Leakage in heat treatment equipment cannot be completely eliminated as it is process-specific, *vis-à-vis* the thermal expansion of metals, damage to the seal resulting from the operation of the furnace etc.

## Authors' contributions

SK: developing the concept of the article, planning research experiments, conducting research experiments, developing a predictive model for a vacuum furnace leak detection system, writing the content of the article.
JPM: developing the concept of the article, conducting research experiments, developing a predictive model of a vacuum furnace leak detection system, writing the content of the article.
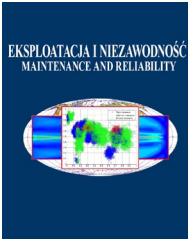MB: carrying out research experiments.

# References

1. Ahmed U, Carpitella S, Certa A. An integrated methodological approach for optimising complex systems subjected to predictive maintenance. Reliability Engineering and System Safety 2021; 216: 108022, https://doi.org/10.1016/j.ress.2021.108022

2. Bishop C. Training with noise is equivalent to Tikhomov regularisation, Neural Computation 1995, 7 (1): 108-116. https://doi.org/10.1162/neco.1995.7.1.108

3. Bousdekis A, Lepenioti K, Apostolou D, Mentzas G. Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0. IFAC-Papers On Line 2019, 52 (13): 607-612. https://doi.org/10.1016/j.ifacol.2019.11.226

4. Calcatelli A, Bergoglio M, Mari D. Leak detection, calibrations and reference flows: Practical example. Vacuum 2007; 81(11–12): 1538–1544, https://doi.org/10.1016/j.vacuum.2007.04.019.

5. Cline B, Niculescu RS, Huffman D, Deckel B. Predictive maintenance applications for machine learning. Proceedings - Annual Reliability and Maintainability Symposium 2017. 1-7, https://doi:10.1109/RAM.2017.7889679.

6. Dalzochio J, Kunst R, Pignaton E, Binotto A, Sanyal S, Favilla J, Barbosa J. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. Computers in Industry 2020; 123: 103298. https://doi.org/10.1016/j.compind.2020.103298.

7. Efthymiou K, Papakostas N, Mourtzis D, Chryssolouris G. On a predictive maintenance platform for production systems. Procedia CIRP 2012; 3: 221-226, https://doi:10.1016/j.procir.2012.07.039.

8. Fradette R J, Jones W R. Vacuum Furnace Leaks and Detection Techniques; https://www.industrialheating.com/articles/95173-vacuum-furnace-leaks-and-detection-techniques, 2019.

9. Gawlik J, Kiełbus A. Zastosowania metod sztucznej inteligencji w nadzorowaniu urządzeń technologicznych i jakości wyrobów. Praktyka zarządzania jakością w XXI wieku, 2012.

10. Gu B, Huang X. Investigation of leak detection method by means of measuring the pressure increment in vacuum. Vacuum 2006; 80(9): 996–1002, https://doi.org/10.1016/j.vacuum.2006.01.005.

11. Haripriya M, Saravanan S, Rejul M. Iot Enabling of Vacuum Heat Treatment Chambers for Data Acquisition and Analytics. 3rd International Conference on Computing Methodologies and Communication (ICCMC) 2019; 18958316, https://doi.org/10.1109/ICCMC.2019.8819829

12. Hesabi H, Nourelfath M, Hajji A. A deep learning predictive model for selective maintenance optimisation. Reliability Engineering & System Safety 2021; 219: 108191, https://doi.org/10.1016/j.ress.2021.108191.

13. Li Z, Wang K, He Y. Industry 4.0 - Potentials for Predictive Maintenance. International Workshop of Advanced Manufacturing and Automation (IWAMA) 2016, https://doi.org/10.2991/iwama-16.2016.8

14. Meng D, Sun L, Yan R, Shao R, Yu X, Li X, Zhang H, Zhao Y. Effects of cryopump on vacuum helium leak detection system. Vacuum 2017; 143: 316–319. https://doi.org/10.1016/j.vacuum.2017.06.036.

15. Mobley R K. An introduction to predictive maintenance. 2nd edition. Butterworth-Heinemann 2002. https://doi.org/10.1016/B978-075067531-4/50006-3

16. Oakes J, Lutz J. Furnace Atmosphere Controls in Heat Treating. Steel Heat Treating Technologies. ASM International 2014; 4B: https://doi.org/10.31399/asm.hb.v04b.a0005928

17. Paolanti M, Romeo L, Felicetti A, Mancini A, Frontoni E, Loncarski J. Machine Learning approach for Predictive Maintenance in Industry 4.0. 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, MESA 2018; 1-6, https://doi:10.1109/MESA.2018.8449150.

18. Patalas-Maliszewska J, Halikowski D. A Model for Generating Workplace Procedures Using a CNN-SVM Architecture. Symmetry 2019; 11: 1-14. https://doi.org/10.3390/sym11091151

19. Ponti M A, Ribeiro L S F, Nazare T S, Bui T, Collomosse J. Everything you wanted to know about deep learning for computer vision but were afraid to ask. 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T) 2017; 17-41, https://doi:10.1109/SIBGRAPI-T.2017.12.

20. Raza A, Ulansky V. Modelling of Predictive Maintenance for a Periodically Inspected System. Procedia CIRP 2016; 59 (TESConf 2016): 95–101, https://doi.org/10.1016/j.procir.2016.09.032.

21. Rottländer H, Umrath W, Voss G. Fundamentals of leak detection. Leybold GMBH (ed) Cat 2016:https://www.leyboldproducts.fr/media/pdf/90/c7/87/Fundamentals_of_Leak_Detection_EN.pdf

22. Ronao C A, Cho S B. Human activity recognition using smartphone sensors with two-stage continuous hidden markov models. Natural Computation (ICNC), 10th International Conference on. IEEE 2014; 681-686, https://doi.org/10.1109/ICNC.2014.6975918.

23. Rumelhart D E, Hinton G E, Williams R J. Learning Internal Representations by Error Propagation in Parallel Distributed Processing. Explorations in the Microstructure of Cognition, Foundations: MIT Press, 1986; Vol. 1, Cambridge MA. https://doi.org/10.7551/mitpress/5236.001.0001

24. Sahba R, Radfar R, Rajabzadeh Ghatari A, Pour Ebrahimi A. Development of Industry 4.0 predictive maintenance architecture for broadcasting chain. Advanced Engineering Informatics 2021; 49: 101324, https://doi.org/10.1016/j.aei.2021.101324.

25. Sakib N, Wuest T. Challenges and opportunities of condition-based predictive maintenance: a review. Procedia CIRP 2018; 78: 267–272, https://doi.org/10.1016/j.procir.2018.08.318

26. Schmidt B, Wang L. Cloud-enhanced predictive maintenance. Int J Adv Manuf Technol. 2018; 99: 5-13, https://doi:10.1007/s00170-016-8983-8

27. Susto G A, Schirru A, Pampuri S, McLoone S, Beghi A. Machine learning for predictive maintenance: A multiple classifier approach. IEEE Trans Ind Informatics 2015; 11(3): 812-820 https://doi:10.1109/TII.2014.2349359.

28. Takeda H. Helium leak detection method using ambient temperature of canister top. Nuclear Engineering and Design 2019; 352: 110135. https://doi.org/10.1016/j.nucengdes.2019.05.031

29. Theissler A, Pérez-Velázquez J, Kettelgerdes M, Elger G. Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. Reliability Engineering and System Safety 2021; 215: 107864, https://doi.org/10.1016/j.ress.2021.107864.

30. Vlasov A I, Echeistov V V, Krivoshein A I, Shakhnov V A, Filin S S, Migalin V S. An information system of predictive maintenance analytical support of industrial equipment. Journal of Applied Engineering Science 2018; 16(4): 515–522. https://doi.org/10.5937/jaes16-18405

31. Wen Y, Fashiar Rahman M, Xu H, Tseng T L B. Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective. Measurement: Journal of the International Measurement Confederation 2022; 187: 110276, https://doi.org/10.1016/j.

measurement.2021.110276.

32. Wuest T, Weimer D, Irgens C, Klaus D T. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research 2016; 4 (1): 23-45, https://doi.org/10.1080/21693277.2016.1192517.

33. Valve gas dosing, EVN 116. [http://www.pfeiffer-vacuum.com/productPdfs/PFI32031.en.pdf. EVN 116, Gas dosing valve with separate shut-off valve, manual].

# Physical and reliability aspects of high-pressure ammonia water pipeline failures

Indexed by:
Web of Science Group

Maciej Michnej [a], Stanisław Młynarski [a], Robert Pilch [b], Wojciech Sikora [b], Maksymilian Smolnik [b], Przemysław Drożyner [c]

[a] Cracow University of Technology, Faculty of Mechanical Engineering, al. Jana Pawła II 37, 31-864 Kraków, Poland
[b] AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics, al. Mickiewicza 30, 30-059 Kraków, Poland
[c] University of Warmia and Mazury in Olsztyn, The Faculty of Technical Sciences, ul. Oczapowskiego 11, 10-719 Olsztyn, Poland

## Highlights

- The research problem was to determine the cause of the damage of the pipeline.

- Several tests of different types of selected parts of the pipeline were carried out.

- Structural analysis of the pipeline compensators was conducted.

- An attempt was made to develop a reliability model of the tested object.

- Practical recommendations on pipeline safe operation were made.

## Abstract

The paper concerns the problem of the occurrence of failures of the high-pressure ammonia water pipeline of the coke oven battery complex, which is affected by chemical and thermal factors as well as the operating pressure occurring during its use. Pipeline failures manifested themselves as leaks (leakage of the medium) due to cracks in the area of the pipeline thermal elongation compensators. The conducted tests included, among others: visual inspection, penetration tests, macroscopic and microscopic tests as well as chemical analysis of the material. The study includes microscopic photographs of the material structure and cracks. The results of the pipeline strength and reliability analysis were also presented. On the basis of the conducted research and analyses conclusions were formulated. The assumed cause of the damage was the incorrectly made welded joints. Formulated recommendations and proposals for actions aimed at avoiding further failures of this and similar pipelines were related to the inspection time and preventive renewal.

## Keywords

coke oven battery, pipeline, welded joint, crack, structural analysis, reliability assessment.

## 1. Introduction

The use and maintenance of complex and operating in difficult conditions technical systems always carries the risk of both forms of wear and failures [5, 6, 21], which the operator did not expect. Such failures, occurring at unforeseen moments, cause the greatest difficulties, often long downtimes increasing financial losses and may pose a threat to people and the environment [11].

The causes of this type of failures may result from errors made at the design stage of the system, e.g. incorrect selection of materials or imperfections of the calculation models used. It is also often impossible to precisely determine the operating conditions, the range of variability of all factors, their impact force and interaction on the object being designed [7]. The reason for the occurrence of unforeseen failures may also be undetected defects in materials and errors in the production process of the objects which constitute the technical system [12]. A failure may be caused by inaccuracies and technological errors made during the final assembly of the system in its target place of operation, where the conditions for the performance of works are far from ideal. The multitude of causes and possibilities leading to the appearance of unexpected failures of objects during their operation means that despite various attempts and treatments, it is not possible to eliminate them and in practice, the possibility of their occurrence should always be taken into account (see [8, 13, 22]).

One of the frequently appearing and analysed problems is corrosion and cracking of pipelines transporting a medium with high temperature and an adverse effect on the pipeline material [3, 16]. The areas of welded joints and sections of pipeline change of direction are particularly vulnerable to damage in these facilities. Welded joints are critical areas in the structure. Damage in this area has been reported many times [1, 2, 14, 15] and is a problem related to stress in and strength of the construction material. Differences in the hardness of the pipeline components can cause damage initiation in the connection area [18].

The study presents the problem and analysis of the failures of a pipeline, the unusual damage of which occurred unexpectedly during its initial operation.

E-mail addresses: M. Michnej (ORCID: 0000-0003-0030-1973): maciej.michnej@pk.edu.pl, S. Młynarski (ORCID: 0000-0002-3374-7739): mlynarski_st@poczta.onet.pl, R. Pilch (ORCID: 0000-0003-2342-5776): pilch@agh.edu.pl, W. Sikora (ORCID: 0000-0002-2953-5653): wosikora@agh.edu.pl, M. Smolnik (ORCID: 0000-0003-0753-5190): smolnik@agh.edu.pl, P. Drożyner (ORCID: 0000-0002-6125-8035): przemyslaw.drozyner@uwm.edu.pl

The aim of the work is to identify the failures of the pipeline, analyse their potential causes and attempt to develop a description of reliability of the tested object. The obtained characteristics may be helpful in the assessment of technological solutions applied in the considered operating conditions.

## 2. Characteristic of the examined object

The object of the study is the high-pressure ammonia water pipeline of the coke oven battery complex. The pipeline is placed on sliding supports with certain movement restrictions and fixed supports. The supporting elements are brackets connected with the trestle bridge common to two other process pipelines. The entire route of the pipeline runs with a slope of 3 ‰ to 8 ‰ between successive supports. The pipeline is operated in the natural ambient conditions of the coke oven battery. The pipeline components are protected against weather conditions with a varnish coat. The ambient temperature of the pipeline is defined as being between -25°C in winter and + 30°C in summer. The structure of the pipeline ensures compensation of changes in its length due to temperature changes. Length change compensators were made by local U-shaped profiling of the pipeline (Fig. 1), which allows for compensation of the length by elastic deformations of the pipeline material within such a section. Solutions of this kind are commonly applied and usually do not cause operational problems. However, in the analysed case, pipeline leaks were observed in the area of the mentioned compensators. The first to be analysed are the factors related to the environment and the operating parameters of the pipeline.



*Fig. 1. View of a part of the ammonia water pipeline*

Operation of the pipeline is carried out under the conditions defined by:
- corrosivity of the environment described by category C4 according to PN-EN ISO 12944,
- maximum operating pressure – 5,5 MPa,
- opearting temperature – approximately 80°C.

After about two years of operation of the pipeline, liquid drops was observed in the area of thermal elongation compensators. The analysis of the condensate and the place of its appearance revealed the pipeline cracks and the leakage of the transported medium. The damages were found around characteristic points of the pipeline compensators. The formation of new cracks in the area of already existing damage was also observed. Due to the need for uninterrupted use of the pipeline, attempts were made to repair the pipeline in damaged places on an ad hoc basis by covering the leaky parts with additional tight external bands. The process of the failures occurrence was subjected to appropriate tests.

## 3. Localisation of the pipeline damage

The failures of the high-pressure ammonia water pipeline of the coke oven battery complex consisted of leaks (medium leaks) caused by cracks in the pipeline thermal elongation compensators, appearing on welded joints and in their immediate vicinity. The analysis of the time of occurrence of the pipeline failures showed that most of them took place during 5 months of the pipeline's operation, after about two years of its use. All the failures occurred in the area of the joints of the bends constituting parts of the pipeline compensators. Fig. 2 shows a scheme of the pipeline route and additionally marked the characteristic places where the damage was located.
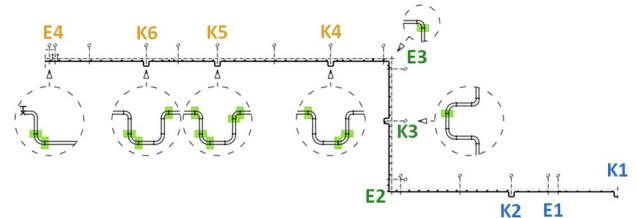


*Fig. 2. Scheme of the pipeline with marked compensators (K1-6) and changes to its route (E1-4) [10]*

The location of the damage may indicate the relationship of their formation with the local nature of the operation of a compensator profile. However, as schown in the scheme, the damage was not found in all of the affected areas. Therefore, the pipeline components were further tested.

## 4. Tests and their results

First, information was collected on the operating time of the pipeline during which the considered failures occurred. Subsequent tests were aimed at verifying the primary indications of causes and showing other possible causes of failures.

Further tests of the pipeline included an external visual inspection, during which the damaged elements were checked and the scope of the damage was determined. Then, penetrant tests, macroscopic tests, microscopic tests as well as chemical analysis and hardness measurements of the construction material of the elements were carried out. The welds, bends and pipe sections from the part of the high-pressure ammonia water pipeline supplied for testing were examined. Macroscopic examination and observation also concerned the remaining pipeline compensators.

### 4.1. External examination, penetration tests and macroscopic tests

Figures 3 and 4 show selected pipeline welds tested, marked with symbols A1, B1 and C1. There were sealing bands on the welds placed on the pipeline during its operation after the occurrence of leakage. The joint A1 was sealed with a double flange with an external band.

The sections of the pipeline with the welded joints were cut longitudinally and a part of the pipeline was cut out of the external band. Then, the parts of the pipeline in the weld zone were cleaned of paint and impurities for testing. Fig. 5 shows the parts of the A1 weld from the root side. The arrow marks the visible crack in the pipeline wall. Penetrant testing was performed to reveal any other cracks not visible to the naked eye.

Figures 6 and 7 show the pipeline cracks visible from the inside, detected by penetrant tests. Dye penetrant has leaked to the outside of the pipe in some places.

Fig. 8. shows a part of the pipeline with C1 joint after it has been cut open. Fig. 9 shows the pipeline cracks visible from the inside, detected by penetrant tests

The welded joints in the area of which cracks were revealed were cut transversely and examined after etching. The existing cracks in the

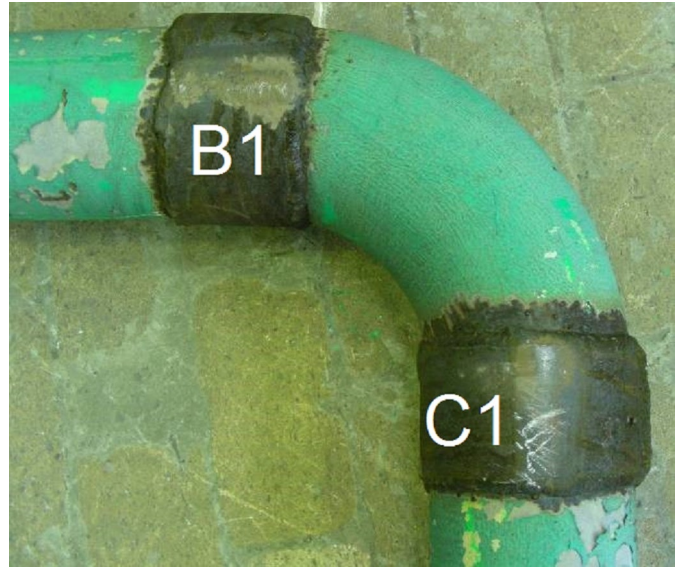*Fig. 3. A1 joint of the high-pressure ammonia water pipeline*



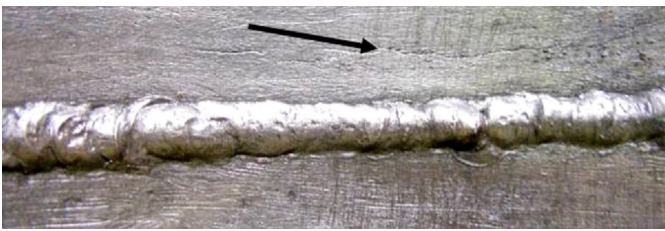*Fig. 4. B1 and C1 joints of the high-pressure ammonia water pipeline*



*Fig. 5.  View of the A1 welded joint from the root side (from the inside of the pipe)*

material of the pipeline in various places around the welded joints A1 and C1 are shown in Fig. 10. These cracks are located in the material outside the joint.

All the cracks in the pipeline near the joints were located in its bends and propagated from the inside of the pipeline. Where the crack reached the outer surface of the pipe, the medium penetrated the gap and the pipeline began to leak. Defects of joints visible in the cross-sections, such as shifts of the joined edges, may be of significant importance for the operation of the joint and the occurrence of the observed defects.



*Fig. 6. View of the A1 welded joint from the root side after penetrant testing*
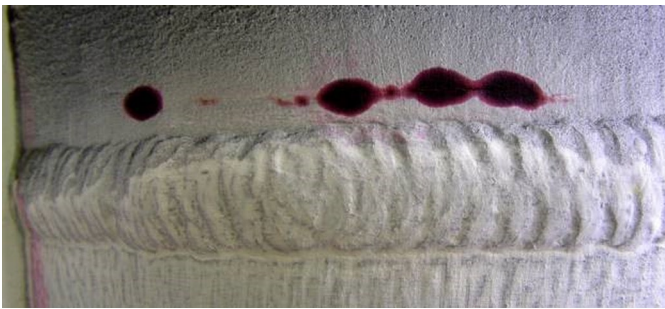


*Fig. 7.  View of the A1 welded joint from the face side after penetration of the dye penetrant from the inside of the pipe*
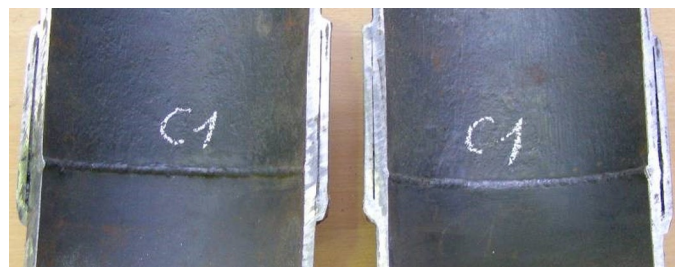


*Fig. 8. View of C1 welded joint from the root side; visible sealing band*



*Fig. 9. View of the C1 welded joint from the inside of the pipeline after penetration testing*
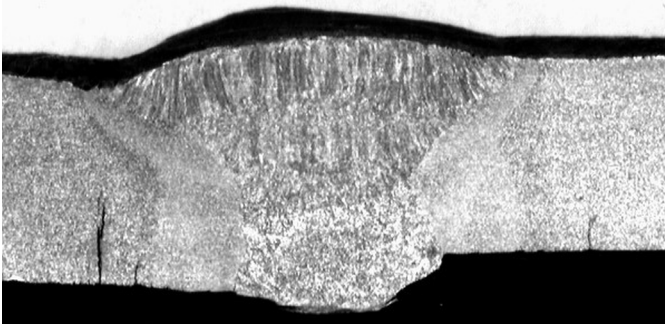
*Fig. 10. Cross-sections of A1 (left) and C1 (right) welded joints*



*Fig. 11. Structure of the pipe material. 100x magnification, nital etching*

## 4.2. Metallographic examination

The examination of the material structure of the pipeline elements was carried out on metallographic specimens obtained from parts of pipes and bends. The aim of the study was to identify the material from which the individual elements of the pipeline had been manufactured. The identification of the structure was necessary to further verify the causes of the failures. The structure of the pipe material is shown in Fig. 11 and 12 at magnifications of 100x and 500x. The observed structure was ferritic-pearlitic. The particles of perlite were lamellar. There were no irregularities in the examined structures in the form of inclusions or discontinuities.

Fig. 13 to 16 show selected damage cases observed in the bend material near the A1 and C1 welded joints.

## 4.3. Chemical analysis of the material

The chemical analysis of the material of the pipeline elements was carried out applying the spectral method. It was necessary to infer about the causes of the damage. The chemical composition has a significant impact on the physical and chemical properties of the material, which in the discussed case were highly relevant to the damage process. The test results are presented in Tab. 1. The table shows the chemical composition of the material from which the pipeline was made (requirements grade 1 boiler pipes) according to the Factory



*Fig. 12. Structure of the pipe material. 500x magnification, nital etching*



*Fig. 13. Crack in the bend material in the vicinity of the A1 welded joint. Approx. 2,5x magnification, nital etching*



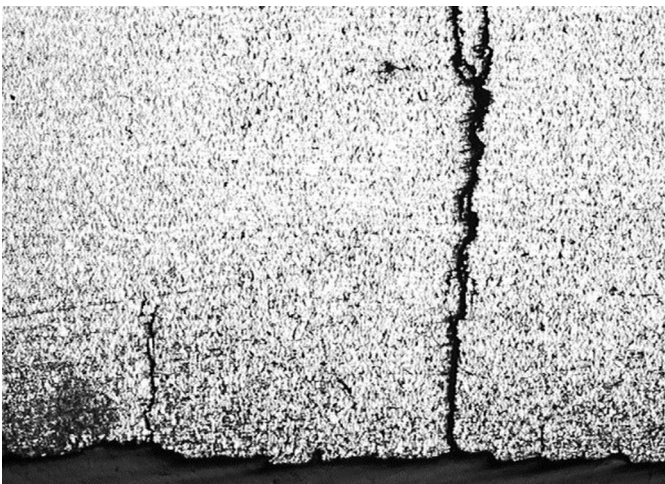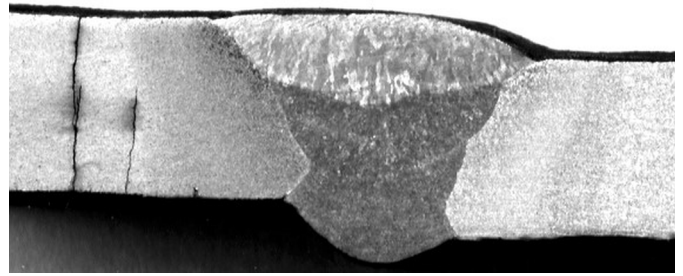*Fig. 14. Crack in the bend material in the vicinity of the A1 welded joint. 100x magnification, nital etching*
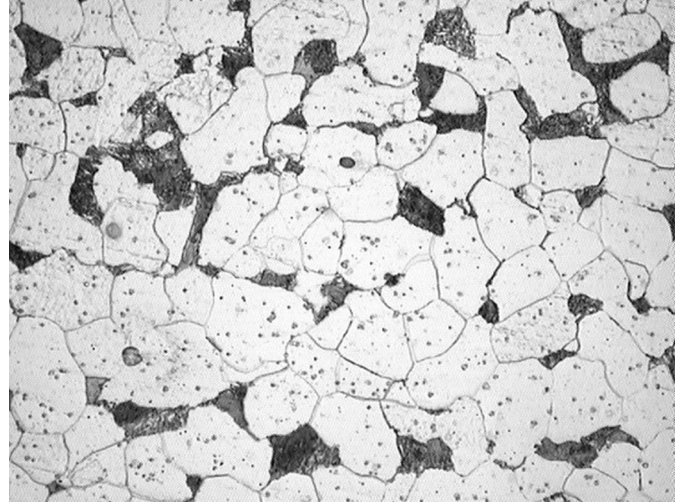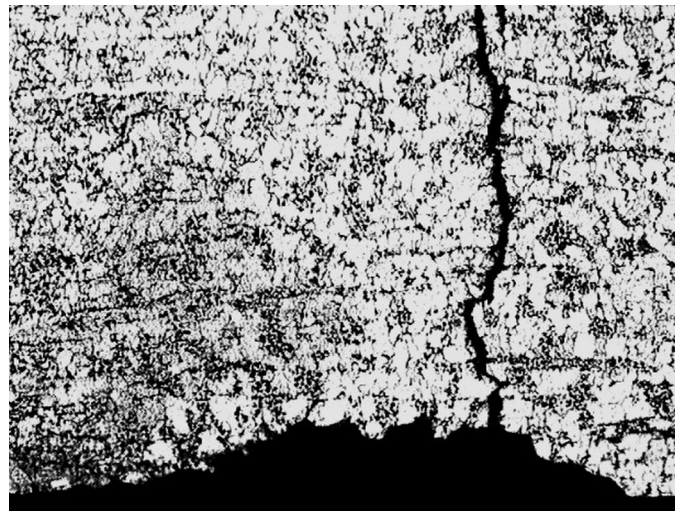


*Fig. 15. Crack in the bend material in the vicinity of the C1 welded joint. Approx. 50x magnification, nital etching*
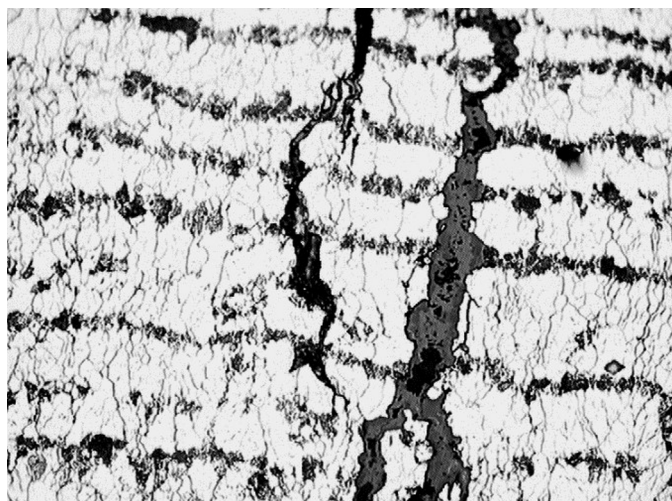
*Fig. 16. Crack in the bend material in the vicinity of the C1 welded joint. Approx. 500x magnification, nital etching*

*Table 1. Chemical compositions of the steel*

| | C | Mn | Si | P | S | Cr | Ni | Cu | Mo | Nb | V | Al | N₂ | B |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Content in % | | | | | | | | | | | | | |
| 3.1.B | 0.14 | 0.48 | 0.22 | 0.012 | 0.008 | 0.05 | 0.09 | 0.15 | 0.014 | 0.002 | 0.002 | 0.025 | 0.0084 | 0.0001 |
| A1 | 0.23 | 0.50 | 0.27 | 0.015 | 0.010 | 0.02 | 0.03 | 0.06 | 0.01 | 0.002 | 0.001 | 0.001 | | |
| B1 | 0.16 | 0.95 | 0.26 | 0.015 | 0.005 | 0.03 | 0.01 | 0.02 | 0.00 | 0.004 | 0.003 | 0.033 | | |
| C1 | 0.21 | 0.45 | 0.25 | 0.015 | 0.020 | 0.05 | 0.03 | 0.13 | 0.01 | 0.002 | 0.006 | 0.001 | | |
| P235GH | max. 0.16 | max. 1.20 | max. 0.35 | max. 0.015 | max. 0.020 | max. 0.30 | max. 0.30 | max. 0.30 | max. 0.08 | max. 0.010 | max. 0.02 | max. 0.020 | | max. 0.0001 |
| P245GH | 0.08 0.20 | 0.5 1.3 | max. 0.40 | max. 0.025 | max. 0.015 | max. 0.30 | | | | | | | | |
| P250GH | 0.18 0.23 | 0.30 0.90 | max. 0.40 | max. 0.025 | max. 0.015 | | max. 0.30 | max. 0.30 | | max. 0.010 | max. 0.02 | 0.015 0.050 | | |

Acceptance Certificate (3.1.B). The table also shows the chemical composition of steel grade P235GH, material number 1.0345 according to PN-EN 10216-2 (old marking St36K according to PN/H-84024 from 1975).

All sections of the tested pipes are made of steel of chemical composition corresponding to the chemical composition of steel grade P235GH.

The material of the bends has a chemical composition corresponding to the chemical composition of steel grade P245GH and material number 1.0352 according to DIN 102222-2 and PN-EN 10222-2 and P250GH material number 1.0460 according to DIN 17243 and PN-EN 10222-2.

As a result of the analysis, the chemical composition of the tested material in the vicinity of A1, B1 and C1 joints was determined. The results presented in the table are in accordance with the adopted standards and recommendations for materials used in the construction of ammonia water pipelines.

## 5. Structural analysis of the pipeline compensators

Pipeline route changes its direction twice, in both cases by an angle of 90°, thus it creates three straightforward segments. In each of them, there are compensators located marked by the letter ''K'' (Fig. 2) and three different colours, depending on which of the segments they are situated. Similarly marked, but with letter ''E'', are local route changes. Positions E1 and E4 refer to slight pipeline bypasses necessary because of a pre-existing infrastructure in the given area. Remaining positions E2 and E3 are the result of the pipeline main route direction

change by an angle 90°. Indicated locations along the pipeline were investigated through structural mechanics calculations.

To assess a level of nominal stress within the pipeline and rule out basic design errors, the pipeline model was prepared in the ANSYS simulation environment based on the finite element method. Using elements PIPE289 and ELBOW290, dedicated for piping calculations, its route was recreated. FEM analysis was performed under assumptions of a static and linear system. The pipeline geometry change due to deformation did not have significant effect on the obtained results. The assumed boundary conditions were as follows: outermost nodes, at the start and end of pipeline, were fixed. At locations of the pipe intermediate physical supports (80 in total), nodes were constrained in directions perpendicular to the pipeline axis. Supports were located according to the technical documentation of the pipeline. Working loads were applied in the form of internal pipe pressure equal to 5.5 MPa and internal temperature of +80°C, both compliant with the documentation.

The major factor contributing to the stress within the pipe are thermal strains. Their magnitude depends of an environmental (reference) temperature and temperature gradient across pipe wall thickness. Final external pipe temperature can reach different values as it depends on the pipe surface emissivity and, also, environmental temperature. Therefore in the performed analysis a possible range of the pipe outer surface temperature was considered (Fig. 18 and 19) which allowed to obtain a potential range of stress occurring in the pipeline. Calculations were conducted for two different environment temperatures: +20°C and −20°C. Due to higher difference between the temperatures in the regard to a transported medium, higher mean stress was observed for the second variant.

The example of the stress distribution in K2 compensator for environmental temperature and pipe outer surface temperature equal to −20°C is shown in Fig. 17. This stress results from the normal stress coming from bending and tension/compression. The figure shows its maximum value in a given location along pipeline axis. In each of the compensators, the stress distribution is very similar as in Fig. 17. The highest values can be observed in bends located by the main pipeline. Calculated specific values are, however, varied and dependent on which of the three pipeline segments given compensator is located.

Based on Fig. 18 and 19, which present the maximum values of stress in locations K and E as a function of temperature of the pipe outer surface, it can be observed that compensators K4-6 are loaded almost in the same way and stress level there is significantly lower compared to the rest of compensators. Next, according to the ascending load order, are compensators K3, K1 and K2. Higher value of stress in these pipeline parts can be explained through the lower number of compensators in the regard to the length of the given pipeline segment. In all these locations and for all temperature variants, maximum stress did not exceed yield stress for steel P235GH ($Re \approx 235\,\text{MPa}$)
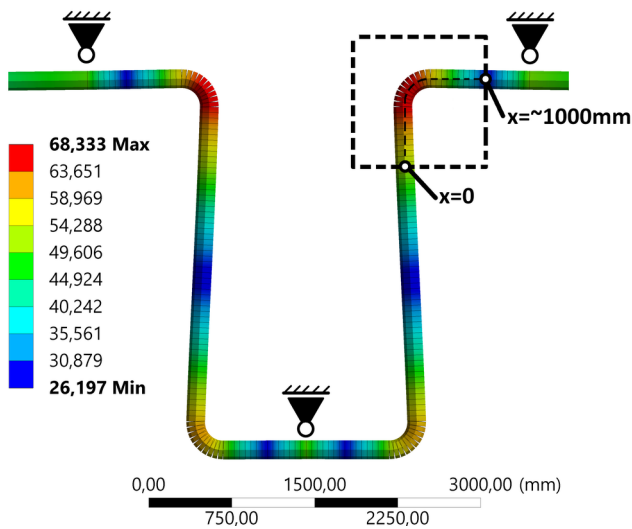
*Fig. 17. Equivalent stress [MPa] distribution in the K2 compensator for the environmental temperature and pipe outer surface temperature both equal to −20°C. Intermediate supports locations are showed. Visible deformation is scaled. Dashed rectangle shows pipeline part from Fig. 20*

from which pipes were manufactured. They remained at a safe level (max. 140 MPa). For locations marked as E, stress values were comparable with those in the compensators.

The stress distribution in the vicinity of a single bend of the compensator K2 (Fig. 2) is showed in Fig. 20. The stress is plotted as a function of the pipe length measured along its route. Maximum stress occurs within the bend, but a bit further from it, in both directions, where the actual welded joints are located, the stress is significant lower. More sharp decline appears in case of the part located on the main pipeline route, which is caused by a close proximity of the pipe support. Therefore, actual value of the stress in the place where welded joints are located should be even lower that it was initially estimated in Fig. 18 and 19.

Possible mechanism behind the failures of the investigated pipeline can be material fatigue. According to the information delivered by the operator, once per day the temperature inside pipes is decreasing to, approximately, an environmental temperature. It can be interpreted as a single cycle of a thermal load of the pipeline. The number of 365 cycles per year resulting from this, excludes possibility of material fatigue occurrence with an exception of a situation when stress in the pipes would exceed the yield stress and it could lead to a low cycle fatigue.
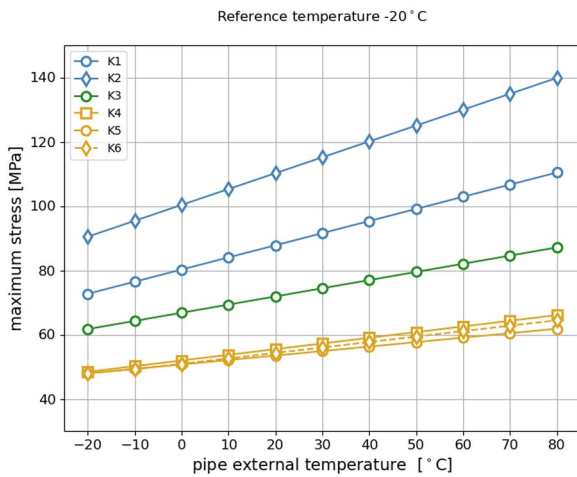


*a) for $T_{env}$=−20°C*
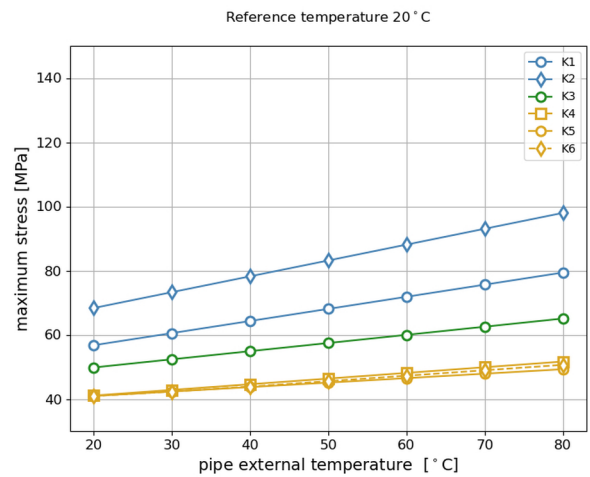


*b) for $T_{env}$=20°C*

*Fig. 18. Maximal equivalent stress [MPa] in pipeline compensators, as a function of pipe outer surface temperature for environmental temperature a)−20°C −20°C, b) 20°C*
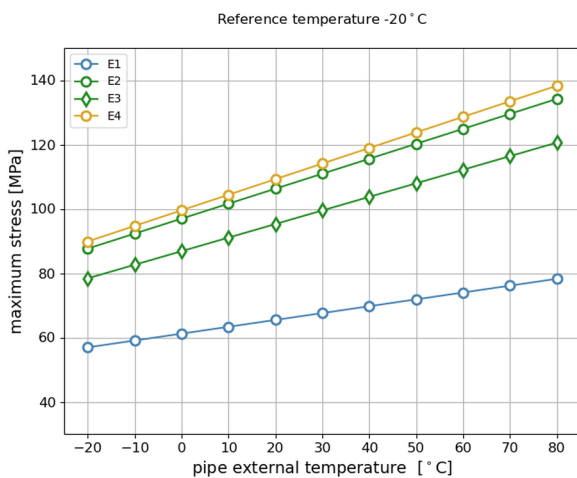


*a) for $T_{env}$=−20°C*
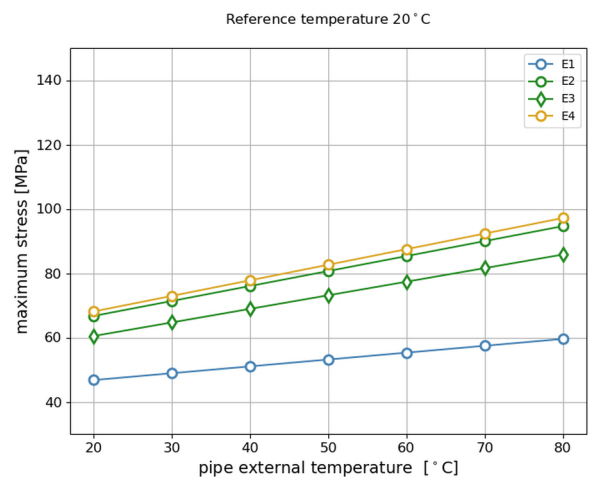


*b) for $T_{env}$=20°C*

*Fig. 19. Maximal equivalent stress [MPa] in local pipeline route changes, as a function of pipe outer surface temperature for environmental temperature a)−20°C, b) 20°C*

The value of fatigue strength for a steel of comparable mechanical properties like the one used in the investigated pipeline, according to literature [19], is ca. 170 MPa for the case of bending. This suggests that, under the assumption of lack of stress concentrators within pipes, fatigue damage in pipes should not occur, especially within very low number of cycles.

Locations of observed leakages in the actual pipeline do not overlap with possible locations that can be selected on the basis of the performed structural analysis. In the K1 and K2 compensators, which are the most stressed, no leakage was found. On the contrary, damage was present in the moderately loaded K3-6 compensators. Better agreement can be found in case of E2-4 locations where higher stress levels coincide with the actual leakages. However, the nominal stress levels should not cause any damage. This might suggest that in the investigated case there are other, not revealed yet, factors that lead to a large intensification of fatigue processes and lead finally to pipeline damage despite the lack of alarming nominal stress levels.
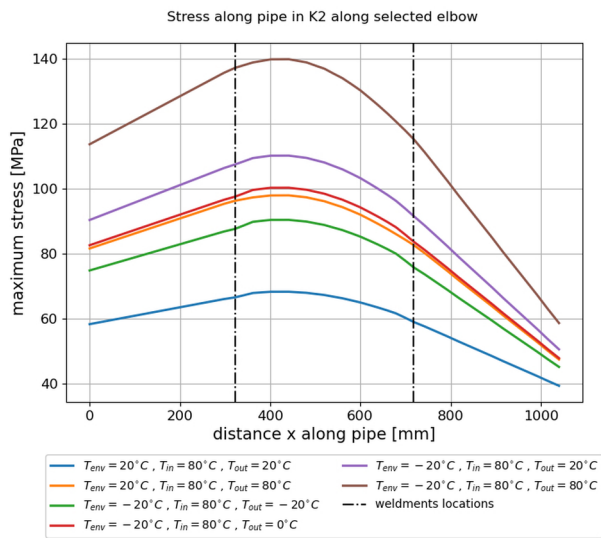


*Fig. 20. Maximal stress [MPa] distribution along pipeline in the part depicted in Fig. 17 for selected cases of temperatures. Dash-dotted lines show actual location of welded joints connecting bend with the straight pipes.*

## 6. Reliability aspects of pipeline operation

The analysis of the damage to the pipeline shows that they appeared in characteristic places and independently of each other. In the analysis, the leakage of the pipeline occurring as a result of the propagation of the crack through the whole thickness of the pipe wall was assumed as a failure. All the noted damage was located in the immediate vicinity of the welded joints connecting the bends with straight sections of the pipeline (Fig. 10). Significantly, none of the damage occurred elsewhere in the pipeline. Therefore, it was assumed that in the reliability analysis, the pipeline will be considered as an object consisting of 56 identical elements, operating in the same conditions and which may fail independently of each other. Each pipeline part in the immediate vicinity of the welded joint connecting the bend with the straight section is treated as a single element. The assumed number of elements of the object results from the fact that the entire pipeline has 28 bends, and each of them is connected to straight sections with two welds, each of which constitutes a potential damage area.

With the data on the time of failures occurrence and the time of commencement of operation of the pipeline, the operating times to failure were determined for those of the 56 identified elements that were damaged (Tab. 3). Until the end of the observation, the remaining elements were not damaged and the whole data set was treated as including truncated data.

*Table 3. Times till failure of the elements of the pipeline*

| Element No. | Time till failure [days] |
|---|---|
| 1 | 945 |
| 2 | 984 |
| 3 | 986 |
| 4 | 1005 |
| 5 | 1014 |
| 6 | 1048 |
| 7 | 1057 |
| 8 | 1064 |
| 9 | 1075 |
| 10 | 1075 |
| 11 | 1075 |
| 12-56 | >1076 (truncated data) |

Based on Tab. 3, the course of the empirical reliability index, which is the failure stream parameter ($\omega^*(\Delta t)$), was determined and presented in Fig. 21.

$$\omega^*(\Delta t) = \frac{n'(\Delta t)}{n_0 \cdot \Delta t} \qquad (1)$$

where:

$n'(\Delta t)$ – number of elements which failed for the first or subsequent time in a given time range $\Delta t$,

$n_0$ – number of elements tested,

$\Delta t$ – length of the adopted time range.
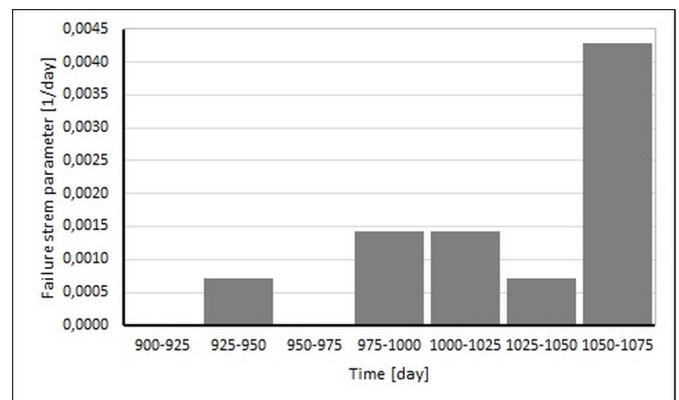


*Fig. 21. Failure stream parameter values*

The observation of its course shows that after the 925th day of operation, the value of the parameter becomes greater than zero and begins to increase. Generally, such a fact is interpreted as the time the technical object enters the aging period. In this case, the beginning of the increasing course over time can be interpreted as the beginning of the period in which the cracks initiated by corrosion and propagating from the inside of the pipeline, in some elements, already reach a length equal to the wall thickness and cause the failures observed as a pipeline leak. It can be assumed that in the further operation, the number of such failures will increase since the subsequent considered elements will most likely be damaged in the same way, and the observed values of failure stream parameter will continue to increase.

Treating the set of 56 distinguished elements as observed till their first failures, the working times till failure of 11 of them were recorded and the rest were considered as truncated data. On this basis,

the Kaplan Meier analysis was performed. The determined graph of the survival function is shown in Fig. 22.
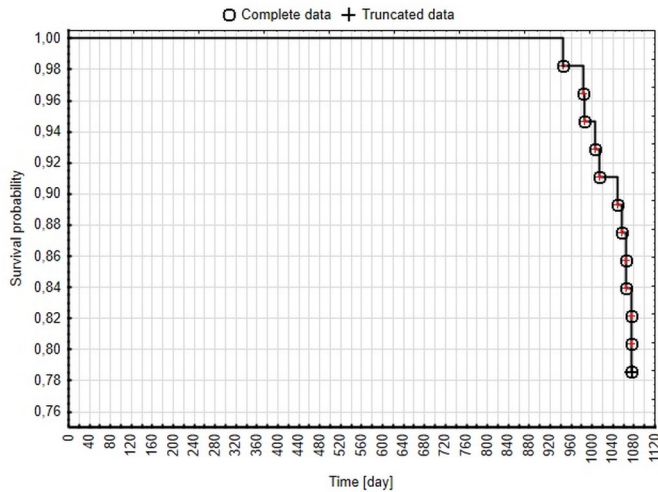


Fig. 22. Survival function for truncated data in the Kaplan-Meier analysis

The graph indicates the probability of surviving without a failure for a certain number of days for each individual element. This value drops very quickly after reaching the operating time of 940 [days]. It can be concluded that each of the remaining available (undamaged) components is less and less likely to survive the next days without a failure. As the probability decreases quickly, it is possible to predict the appearance of many successive failures in the observed set of elements over a short time.

The third stage of the reliability analysis is to determine the probability distribution of operating time till failure for the observed elements. On the basis of the collected data on time till failures and truncated data (Tab. 3), the Weibull distribution was fitted using the Statistica computer program as a mathematical model of the operating time till failure. The parameters determined by the maximum likelihood method are as follows: $\alpha = 24.96$; $\beta = 1139.6$ [days]. The probability density function is presented as [17, 20]:

$$f(t) = \alpha \cdot \left(\frac{1}{\beta}\right)^{\alpha} \cdot t^{\alpha-1} \cdot e^{\left(-\left(\frac{t}{\beta}\right)^{\alpha}\right)}$$  (2)

where:
    $\alpha$ – shape parameter,
    $\beta$ – scale parameter.

The initial part of the cumulative distribution function F(t) (for complete data) with the determined 95% confidence interval is presented in Fig. 23.

Next, for the obtained distribution, the course of the forecasted failure probability density function f(t), cumulative distribution function F(t) and failure rate $\lambda(t)$ were plotted over a period longer than the time adopted for the complete data (Fig. 24, 25).

The graphs show an accumulation of the failure probability density between 1000 and 1200 operation days and a rapid increase in cumulative distribution function values during this time. The determined course of the failure rate shows its characteristic feature which is the intense increase over time. From this fact it can be directly concluded that the probability of failure of each element increases significantly with each subsequent period that it survives without a failure.

The determined changes in the values of the failure stream parameter, the survival function as well as the cumulative distribution function, the failure rate and the expected value of operation time till failure amounting to 1115 [days] allow one to suggest that in order to avoid subsequent failures of the pipeline, appropriate preventive
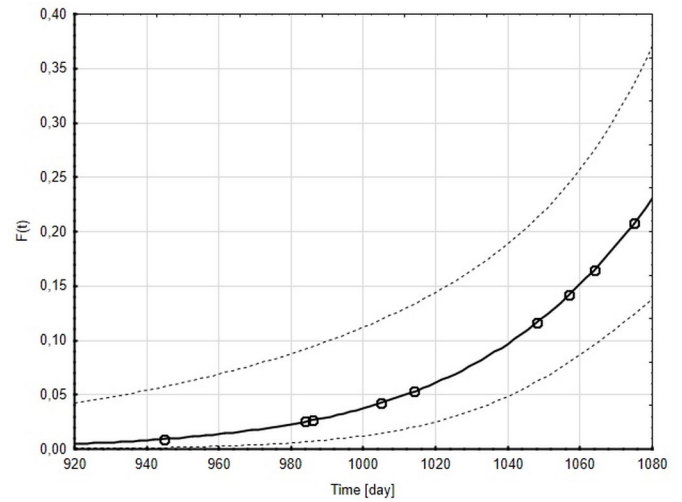


Fig. 23. Cumulative distribution function course for a single pipeline element with 95% confidence interval

a)
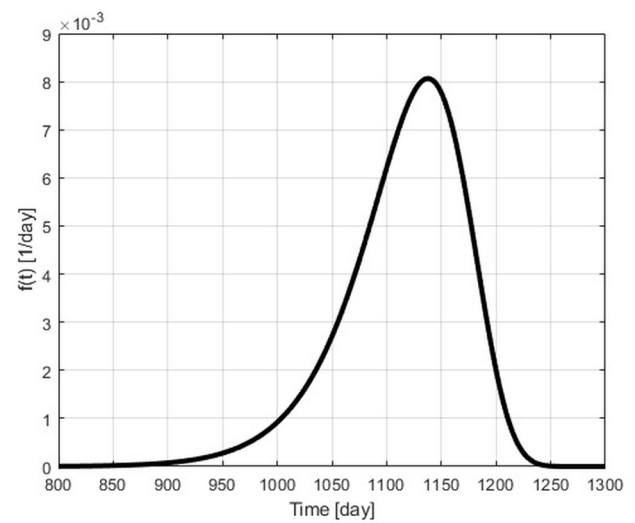


b)
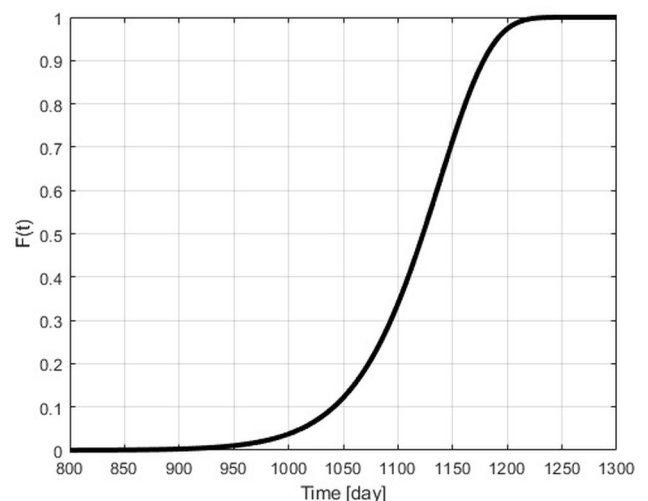


Fig. 24. Failure probability density function a) and cumulative distribution function of the operation time till failure b)

measures should be taken with regard to all available (undamaged) elements, as the probability of failure indicated by the indexes is high and increases rapidly over time. It should be borne in mind that the assumption adopted in the reliability analysis about the crack initiation and propagation in the welded joints areas does not actually have
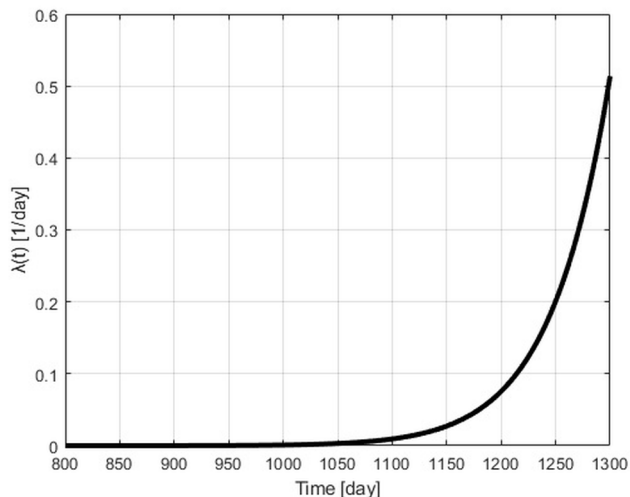
*Fig. 25. Failure rate*

to take place in each of them. However, due to: the method of making welded joints (without stress relief annealing), aggressive affecting of ammonia water on the pipeline material, high operating temperature and variable stresses, the appearance of a microcrack and its propagation from the inside to the outside of the wall is very likely and confirmed by observations in a number of objects of this type operated in very similar conditions [4, 9]. Hence, the suggested preventive measures in areas where leaks have not yet occurred seem to be the most justified, especially if the costs of carrying them out are lower than the later losses resulting from the failures.

Another possible action is to verify whether a propagating fracture has already appeared in a given area, for which non-destructive ultrasonic testing can be applied. Such tests may be repeated periodically.

## 7. Conclusions

As a result of the tests of the delivered parts of the high-pressure ammonia water pipeline of the coke oven battery complex, the following conclusions were formulated.

The pipe sections of the tested part of the pipeline were made of low-carbon steel of a chemical composition corresponding to the

chemical composition of the steel grade P235GH. The bends in the tested section of the pipeline were manufactured from a material of a different chemical composition than the pipe sections. The B1 bend material has a chemical composition corresponding to the chemical composition of steel grade P245GH. The material of bends A1 and C1 has a chemical composition corresponding to the chemical composition of steel grade P250GH.

The circumferential welded joints were made by arc welding. The external appearance of the welds raises the following reservations: variations in the shape of the weld face are observed, the occurrence of asymmetry of the joint was found (it is bound with unequal folding of the sections of the pipeline welded together), metal spattering appears on the outer surface of the pipes welded, and unevenness of the weld root was found.

Transverse cracks (in relation to the pipeline axis) appear in the area of the welded joints. They are located in the material of the bends outside the heat-affected zone. Cracks propagated from the inside of the bends, and some of them covered the entire thickness of the pipeline wall. As a result, leaks appeared in these places and the necessity to put on sealing bands.

It may be assumed that it was the incorrectly made welded joints that caused excessive tensile stresses in the surrounding material. These stresses present in the object operating in the corrosive environment, which is the medium flowing through the pipeline, led to the formation of damage and leakage of the pipeline. On the basis of the conducted tests and the obtained results, it is difficult to unequivocally state what was the cause and whether there was only one cause of the failures. The structural analysis of pipeline compensators led to similar conclusions.

The statistical and reliability analyses of the failures occurrence suggest that in the considered case certain preventive actions should be performed to avoid the subsequent failures of the pipeline.

A recommendation affecting the safe operation of ammonia water pipelines is monitoring the tightness of compensators after 1000 days of operation.

Another recommendation, developed on the basis of the reliability characteristics of the tested pipeline structure, is the necessity to quickly repair all endangered parts of the structure after detecting the first damage to the pipeline.

## References

1. Abedi S. Sh., Abdolmaleki A., Adibi N. Failure analysis of SCC and SRB induced cracking of a transmission oil products pipeline. Engineering Failure Analysis, 2007; 14 (1): 250–261, https://doi.org/10.1016/j.engfailanal.2005.07.024.
2. Aljoboury A., Mourad A., Alawar A., Abou Zour M., Abuzeid O. Stress corrosion cracking of stainless steels recommended for building brine recirculation pumps. Engineering Failure Analysis, 2010; 17 (6): 1337–1344, https://doi.org/10.1016/j.engfailanal.2010.03.008.
3. ASM Handbook. Volume 13C, Corrosion: Environments and Industries. ASM INTERNATIONAL, 2006, doi: 10.31399/asm.hb.v13c.9781627081849.
4. Banaszek A., Łosiewicz Z., Jurczak W. Corrosion Influence on Safety of Hydraulic Pipelines Installed on Decks of Contemporary Product and Chemical Tankers. Polish Maritime Research, 2018; 25 (2): 71–77, https://doi.org/10.2478/pomr-2018-0056.
5. Borucka A. Method of testing the readiness of means of transport with the use of semi-Markov processes, Transport 2021; 36 (1), https://doi.org/10.3846/transport.2021.14370
6. Borucka, A. Three-state Markov model of using transport means. Business Logistics In Modern Management, 2018; 3-19.
7. Campione G., Giambanco G. Influence of design mistakes and material degradation on the collapse of a long-span RC roof in South Italy. Engineering Failure Analysis, 2020; 111, https://doi.org/10.1016/j.engfailanal.2019.104257.
8. Cao Y., Chang Q., Zhen Y. Numerical simulation of fracture behavior for the pipeline with girth weld under axial load. Engineering Failure Analysis, 2022; 136, https://doi.org/10.1016/j.engfailanal.2022.106221.
9. Dobosiewicz J., Brunné W. Przyczyny nieszczelności rurociągu wody amoniakalnej w obszarze połączeń spawanych. Causes of pipeline leakage ammonia water near to welded joints. Przegląd Spawalnictwa, 2011; 83 (7): 14–16, https://doi.org/10.26628/ps.v83i7.530
10. Documentation of the project U-28062. Object 251 Inter-raw gas network. High ammonia-water piping.
11. Hafez K. M. The role of a plain dent on the failure mode of a crude oil pipeline. Engineering Failure Analysis, 2021; 122, https://doi.org/10.1016/j.engfailanal.2021.105291.
12. Khosravani M., Božić Ž., Zolfagharian A., Reinicke T. Failure analysis of 3D-printed PLA components: Impact of manufacturing defects and thermal ageing. Engineering Failure Analysis, 2022; 136, https://doi.org/10.1016/j.engfailanal.2022.106214.
13. Kozłowski E., Borucka A., Swiderski A., Gil, L. Predicting the Fatigue Life of a Ball Joint. Transport and Telecommunication, 2021; 22(4): 453-460, doi:10.2478/ttj-2021-0035

14. Mistur L. Spawanie i napawanie w naprawach części maszyn i konstrukcji metalowych. Wydaw. i Handel Książkami "KaBe", Krosno 2003.

15. Movafeghi A., Mohammadzadeh N., Yahaghi E. et al. Defect Detection of Industrial Radiography Images of Ammonia Pipes by a Sparse Coding Model. Journal of Nondestructive Evaluation, 2018; 37(3), https://doi.org/10.1007/s10921-017-0458-9.

16. Nyborg R., Lunde, L. Measures for reducing SCC in anhydrous ammonia storage tanks. Process Safety Progress, 1996; 15 (1): 32–41, doi: 10.1002/prs.680150110.

17. Pham H. (ed.). Handbook of Reliability Engineering. Springer-Verlag, London, 2003. https://doi.org/10.1007/b97414

18. Schweitzer P. (ed.). Carbon steel and low alloy steel in corrosion and corrosion protection handbook. Marcel Dekker, New York 1983.

19. Skoć A., Spałek J. Podstawy Konstrukcji Maszyn, Vol. 1, WNT, Warszawa 2006.

20. Tobias P. A., Trindade D. C. Applied Reliability. Third Edition. CRC Press Taylor & Francis Group, Boca Raton, 2012.

21. Wang W., Zhang Y., Li Y., Hu Q., Liu C., Liu C. Vulnerability analysis method based on risk assessment for gas transmission capabilities of natural gas pipeline networks. Reliability Engineering & System Safety, 2022; 218 (B), https://doi.org/10.1016/j.ress.2021.108150.

22. Zhu L., Luo J., Wu G., Han J., Chen Y., Song C. Study on strain response of X80 pipeline steel during weld dent deformation. Engineering Failure Analysis, 2021; 123, https://doi.org/10.1016/j.engfailanal.2021.105303.
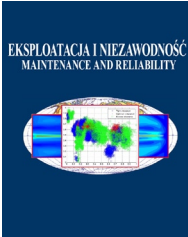
# Virtual–real fusion maintainability verification based on adaptive weighting and truncated spot method

Zhexue Ge[a,*], Yi Zhang[b], Fang Wang[a], Xu Luo[a], Yongmin Yang[a]

[a] National University of Defense Technology, Laboratory of Science and Technology on Integrated Logistics Support, School of Intelligent Science and Technology, De Ya Road, 109, Changsha, Hunan 410073, P. R. China
[b] China Airborne Missile Academy, Luoyang, Henan 471009, P. R.China

## Highlights

- Maintainability test is carried out in virtual–real fusion scenario with lower cost.

- The error of virtual–real fusion maintainability evaluation is reduced.

- The weights of real and virtual person test data are adaptively determined.

- Less time is consumed than the traditional virtual maintainability test.

## Abstract

Maintainability is an important general quality characteristic of products. Insufficient maintainability will lead to long maintenance time and high maintenance cost, thus affecting the availability of products. Maintainability verification is an important means to ensure maintainability meets design requirements. However, the cost of traditional real maintainability verification method is very high, and the virtual maintenance method has insufficient verification accuracy due to the lack of large maintenance force feedback when the human body is moving. In order to reduce the evaluation error and test sample size, the paper conducts maintainability verification based on the mixed physical and virtual maintainability test scenarios. Aiming at the problem that traditional methods are difficult to deal with the real test information and synchronous virtual simulation information in the test process, this study proposes a virtual–real fusion maintainability evaluation algorithm based on adaptive weighting and truncated SPOT (Sequential Posterior Odd Test) method. It can weigh real test information and virtual human simulation information adaptively to obtain a virtual–real fusion maintainability test sample. Then, the SPOT method is used to evaluate the maintainability of small samples. The adjustment of valve clearance, replacement of air filter element and replacement of starting motor maintenance tasks of ship engine are taken as examples for demonstration. The virtual–real fusion and virtual maintainability verification methods are respectively used for verification, and compared with the physical maintenance scenario constructed by 3D printing, indicating that the accuracy of virtual–real fusion maintainability test verification is 89%, while the virtual maintainability verification is only 33%.

## Keywords

virtual–real fusion maintainability, maintainability verification, adaptive weighting method, truncated SPOT method.

## 1. Introduction

### 1.1. Requirement analysis

Maintainability, like reliability, is an important general quality characteristic of products [23]. Insufficient maintainability will lead to long maintenance time and high maintenance cost, thus affecting the availability of products [5, 25]. Maintainability verification is an important means to discover the defects of product maintainability design and ensure that the qualitative and quantitative requirements of maintainability are met [16]. The traditional maintainability test verification is carried out on the physical equipment and real maintenance environment, so the verification is accurate, but the test cost is high and the test cycle is long [2], as shown in Fig.1. The virtual maintainability test is carried out on the digital prototype of the product as shown in Fig.2, which can reduce the requirements of the physical test prototype, which has become a highly concerned maintainability verification method [12]. Virtual reality technologies such as motion capture and data glove technology can be used to achieve the virtual maintenance operation by real human. However, in the process of maintainability test, there is always a lack of force feedback mechanism that can adapt to large scene and large maintenance force, and there will be

(*) Corresponding author.

E-mail addresses: Z. Ge (ORCID :0000-0002-9199-8042): gzx@nudt.edu.cn, Y. Zhang (ORCID :0000-0002-9823-1792): zhangyi1466@gmail.com, F. Wang (ORCID :0000-0003-4925-7002): 18739002732@163.com, X. Luo (ORCID :0000-0002-7592-6962): luoxu2002@gmail.com, Y. Yang (ORCID :0000-0002-2114-1078): yangyongmin@163.com

various errors such as virtual environment positioning error, motion capture delay, collision feedback delay, etc. [27], which will greatly affect the accuracy of maintainability evaluation [6]. How to reduce the test cost while ensuring the maintainability verification accuracy is an important problem to be solved urgently.

## 1.2. Overview

In the past 20 to 30 years, many scholars have carried out a lot of new technology research to solve the problem that traditional maintainability verification depends on physical test and requires a high sample size. They mainly focus on two aspects.



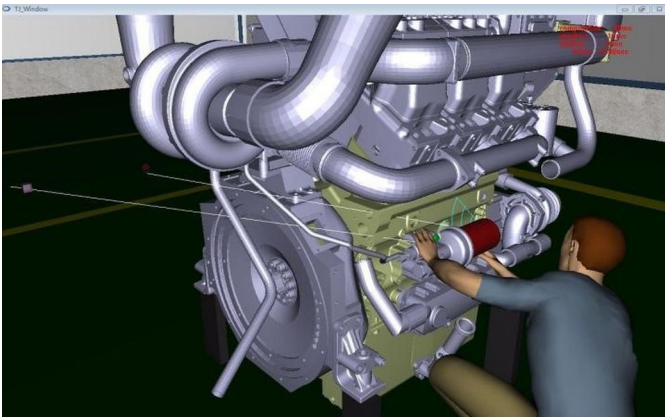*Fig. 1. The real maintainability verification*



*Fig. 2. The virtual maintainability verification*

First, the maintainability test still uses the physical prototype, but the cost is reduced by less test times, and several new maintainability test data processing and verification methods are studied. The number of samples given in the maintainability standard is at least 30 [8, 20]. Miao et al. adopt the idea of segmentally weighted verification and propose the segmentally weighted verification (SWV) method to realize in-lab data verification. Then, the Dempster–Shafer evidence theory based integrative verification method is presented to solve the problem of in-lab and field data combination [19]. Wu et al. propose a novel prior distribution elicitation method for MTTR Bayesian demonstration. The test requires fewer samples than traditional methods that require no less than 30 samples relies heavily on expert experience and can be time consuming if performed manually [31].

The second is to adopt the virtual maintenance based test method, mainly focusing on how to improve the fidelity of human-computer interaction [10-12, 15, 24]. Desktop virtual maintenance is relatively simple [29]. In reference [24], a desktop virtual reality-based integrated system is developed for complex product maintainability verifica-

tion. Guo et al. review the application of virtual reality technology in product maintenance, and deeply analyze the application field and effect, virtual reality hardware, development platform and current main research focus [12]. In reference [17], Luo et al. propose a method for quantitative evaluation of maintainability based on qualitative attributes of maintainability. The development of virtual reality technology can enhance the immersion and simulation fidelity of the maintenance process; thus, [10] and [11] combine the operation information of a real person and virtual information to carry out the maintainability evaluation and obtain higher evaluation accuracy. The difficulty of haptic and haptic interaction is a key problem in virtual reality [22]. The tactile feedback based on the data glove and the small force feedback of the hand under the fixed position are relatively mature [18]. Overtoom etc. provide a systematic overview of the literature assessing the value of haptic and force feedback in simulators teaching laparoscopic surgical skills [22]. It is still very difficult to apply the tactile feedback when the human body is moving [27]. The maximum feedback force is generally only 20N, which is difficult to meet the feedback needs of maintenance operations [21, 26]. So there are some researches on the modification of human model data by compensating for the influence of external factors on human motion[9, 28, 33]. For example, Grochow et al. propose an inverse kinematics method based on physical kinematics characteristics, which combine global nonlinear dimensionality reduction technology, Gaussian process latent variable model (GPLVM), and a priori kinematics model. It is suitable for correcting similar small-scale human motion data [9]. Seemann et al. propose a method to generate a modified new trajectory by projecting the observed position, velocity, and acceleration on the corresponding constrained manifold, ensuring the consistency of motion parameters [28]. Reference [33] proposes a hybrid method of real-time human motion capture using simplified marker sets and monocular video, then an improved inverse motion solver is used to estimate pose based on marker positions.

## 1.3. Focused questions

Through the analysis of the current situation, we can find that the current methods still have some limitations. For complex equipment, such as ship and aircraft, when the traditional physical maintainability verification method is used, even if only a small number of samples are needed, a very complex test scenario must be built in order to reflect the impact of complex cabin environment on maintainability with unbearable cost [12]. In addition, the current methods still lack universality in solving the problem of virtual maintenance fidelity, and are difficult to solve the impact of lack of force perception on maintenance time and comfort. Therefore, considering the economy and accuracy of product test, the combination of virtual and physical test has received more and more attention [1, 30, 32, 34]. In the early stage, we carried out research on the construction technology of maintainability test environment based on virtual–real integration scenarios [6]. The maintenance test operation is carried out on the physical equipment, whereas the maintenance obstacles and surrounding equipment with less operations use virtual prototypes. All virtual prototypes are presented through AR glasses, which can effectively simulate the real maintenance scene and produce a good sense of immersion, as shown in Fig.3.

In the previous stage, a virtual maintainability test information fusion method based on t-test and F-test was proposed [7]. This study is to solve the problem of large maintainability evaluation error due to the introduction of virtual prototypes based on the virtual–real fusion maintainability test mechanism. The evaluation information of real human and the evaluation results of synchronous virtual human in the process of immersive and virtual–real fusion maintainability test is fully considered. The fusion method is used to reduce the uncertainty, the virtual and real data-level fusion is realized through adaptive weighting, and the truncated SPOT method is used for verification.

*Fig. 3. Schematic of virtual reality integration*

The rest of this paper is organized as follows: In Section 2, the proposed maintenance index verification method based on virtual and real information fusion is presented and discussed. In section 3, the maintenance tasks of ship engine are taken as examples for demonstration. Finally, in Section 4 the conclusions are provided.

## 2. Maintenance index verification method based on virtual and real information fusion

By carrying out the virtual–real integration maintainability evaluation test, the real maintainability test information and the virtual human maintainability test information can be obtained. In the test based on virtual–real fusion scene, the maintenance human carries out maintenance tasks on real maintenance object, so the obtained maintainability data is close to the real maintenance data. Meanwhile, because other equipment and environments around the real maintenance object are presented in the maintainer's field of vision through AR glasses, there will be some real registration errors and time delays, so the maintainability data has some deviations. The errors can be effectively avoided by using the virtual human test information obtained by motion capture during the test process. Therefore, the real and virtual human test data in the virtual–real integration maintainability test has strong complementarity.

The paper fully fuses the multi-source information in the test process to obtain the fused test samples, reducing the uncertainty of the underlying test data. In addition, in order to reduce the requirement for the number of test samples, the maintainability verification method based on small samples is studied. A truncated SPOT maintainability verification method based on the virtual–real fusion data is proposed, as shown in Fig. 4 to avoid the influence of motion capture error, virtual–real registration error, and tactile feedback error on the maintainability data in the virtual–real fusion maintainability evaluation test as shown in Fig. 4. Initially, for the same maintenance task, multiple maintenance tests are carried out in the virtual–real fusion scene, and the real test data and virtual human test data are obtained at the same time. Then, the adaptive weighted fusion algorithm is used to effectively fuse the two maintainability test datasets. Finally, the fused data are used as the field data, and the truncated spot maintainability verification method is used to analyze and verify the virtual–real fusion maintainability results to judge the rationality and effectiveness of the virtual–real fusion maintainability evaluation test scheme.
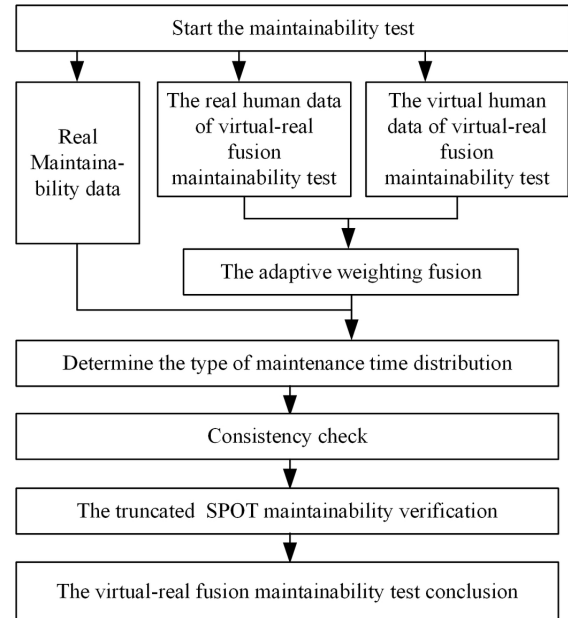


*Fig. 4. Virtual–real fusion maintainability evaluation test method*

### 2.1. Fusion method of virtual and real test data based on adaptive weighting algorithm

Inspired by the adaptive weighted fusion algorithm in the literature [13], this study regards the maintenance test results in different scenarios as the results obtained by different sensors measuring the same maintainability index in the maintenance process. The measurement results of the indicators are different. The adaptive weighted fusion method is used to perform data fusion obtained from different maintenance tests.

The real value of the maintainability index is assumed to be $X$. In the virtual–real fusion maintainability test, the maintainability index obtained by the real human evaluation is $Xr$, and the virtual human evaluation index obtained synchronously is $Xx$; their variances are $\sigma_r^2$ and $\sigma_x^2$, respectively. The corresponding fusion weights are $\varphi_r$ and $\varphi_x$. According to the fusion model structure, we can obtain the following:

$$\begin{cases} X_R = \varphi_r X_r + \varphi_x X_x \\ 1 = \varphi_r + \varphi_x \end{cases}$$

The total variance is as follows:

$$\sigma^2 = E\left[\left(X - \overline{X}\right)^2\right] = E\left[\begin{matrix} \varphi_r^2(X - X_r)^2 + \varphi_x^2(X - X_x)^2 \\ + 2\varphi_r\varphi_x(X - X_r)(X - X_x) \end{matrix}\right]$$

$X_r$ and $X_x$ are the maintainability data obtained in different maintenance scenarios; thus, they are independent of each other and are unbiased estimates of X, as follows:

$$E(X - X_r)(X - X_x) = 0$$

Therefore,

$$\sigma^2 = E\left[\varphi_r^2(X - X_r)^2 + \varphi_x^2(X - X_x)^2\right] = \varphi_r^2\sigma_r^2 + \varphi_x^2\sigma_x^2$$

The accuracy of the maintenance test data fusion results is inversely proportional to the size of the total variance. Therefore, when the fusion variance is the smallest, the accuracy of the fusion result is the highest. That is, the weights should satisfy the following:

$$\min \sigma^2 = \min\left( (\sigma_r^2 + \sigma_x^2)\varphi_r^2 - 2\sigma_x^2\varphi_r + \sigma_x^2 \right)$$

By Lagrange multiplier method, we can obtain the following [3]:

$$\varphi_r = \frac{\sigma_x^2}{\sigma_r^2 + \sigma_x^2} \ , \ \ \varphi_x = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_x^2}$$

Subsequently, $X_r(i)$ represents the result of the $i$-th group of maintainability tests based on the virtual–real fusion scenario, and $X_x(i)$ represents the result of the $i$-th group of simulated maintainability tests based on the virtual maintenance scenario. The average value of the first $k$ groups of the maintainability test data is calculated as follows:

$$\overline{X}(k) = \frac{1}{2k} \sum_{i=1}^{k} (X_r(i) + X_x(i))$$

Then, the estimated variance of the $k$-th group of maintainability tests based on the virtual–real fusion scenario can be expressed as follows:

$$\sigma_{re}^2(k) = (\overline{X}(k) - X_r(k))^2$$

The final variance of $k$ groups of maintainability tests based on the virtual–real fusion scenario is obtained as follows:

$$\sigma_r^2(k) = \frac{1}{k} \sum_{i=1}^{k} \sigma_{re}^2(i)$$

Similarly, the final variance value of the $k$ groups of maintainability test based on the virtual maintenance scenario can be obtained as follows:

$$\sigma_x^2(k) = \frac{1}{k} \sum_{i=1}^{k} \sigma_{xe}^2(i)$$

Then, the fusion data $X_R$ of $k$ groups of maintainability test are obtained as follows:

$$X_R(k) = \frac{\sigma_x^2(k)}{\sigma_r^2(k) + \sigma_x^2(k)} X_r(k) + \frac{\sigma_r^2(k)}{\sigma_r^2(k) + \sigma_x^2(k)} X_x(k)$$

## 2.2. Maintenance time distribution type determination

Generally, the maintenance time obeys the log-normal distribution [14, 31], and the Kolmogorov method is used to test and analyze the real maintenance test data and the virtual–real fusion maintainability data to judge whether the log-normal distribution is obeyed. In the truncated SPOT method, the accumulated maintenance test data isused as the pre-test historical data, and the virtual–real fusion maintainability data is used as the field test data to verify the maintainability.

## 2.3. Consistency check

The variance and mean test method is used to test whether a significant difference between the real maintenance test data and the virtual–real fusion maintainability data. The two data parameters must be consistent to carry out maintainability verification.

## 2.4. Maintainability Verification

Let the maintenance time be $Y$, assuming that $Y_d = \ln Y$ obeys a normal distribution $Y_d \sim N(\theta, \sigma^2)$, where $\sigma^2$ known, or an estimate of its appropriate accuracy can be obtained from previous data. $\theta$ is an unknown parameter of the overall distribution and can be known by analysis and calculation based on real maintenance test data. According to the contract, the index value of the mean repair time (MTTR) is $\theta_0$, the risk of the contractor is $\alpha$, and the risk of the subscriber is $\beta$. The MTTR can be verified by the following methods.

The following assumptions are made:

$H_0: \theta = \theta_0 \ \ H_1: \theta = \theta_1 = \lambda\theta_0 > \theta_0$, $\lambda > 1$, where $\lambda$ is the detection ratio agreed by the manufacturer and the subscriber. In general, $1.2 \le \lambda \le 1.5$.

For virtual–real fusion maintainability data time samples $X_d = (X_{d1}, \cdots, X_{dn})$, the post-test weighting ratio is obtained as follows:

Suppose: $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1 = \lambda\theta_0 > \theta_0$,

where $\lambda$ is the detection ratio, and $\lambda > 1$. In general, $1.2 \le \lambda \le 1.5$.

For virtual–real fusion maintainability data time samples $X_d = (X_{d1}, \cdots, X_{dn})$, the posteriori weighting ratio is obtained as follows:

$$O_n = \frac{P_1}{P_0} \cdot \frac{L(X_d \mid \theta_1)}{L(X_d \mid \theta_0)}$$

where $P_0 = \int_{-\infty}^{\theta_0} \frac{1}{\sqrt{2\pi}v} e^{-\frac{(\theta - \overline{Z_d})^2}{2S_{Z_d}^2}} \, d\theta$, $P_1 = 1 - P_0$.

The introduction of constants $A, B$, $0 < A < 1 < B$. According to Wald's point of view, $A = \frac{\beta}{P_0 - \alpha}$, $B = \frac{P_1 - \beta}{\alpha}$. The following judgment rules shall be adopted [4]:

(1) If $O_n \le A$, then the virtual–real fusion maintainability data satisfy the maintainability requirements, and the virtual–real fusion maintainability evaluation scheme is feasible.

(2) If $O_n \ge B$, then the virtual–real fusion maintainability data do not satisfy the maintainability requirements, and the feasibility of the virtual–real fusion maintainability evaluation scheme is poor.

(3) If $A < O_n < B$, then proceed to the next step.

① If $A < O_n < C$, then the virtual–real fusion maintainability data satisfy the maintainability requirements, and the virtual–real fusion maintainability evaluation scheme is feasible.

② If $C < O_n < B$, then the virtual–real fusion maintainability data do not satisfy the maintainability requirements, and the virtual–real fusion maintainability evaluation scheme is poor.

## 3. Demonstration

### 3.1. System construction

Taking a ship engine maintainability verification as a typical research case, the simulation environment of the real rear auxiliary engine room is shown in Fig. 5. The diesel engine is shown in Fig. 6, which is mainly composed of crank connecting rod mechanism, valve structure, fuel system, lubrication system, cooling system, and starting system.

The diesel engine needs to replace the fuel filter, air filter element, and other consumables, and the cylinder head needs to be opened to adjust the valve clearance. In addition, the starting motor has a certain failure rate; thus, it should be designed with good maintainability to ensure the rapid maintenance of the crew. Here, three maintenance tasks are selected to carry out the maintainability test: replacing the

air filter element, replacing the starting motor, and adjusting the valve clearance.
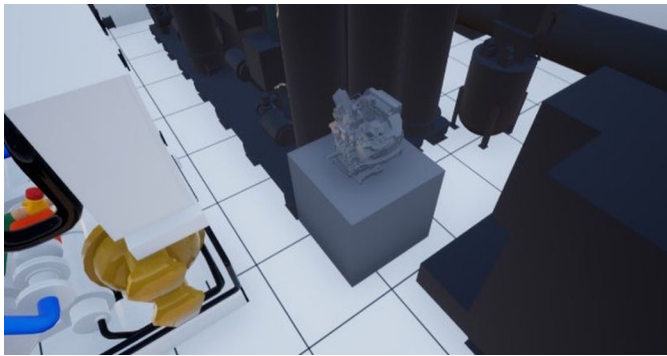

*Fig. 5. Simulation of real rear auxiliary engine room environment*


*Fig. 6. Diesel engine to be studied*

The main purpose of the example is to verify the maintainability verification method of virtual–real fusion. Three test methods are compared. The first is the real physical test. The real prototype is maintained and operated by real human, which represents the most accurate test conclusion. The second is the virtual–real fusion test verification, which uses the data obtained from the real operation of the physical prototype and virtual environment, and carries out the sequential verification of the data fusion method proposed above. The third method is to operate the virtual prototype by human. Although the tactile and force senses are not mature enough, this test can simulate the operation process to a certain extent.

Due to the high cost of carrying out live maintenance tests, and the test operation also has certain safety risks, the main equipment of the ship's auxiliary engine cabin is obtained in the laboratory by using the complete digital model of the ship's auxiliary engine compartment and 3D printing. The space layout is carried out according to the relative position relationship of each equipment in the real cabin, so as to simulate the real ship's cabin maintainability test scenario, as shown in Fig.7.

## 3.2. Maintainability verification of valve clearance adjustment task

### (1) Test operation and data acquisition


*Fig. 7. 3D printed cabin equipment*

Prior to adjusting the valve clearance, the cylinder head, which is connected with the engine body through three No. 10 fixing screws, is removed. Therefore, the maintenance tools include No. 10 hex wrench and slotted screwdriver. The operation steps of the maintenance process are shown in Fig.8.
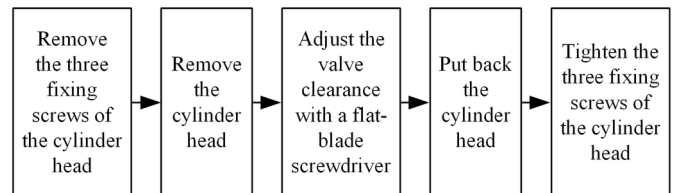

*Fig. 8. Operation steps of adjusting the valve clearance maintenance*

The operation method of the main process is shown in Fig.9.

The human and virtual human maintenance data can be obtained at the same time by carrying out the virtual–real fusion maintainability test, the process is shown in Fig.10. The virtual maintenance data can be obtained by correcting the errors of the virtual human data, con-
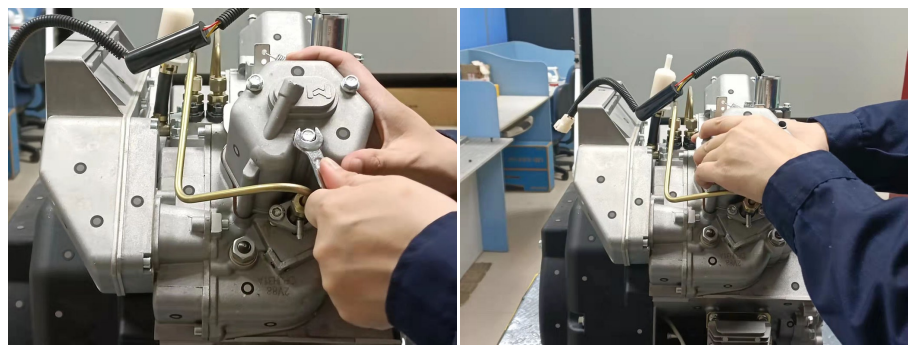

*a) Remove the screws*


*b) Remove the cylinder head*
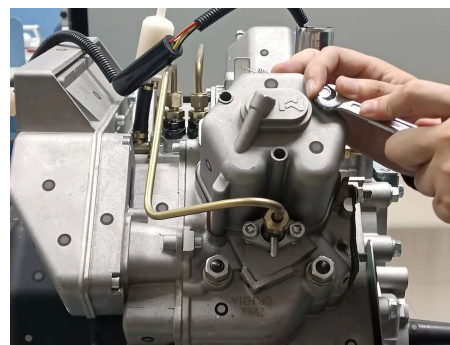

*c) Adjust the valve clearance*
*Fig. 9. Adjustment of valve clearance Service procedure*

sidering the various errors in the virtual–real fusion test. At the same time, the real maintainability test and virtual simulation test are carried out to compare with the virtual–real fusion test, as shown in Fig.11 and Fig 12.

Fig. 10. The virtual–real maintainability test



Fig. 11. The real maintainability test



Fig. 12. The virtual maintainability test

The virtual–real fusion maintainability test is conducted repeatedly for 10 times to reduce the randomness of the single maintenance test and ensure the reliability of the test results. At the same time, 10 real maintainability tests and virtual simulation tests are carried out to facilitate comparison and analysis with the virtual real fusion test. The maintainability test results are shown in Table 1.

**(2) Virtual and real data fusion**

Table 1.  Adjustment of valve clearance maintainability test results

| | Real mainte-nance time /s | Virtual–real fusion maintain-ability data | | Virtual data time /s |
|---|---|---|---|---|
| | | Human maintenance time /s | Virtual hu-man mainte-nance time /s | |
| 1 | 342 | 355 | 353 | 403 |
| 2 | 358 | 379 | 367 | 394 |
| 3 | 359 | 307 | 331 | 440 |
| 4 | 360 | 329 | 342 | 469 |
| 5 | 348 | 363 | 359 | 431 |
| 6 | 353 | 398 | 377 | 373 |
| 7 | 364 | 350 | 353 | 356 |
| 8 | 356 | 324 | 340 | 413 |
| 9 | 347 | 359 | 357 | 410 |
| 10 | 356 | 387 | 371 | 378 |

According to the adaptive weighted fusion algorithm in Section 2.2, the fusion test data are calculated according to the real maintenance data and the virtual maintenance data, and the fusion results are shown in Table 2.

Table 2. Fusion results of maintainability test data

| | $X_1(k)$ | $X_2(k)$ | $\overline{X}(k)$ | $\sigma_1^2(k)$ | $\sigma_2^2(k)$ | $\varphi_1(k)$ | $\varphi_2(k)$ | $X_r(k)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 355 | 353 | 354.0 | 1 | 1 | 0.5 | 0.5 | 354.0 |
| 2 | 379 | 367 | 363.5 | 120.6 | 6.6 | 0.052 | 0.948 | 367.6 |
| 3 | 307 | 331 | 348.7 | 660.0 | 108.8 | 0.142 | 0.858 | 327.6 |
| 4 | 329 | 342 | 345.4 | 562.2 | 84.5 | 0.131 | 0.869 | 340.3 |
| 5 | 363 | 359 | 348.5 | 491.8 | 89.7 | 0.154 | 0.846 | 359.6 |
| 6 | 398 | 377 | 355 | 718.0 | 155.4 | 0.178 | 0.822 | 380.7 |
| 7 | 350 | 353 | 354.5 | 618.3 | 133.5 | 0.178 | 0.822 | 352.5 |
| 8 | 324 | 340 | 351.7 | 636.9 | 133.9 | 0.174 | 0.826 | 337.4 |
| 9 | 359 | 357 | 352.4 | 571.0 | 121.4 | 0.175 | 0.825 | 357.4 |
| 10 | 387 | 371 | 355.1 | 615.7 | 134.5 | 0.179 | 0.821 | 373.9 |

The comparison chart is drawn according to the test data and the fused data, as shown in Fig. 13.
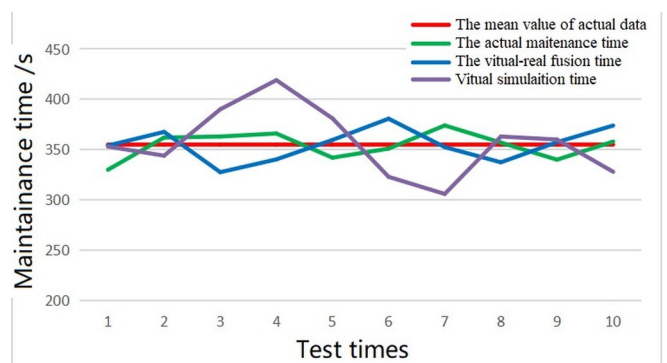


Fig.13. Maintenance data comparison

Fig. 13 shows that the virtual–real fusion data are closer to the actual maintenance data than the virtual simulation data, and the virtual–real fusion maintainability test can reflect the maintenance process more truly. Thus, the accuracy of the virtual real test data fusion method using the adaptive weighting algorithm is confirmed.

**(3) Time distribution check**

The real maintenance data in Table 1 are considered historical data, and the data $X_r(k)$ obtained by fusion in Table 3 are considered

| | $X_1(k)$ | $X_2(k)$ | $\overline{X}(k)$ | $\sigma_1^2(k)$ | $\sigma_2^2(k)$ | $\varphi_1(k)$ | $\varphi_2(k)$ | $X_R(k)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 355 | 353 | 354.0 | 1 | 1 | 0.5 | 0.5 | 354.0 |
| 2 | 379 | 367 | 363.5 | 120.6 | 6.6 | 0.052 | 0.948 | 367.6 |
| 3 | 307 | 331 | 348.7 | 660.0 | 108.8 | 0.142 | 0.858 | 327.6 |
| 4 | 329 | 342 | 345.4 | 562.2 | 84.5 | 0.131 | 0.869 | 340.3 |
| 5 | 363 | 359 | 348.5 | 491.8 | 89.7 | 0.154 | 0.846 | 359.6 |
| 6 | 398 | 377 | 355 | 718.0 | 155.4 | 0.178 | 0.822 | 380.7 |
| 7 | 350 | 353 | 354.5 | 618.3 | 133.5 | 0.178 | 0.822 | 352.5 |
| 8 | 324 | 340 | 351.7 | 636.9 | 133.9 | 0.174 | 0.826 | 337.4 |
| 9 | 359 | 357 | 352.4 | 571.0 | 121.4 | 0.175 | 0.825 | 357.4 |
| 10 | 387 | 371 | 355.1 | 615.7 | 134.5 | 0.179 | 0.821 | 373.9 |

field data to evaluate the virtual–real fusion maintenance scheme. Assuming that the historical maintenance time is $Y = (y_1,...,y_n)$, let $X = \ln Y = (x_1,...,x_n)$. Whether $X$ follows a normal distribution should be checked.

The following can be calculated: $\hat{\mu} = \overline{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = 5.869$, $\hat{\sigma}^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{X})^2 = 0.001396$. The sample values are arranged from small to large (repeated data are combined), and the frequency corresponding to each order statistic is $n_i$. Table 3.3 shows the calculation detail. The table indicates that $\hat{D}_n = 0.1783$, and the significance level $\alpha = 0.2$. The critical value table of Kolmogorov test indicates that $\hat{D}_{n\alpha} = 0.3226$. Thus, it obeys the normal distribution, that is, the maintenance time obeys the log-normal distribution. In addition, the field maintenance time obeys the log-normal distribution.

**(4) Consistency check**

The maintenance time has been shown to follow a log-normal distribution; thus, only a parametric test is required. For the convenience of research, the logarithm of historical data and field data is considered (denoted as $X_1, X_2$, respectively) and transformed into normal distribution for further research.

As mentioned, $n_1 = 10$, $n_2 = 10$, $\overline{X}_1 = 5.869$ $S_1^2 = 0.001396$ $\overline{X}_2 = 5.871$ $S_2^2 = 0.002457$, and $\alpha = 0.1$.

The variance $F' = 0.5682$ is tested, and because $F_{0.95}(9,9) < F' < F_{0.05}(9,9)$, no difference is found in their variances. The mean value $t' = -0.0967$ is also tested, and because $-t_{0.95}(18) \le t' \le t_{0.95}(18)$, no difference is found in their mean value. Therefore, the historical data and the field data passed the consistency test.

**(5) Maintainability verification**

In $X \sim N(\theta, \sigma^2)$, the variance is estimated from the field data as $\bar{\sigma}^2 = 0.002457$ and $\theta \sim \pi(\theta) = N(5.869, 0.001396)$. The index value of MTTR is 355 seconds, and then $\theta_0 = 5.871$. Let $\alpha = 0.2$, $\beta = 0.15$, and $\lambda = 1.4$, that is, $\theta_1 = 1.4\theta_0$.

The average repair time is verified according to the Bayes sequential probability ratio test method. The following can be calculated: $P_{H_0} = 0.5199$ and $P_{H_1} = 0.4801$. Then, we can obtain $\Lambda(X) = 0.1821$ and $O_n = 0.1682$. Then, $A = 0.4519$ and $B = 1.5905$. The fusion data satisfy the maintainability requirements because $O_n \le A$, and the feasibility of the virtual reality fusion maintainability evaluation scheme is good.

**(6) Results comparison**

The above are the virtual–real fusion maintainability verification results, and its correctness and superiority need to be compared with the real and virtual maintainability verification results. Both real and virtual test data are from Table 1. The maintainability verification adopts the method specified in the literature [20]. By calculation, the real test passes verification, while the virtual test data fails to pass the consistency inspection due to the scattered data, so it is impossible to judge the maintainability level of valve clearance adjustment. This shows that the result of virtual–real fusion is consistent with the real experiment and the virtual experiment can not get the correct conclusion.

### 3.3. Maintainability verification and comparison of other tasks

Using the same test process and method as above, the maintainability test is carried out for the replacement of air filter element and the replacement of starting motor maintenance tasks respectively, and the virtual–real fusion maintainability test is compared with the real test and virtual simulation test. The operation processes pictures are shown in Fig.14 and Fig.15.

Since only one verification conclusion can be obtained from 10 test repairs and corresponding data, 3 groups of tests are carried out for each maintenance task to verify the stability of the method, and each group of tests should compare the three test modes. The real maintainability test is the benchmark, and it is the correct data no matter whether the maintainability meets the requirements. If the virtual–real fusion maintainability verification and virtual maintainability verification are the same as the real test conclusion, then the judgment result is correct, otherwise, it is wrong. Table 4 lists the maintainability verification results of different maintenance tasks.

Table 4 shows that the accuracy rate of the virtual–real fusion test maintainability verification results is 88.9%, whereas that of the virtual simulation test maintainability verification results is only 33.3%. Evidently, the accuracy of the virtual–real fusion maintainability evaluation results is significantly higher than the latter and is closer to the real maintainability test evaluation results. Therefore, the virtual–real fusion maintainability evaluation for ship equipment has strong feasibility.

Further analysis shows the main reasons for the higher accuracy of the virtual–real fusion maintainability test verification are: (1) The real operation is carried out on the real object, and its force touch is consistent with the actual. (2) During the test, the maintenance environment is superimposed, which can reflect the influence of maintenance space on maintenance operation. (3) The fusion of virtual and real data are adopted to reduce data error further.

The main reasons for the low accuracy rate of virtual maintainability verification are: (1) Although it can simulate the maintenance environment and space, and has good test vision, it is difficult to establish touch and force sense, resulting in a large error compared with the actual maintenance operation. (2) Single test is difficult to reflect detailed operations accurately, such as screwing, wiping, etc. The test is subjective and unstable. (3) Just like the virtual-real fusion

*a) real test*



*b) virtual–real fusion test*



*c) virtual simulation test*

Fig. 14. Maintainability test of air filter replacement



*a) real test*



*b) virtual–real fusion test*



*c) virtual simulation test*

Fig. 15. Maintainability test of starting motor replacement

Table 4.  Maintainability verification results of different maintenance tasks

| Maintenance task | No. | Real test | virtual–real fusion test | virtual test |
|---|---|---|---|---|
| valve clearance adjustment | 1 | √ | √ | × |
| | 2 | √ | √ | √ |
| | 3 | √ | √ | × |
| the air filter replacement | 1 | √ | √ | √ |
| | 2 | √ | √ | × |
| | 3 | √ | √ | √ |
| starting motor replacement | 1 | × | √ | √ |
| | 2 | × | × | √ |
| | 3 | × | × | √ |
| accuracy rate | | 100% | 88.9% | 33.3% |

*Note: "√" indicates that the maintainability satisfies the requirements, and "×" indicates that the maintainability does not satisfy the requirements.*

maintainability verification, the complementary data cannot be fused, which leads to large errors in the verification data.

## 4. Conclusion

In this study, real and virtual maintenance data are fused by adaptive weighting algorithm, thereby reducing the influence of the errors in the virtual–real fusion maintainability test on the maintenance results. The experimental results show that the fused data are closer to the results of the real maintenance test. Then, the fusion data is evaluated and verified using the truncated spot method. The results show that the virtual–real fusion maintainability verification method has higher accuracy and stronger feasibility than the virtual simulation test.

## References

1. Bernard F, Zare M, Sagot J C, Paquin R. Using Digital and Physical Simulation to Focus on Human Factors and Ergonomics in Aviation Maintainability, Human Factors, 2019, 62(1): 37-54, https://doi.org/10.1177/0018720819861496.
2. Brown J, Kelly. Maintainability Verification for Cost-Effective Execution, Annual Reliability and Maintainability Symposium (RAMS),

2018:1-4, https://doi.org/10.1109/RAM.2018.8463020.

3.  Chen Z Y, Zhang X W, Ye L Y. Multi-sensor data weighted fusion method based on LMS algorithm. Computer engineering and application, 2014, 50(20): 86-90, https://doi.org/10.3778/j.issn.1002-8331.1401-0273.

4.  Dong B C, Song B W, Liang Q W, Mao Z Y. Research on Small Sample Maintainability Experimentation and Evaluation of Weapon System, Acta Armamentar II, 2011, 32(3): 327-330.

5.  Ge X Y, Zhou Q X, Liu Z Q. Assessment of Space Station On-Orbit Maintenance Task Complexity. Reliability Engineering & System Safety, 2019, 106661, https://doi.org/10.1016/j.ress.2019.106661.

6.  Ge Z X, Zhang Y, Yang Y M, Luo X. A New Maintainability Evaluation Method Based on Virtual–real Fusion Scene Construction, Scientific Programming, 2022: 6547225, https://doi.org/10.1155/2022/6547225.

7.  Ge Z X, Qi Z Q, Luo X, Yang Y M, Zhang Y. Multistage Bayesian fusion evaluation technique endorsing immersive virtual maintenance, Measurement, 2021,177: 109344, https://doi.org/10.1016/j.measurement.2021.109344.

8.  Goulden E C. An analytic approach to performing a maintainability demonstration. IEEE Transactions on Reliability, 1990, 39(1), 19-22, 25, https://doi.org/10.1109/24.52628.

9.  Grochow K, Martin S L, Hertzmann A. Style-based inverse kinematics. ACM Trans on Graphics, 2004, 23(3):522-531, https://doi.org/10.1145/1015706.1015755.

10. Guo Z Y, Zhou D, Chen J Y, Geng J. Using virtual reality to support the product's maintainability design: Immersive maintainability verification and evaluation system, Computers in Industry, 2018, 101:41-50, https://doi.org/10.1016/j.compind.2018.06.007.

11. Guo Z Y, Zhou D, Zhou Q, Meia S, Zeng S, Yu D, Chen J. A hybrid method for evaluation of maintainability towards a design process using virtual reality. Computers & Industrial Engineering, 2020, 140(1): 106227 , https://doi.org/10.1016/j.cie.2019.106227.

12. Guo Z Y, Zhou D, Zhou D D, Zhang X, Geng J, et al. Applications of virtual reality in maintenance during the industrial product lifecycle: A systematic review. Journal of Manufacturing Systems, 2020, 56, 525-538, https://doi.org/10.1016/j.jmsy.2020.07.007.

13. Hao H J, Wang M L, Xu M, et al. Adaptive weighted data fusion of Muti-sensor based on fuzzy preference relation, IEEE International Conference on Information and Automation, 2016: 195-199, https://doi.org/10.1109/ICInfA.2016.7831821.

14. Kline M B. Suitability of the lognormal distribution for corrective maintenance repair times. Reliability Engineering, 1984, 9(2): 65-80, https://doi.org/10.1016/0143-8174(84)90041-6.

15. Lu Z, Zhou J, Li N X. Maintainability fuzzy evaluation based on maintenance task virtual simulation for aircraft system. Maintenance and Reliability, 2015, 17 (4): 504–512, http://dx.doi.org/10.17531/ein.2015.4.4.

16. Lu Z, Liu J, Li D, Liang X H. Maintenance Process Simulation Based Maintainability Evaluation by Using Stochastic Colored Petri Net. Applied Sciences, 2019, 9(16), 3262, https://doi.org/10.3390/app9163262.

17. Luo X, Ge Z X, Zhang S G, Yang Y M.  A method for the maintainability evaluation at design stage using maintainability design attributes, Reliability Engineering & System Safety, 2021, 210: 107535, https://doi.org/10.1016/j.ress.2021.107535.

18. MA Z, Ben-Tzvi P. RML glove--an exoskeleton glove mechanism with haptics feedback. IEEE/ASME Transactions on Mechatronics, 2015, 20(2), 641–652, https://doi.org/10.1109/tmech.2014.2305842.

19. Miao Q, Liu L, Yuan F, Michael P. Complex system maintainability verification with limited samples, Microelectronics Reliability, 2011, 51(2): 294-299, https://doi.org/10.1016/j.microrel.2010.09.012.

20. MIL-STD-471A. Maintainability verification/demonstration/evaluation; 1973.

21. Osafo-Yeboah B, Jiang S, Delpish R, Jiang Z, Ntuen C. Empirical study to investigate the range of force feedback necessary for best operator performance in a haptic controlled excavator interface. International Journal of Industrial Ergonomics, 2013, 43(3), 197–202, https://doi.org/10.1016/j.ergon.2013.02.005.

22. Overtoom E M, Horeman T, Schreuder H W R. Haptic Feedback, Force Feedback, and Force-Sensing in Simulation Training for Laparoscopy: A Systematic Overview. Journal Of Surgical Education, 2019, 76 (1) , 242-261, https://doi.org/10.1016/j.jsurg.2018.06.008.

23. Pedro M D L, Vicente G P, Luis B M, Adolfo C M. A practical method for the maintainability assessment in industrial devices using indicators and specific attributes. Reliability Engineering & System Safety, 2012, 100, 84–92, https://doi.org/10.1016/j.ress.2011.12.018.

24. Peng G L, Yu H, Liu X H, Jiang Y, Xu H. A desktop virtual reality-based integrated system for complex product maintainability design and verification, Assembly Automation, 2010, 30(4): 333-344(12), https://doi.org/10.1108/01445151011075799.

25. Retterer B L, Kowalski R A. Maintainability: A historical perspective, IEEE Transactions on Reliability, 1984, R-33(1): 56-61, https://doi.org/10.1109/TR.1984.6448275.

26. Sagardia M, Hertkorn K, Hulin T, Schätzle S, et al., VR-OOS: The DLR's virtual reality simulator for telerobotic on-orbit servicing with haptic feedback, IEEE Aerospace Conference, 2015: 1-17, https://doi.org/10.1109/AERO.2015.7119040.

27. Shao B C. Research on mobile robot training and control technology based on force feedback and virtual reality, Southeast University, 2021, https://doi.org/10.27014/d.cnki.gdnau.2021.001716.

28. Seemann W, Stelzner G, Simonidis C. Correction of motion capture data with respect to kinematic data consistency for inverse dynamic analysis. ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, 2005:187-194, https://doi.org/10.1115/DETC2005-84964.

29. Tu M X, Lv C, Wang M H, Zhou D, Xu Y L, Wan B L, He W X. Maintainability analysis and evaluation of flexible cables based on DELMIA. Transactions of the Canadian Society for Mechanical Engineering. 40(5): 995-1005, https://doi.org/10.1139/tcsme-2016-0082.

30. Wang X, Di P. Testability Evaluation Method of Equipment Based on Data Fusion for Virtual and Real Test Data Fusion. Ship Electronic Engineering, 2021, 41(06):131-134.

31. Wu Z Y, Hao J P. A Maintenance Task Similarity-Based Prior Elicitation Method for Bayesian Maintainability Demonstration. Mathematical Problems in Engineering, 2020, 1–19, https://doi.org/10.1155/2020/2730691.

32. Yang X, Su W, Deng J, Jin X, Tan G, Pan Z. Real-virtual fusion model for traffic animation. Computer Animation and Virtual Worlds, 2016, 28(6), e1740, https://doi.org/10.1002/cav.1740.

33. Zhang L, Brunnett G, Rusdorf S. Real-time human motion capture with simple marker sets and monocular video. Journal of Virtual Reality and Broadcasting, 2011, 8(1), https://doi.org/10.20385/1860-2037/8.2011.1.

34. Zhu W Y, Zhou S Y. 3D Reconstruction Method of Virtual and Real Fusion Based on Machine Learning, Mathematical Problems in Engineering, 2022, 2022, 1-11, https://doi.org/10.1155/2022/7158504.

# The influence of pressure drop on the working volume of a hydraulic motor

Indexed by:
Web of Science Group

## Pawel Śliwiński [a]

[a] Gdansk University of Technology, Faculty of Mechanical Engineering and Ship Technology, ul. Narutowicza 11/12, 80-233 Gdansk, Poland

## Highlights

- New method of determination ot the working volume of a hydraulic motor has been proposed.

- The actual working volume is a function of the pressure drop in motor working chambers.

- The actual working volume is bigger than the theoretical working volume.

- The actual working volume should be taken to evaluate losses in a hydraulic motor.

## Abstract

Reliability and maintenance analysis of hydraulic positive machines basicly focused on the processes of their wear and failure. But in order to correctly assess the mechanical and volumetric efficiency of a hydraulic motor, both at the stage of development research or at the stage of control tests during its exploitation, the working volume of this motor must be correctly determined. Therefore this paper proposes a new method of assessment of the size of the working volume of a hydraulic motor. It has been shown that the hydraulic motor absorbency per one revolution of this motor shaft is a non-linear function of pressure drop in working mechanism of the motor and non-linear function of motor rotational speed. Thus the relation between the working volume of a hydraulic motor and the pressure drop in the motor working mechanism is a non-linear function. This working volume as a funiction of pressure drop has been called the actual working volume. The correctness of the proposed method was confirmed experimentally.

## Keywords

theoretical working volume, actual working volume, hydraulic motor, satellite motor, volumetric efficiency, mechanical efficiency.

## 1. Introduction

Reliability and maintenance analysis of hydraulic drives is mainly oriented to hydraulic positive displacement machines (pump, motor, cylinder) [4, 5, 8, 17, 26, 27, 42]. This machines are the core of modern hydraulic system [1, 9, 13, 14, 19, 29]. In overall performance of the hydraulic system the behavior of hydraulic motor has an important meaning [11, 12]. The basic characteristics of the hydraulic motor and pump, such as characteristics of volumetric losses, mechanical losses and pressure losses are important to assess its volumetric efficiency and mechanical-pressure efficiency [28, 31, 53, 54]. Thus it is important for designers of a hydraulic system [52]. The knowledge of the above mentioned characteristics is also very important from the point of view of exploitation not only of the hydraulic motor itself, but also the entire hydraulic system [16, 46]. So far, the correct assessment of the losses in hydraulic motor, and thus partial efficiencies, depends on theoretical working volume $q_t$. The value of theoretical working volume $q_t$ is taken constant in the whole range of the pressure drop $\Delta p$ in the motor. Furthermore the theoretical working volume $q_t$ is independent on the motor speed n [2, 3]. Thus, improper adoption of the theoretical working volume $q_t$ may, during the exploitation of

the hydraulic system at different pressure drop in the motor inflow port, results in a different than required shaft rotational speed n than expected [39, 40, 41].

So far, in industrial practice, but also in laboratories, the theoretical working volume $q_t$ is determined in a very simplified way. The liquid compressibility is omitted. Similarly, the pressure drop $\Delta p_{ich}$ in the internal channels of the motor is also treated as insignificant and is negligible. It means that the pressure drop $\Delta p$ in a motor is the same as the pressure drop $\Delta p_i$ in the working chambers of this motor [2]. This is, of course, a considerable simplification.

The theoretical working volume $q_t$ is not the same as the so-called geometric working volume $q_g$. The geometric working volume $q_g$ results from the geometrical dimensions of the working chambers and is determined by constructors of a hydraulic motor, pump or other positive displacement devices like a rotational hydraulic dumper [43, 56, 57]. The formulas describing the geometric working volume $q_g$ are not the same for all types of positive displacement machines and very often these formulas are derived with some simplifications. In this way, the simplifications result in error of up to 3% for gear pumps [3, 43]. Furthermore, in the real (manufactured) motor, the geometric

E-mail addresses: P. Śliwiński (ORCID: 0000-0003-1332-3661): pawel.sliwinski@pg.edu.pl

volume $q_g$ differs from that determined based on the design documentation. The cause are loosenesses between working mechanism elements, machining errors, assembly errors, etc [2, 3].

As a result of the motor load by torque M in the motor working chambers is generated pressure differential $\Delta p_i$. This causes elastic deformation of the engine working chambers. Thus, the working volume of the loaded motor is bigger than motor without load [36, 37]. Similar phenomena observed Osiecki in a prototype of axial piston pump [25]. In this way the working volume depends on the pressure drop $\Delta p_i$ in the working chambers of the motor and was called the actual working volume $q_r$ [36, 37].

If $\Delta p_i = 0$ then $q_r = q_t$, where $q_t$ is the theoretical working volume [36, 37]. Therefore, both the geometric volume $q_g$ and the theoretical working volume $q_t$ are different from the actual working volume $q_r$. Booth $q_t$ and $q_g$ are should not be used to determine losses in motor and thus to calculate partial efficiencies of this motor [36, 37].

The first works on the methodology of determining the theoretical working volume appeared in the 1940s and 1960s. These were the work of Wilson (1949) [51], Schlosser and Hilbrands (1963) [32, 33]. Then the works of Toet and Balawender in 1970s and again in 2019 [2, 3, 44, 45]. The Toeth method was also described by Post in [30]. The methodology proposed by Toet and Balawender is based on characteristics of the flow rate Q to motor in the function the rotational speed n at constant pressure drops $\Delta p$ in the motor. In work [36] Sliwinski presented a new look on the Balawender method. Moreover, in [37] Sliwinski proposed another new method for determining the working volume. In Sliwinski's method the characteristics of effective absorbency of the motor were used.

The adoption of the theoretical working volume $q_t$ for assessment of the losses in displacement machines introduces a significant error [36, 37]. Furthermore in [36, 37] has been demonstrated that:
– the liquid compressibility has the influence on the theoretical working volume $q_t$ and the actual working volume $q_r$;
– to assess volumetric and mechanical losses in a hydraulic motor should be use the actual working volume $q_r$.

The simplest method of assessment the theoretical working volume has been shown in the ISO standard [7, 10]. But this method gives an inaccurate result and in scientific considerations is rather not used. According to Kim, the theoretical working volume can be determined by analysis a flow rate only in a single chamber of the working mechanism [15]. Similar method is proposed by Manring and Williamson [21]. These methods have no practical application in real tests of hydraulic motors and also pumps. The other method, called Latin Hyperspace Sampling (LHS), is proposed by Michael and Garcia-Bravo [22].

Balawender proved that based on the characteristics of the hydraulic motor absorbency $q_e$ per one revolution of the motor shaft vs pressure drop $\Delta p$ in this motor at n=const is possible to determine the theoretical working volume $q_t$ [2, 3]. This method will be briefly presented later in this article. Based on Balawender method, the proposed new method of determination the actual working volume $q_r$ and the theoretical working volume $q_t$ of a hydraulic motor is described in this article. The correctness of this method was experimentally verified.

Hydraulic motors with a satellite mechanism are the object of interest of many researchers and set a new trend in research in the field of hydraulic positive displacement machines. This is evidenced by the works [47, 48, 49, 50, 56]. These motors are also the main object of the author's scientific interests. Therefore a satellite hydraulic motor was selected for the experimental test. Its construction is presented in Section 5.1.

## 2. Flow rate in hydraulic motor

The basis for determining the theoretical working volume $q_t$ and the actual working volume $q_r$ of a hydraulic motor is the precise measurement of the flow rate into or out of the motor and the measurement of external leakage (if any). Generally, is recommended to use two flow meters – in the inflow line and in the outflow line of the tested motor (Fig. 1) [2,3]. But Toeth used only one flow meter installed in the motor inlet line [44, 45].
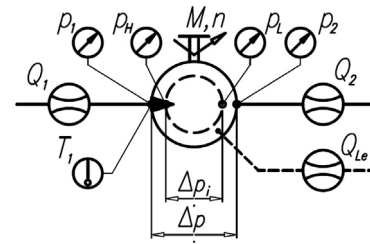


*Fig. 1. Parameters measured in a hydraulic motor [37]: $Q_1$ – flow rate to the motor, $Q_2$ – flow rate from the motor, $Q_{Le}$ – external leakage; $p_1$ – pressure in the inflow port, $p_2$ – pressure in the outflow port, $p_H$ – pressure in the high-pressure working chamber, $p_L$ – pressure in the low-pressure working chamber, M – torque, n – rotational speed, $T_1$ – liquid temperature in inflow port, $\Delta p$ – pressure drop in the motor, $\Delta p_i$ – pressure drop in the motor working chambers.*

To correctly determine the theoretical and actual working volume, it is necessary to analyze the flow balance in the hydraulic motor. The general model of the flow rate in a hydraulic motor is proposed by Balawender and is described by [2, 3]:

$$Q_1 = \underbrace{Q_g + Q_u + Q_k + Q_C + Q_{Li}}_{Q_2} + Q_{Le} \qquad (1)$$

where [2, 3]:
  $Q_g$ – the flow rate dependent on the geometric working volume $q_g$ and rotational speed n,
  $Q_u$ – the flow rate lifted in the gaps,
  $Q_k$ – the flow rate dependent on the working chambers cyclic elastic deformation,
  $Q_C$ – the flow rate caused by liquid compressibility,
  $Q_{Li}$ – the flow rate from high-pressure working chambers to low-pressure working chambers (internal leakage),
  $Q_{Le}$ – the external leakage.

Another model is proposed by Sliwinski. This model is described by general formula [37, 39, 40]:

$$Q_1 = \underbrace{Q_t + \Delta Q_L + Q_{Lfg} + Q_{CU} + Q_{Le}}_{Q_L} \qquad (2)$$

where:
  $Q_t$ – the theoretical flow rate in the motor:

$$Q_t = q_t \cdot n \qquad (3)$$

  $\Delta Q_L$ – the component of volumetric losses depends on liquid compressibility and rotational speed of the motor [39, 40]:

$$\Delta Q_L = \underbrace{\left( C_q \cdot \Delta p_i + \frac{C_{id}}{n^{0,5}} \right) \cdot m^2 \cdot H \cdot n}_{\Delta q_L} \qquad (4)$$

  m – the teeth module in the satellite mechanism,
  H – the height of the satellite mechanism,
  $C_q$, $C_{id}$ – coefficients,
  $\Delta p_i$ – the pressure drop in working mechanism of the motor, defined as [35, 38]:

$$\Delta p_i = \Delta p - \Delta p_{ich} \qquad (5)$$

$\Delta p_{ich}$ – the pressure drop in the motor internal channels [35, 38]:

$$\Delta p_{ich} = C_t \cdot \rho \cdot Q_2{}^2 + C_l \cdot \nu \cdot \rho \cdot Q_2 \qquad (6)$$

$C_l, C_t$ – the constant of the laminar and turbulent flow component respectively,
$\nu$ – the kinematic viscosity,
$\rho$ – the density of liquid,
$Q_{Lfg}$ – the flow rate in flat clearances of working mechanism,
$Q_{CU}$ – the flow rate in commutation unit clearances.

The external leakage $Q_{Le}$ has been defined as [37, 39, 40]:

$$Q_{Le} = Q_{Le1} + Q_{Le2} \qquad (7)$$

$Q_{Le1}$ – the leakage from the high-pressure working chambers (with the pressure $p_H$) to the outside of the motor (with the pressure equal zero),
$Q_{Le2}$ – the leakage from the low-pressure working chambers (with the pressure $p_L$) to the outside of the motor (with the pressure equal zero).

The methodology of determination of $\Delta p_i$ and $\Delta p_{ich}$ is widely described in the works [35, 36, 38, 39]. Similarly, each component of the above mathematical model has been described in detail in publications [35, 38, 39] and will not be discussed in more detail here.

## 3. Balawender's method of determining the theoretical working volume

The method of determining the theoretical working volume based on characteristic $q_e = f(\Delta p)_{n=const.}$ was developed by Balawender in 1974 [3]. $q_e$ is called the effective absorbency of the motor per one revolution of its shaft and is defined as:

$$q_{e1} = \frac{Q_1}{n} \qquad (8)$$

$$q_{e2} = \frac{Q_2}{n} \qquad (9)$$

that is:

$$q_{e1} = \underbrace{q_g + q_u + q_k + q_{C1}}_{q_{w1}} + \underbrace{\frac{Q_{Le1}}{n}}_{q_{L1}} \qquad (10)$$

$$q_{e2} = \underbrace{q_g + q_u + q_k + q_{C2}}_{q_{w2}} + \underbrace{\frac{Q_{Le2}}{n}}_{q_{L2}} \qquad (11)$$

where $q_g = Q_g/n$, $q_u = Q_u/n$ etc.

If the $\Delta p_i$ decreases to zero then for a given constant motor speed n [3]:

$$\lim_{\Delta p_i \to 0} (q_{e1})_{(n)} = q_{t1(n)} + \frac{1}{n} \cdot \lim_{\Delta p_i \to 0} Q_{Le1} \qquad (12)$$

$$\lim_{\Delta p_i \to 0} (q_{e2})_{(n)} = q_{t2(n)} - \frac{1}{n} \cdot \lim_{\Delta p_i \to 0} Q_{Le2} \qquad (13)$$

and:

$$\lim_{\Delta p_i \to 0} Q_{Le1} = \lim_{\Delta p_i \to 0} Q_{Le2} = \frac{1}{2} \cdot \lim_{\Delta p_i \to 0} Q_{Le} = 0 \qquad (14)$$

Then from the equations (12) and (13) is:

$$q_{t1(n)} = \lim_{\Delta p_i \to 0} (q_{e1})_{(n)} - \frac{1}{2 \cdot n} \cdot \lim_{\Delta p_i \to 0} Q_{Le} \qquad (15)$$

$$q_{t2(n)} = \lim_{\Delta p_i \to 0} (q_{e2})_{(n)} + \frac{1}{2 \cdot n} \cdot \lim_{\Delta p_i \to 0} Q_{Le} \qquad (16)$$

But the theoretical forking volume for a given constant motor speed n is:

$$q_{t(n)} = \frac{1}{2} \cdot \left( q_{t1(n)} + q_{t2(n)} \right) \qquad (17)$$

According to Balawender the test of motor can be carried out at limited amount of speed n, for example at minimum speed, nominal speed and maximum speed. Then the final theoretical working volume $q_t$ shoult be taken as:

$$q_t = \frac{1}{z} \cdot \left( q_{t(n)1} + q_{t(n)2} + \dots + q_{t(n)z} \right) \qquad (18)$$

Furthermore, Balawender claims that his metod gives a satisfactory result of $q_t$ only at one motor speed n. Then $q_t = q_{t(n)}$.

The graphical implementation of Balawender's method is shown
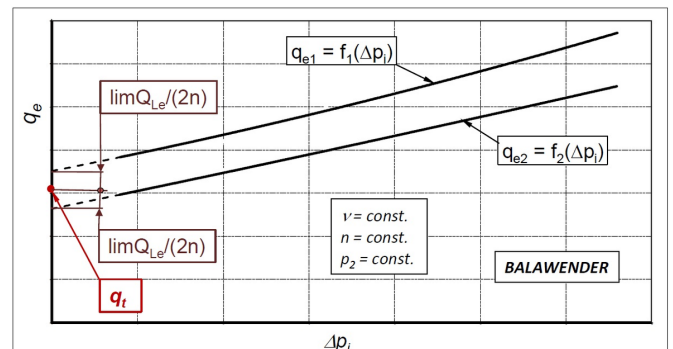


in Fig. 2

*Fig. 2. The graphical implementation of Balawender's method [3]*

Balawender recommends that [2, 3]:
a) the characteristics $q_e=f(\Delta p_i)$ obtained from the experimental data (Fig. 2) should be described by the formula:

$$q_{e1} = q_{e1(\Delta pi=0)} \pm \frac{a_1}{n} \cdot (\Delta p_i)^{b_1} \qquad (19)$$

$$q_{e2} = q_{e2(\Delta pi=0)} \pm \frac{a_2}{n} \cdot (\Delta p_i)^{b_2} \qquad (20)$$

b) the characteristic of external leakage $Q_{Le}$ should be described by the formula:

$$Q_{Le} = Q_{Le(\Delta pi=0)} + a_Q \cdot (\Delta p_i)^{b_Q} \qquad (21)$$

where the values of the constants $a_1$, $a_2$, $a_Q$, $b_1$, $b_2$ and $b_Q$ should be determined by the least squares method.

## 4. Proposed new method of determining the actual and theoretical working volume

The bigges disadvantage of the Balawender method is the assumption of the theoretical working volume $q_t$ as the average of the working volumens calculated for various constant rotational speeds. In addition, Balawender argues, that the theoretical working volume $q_t$ can be determined even with the one constant rotational speed n [2, 3]. Thus, Balawender does not specify the effect of the motor speed n on the theoretical working volume $q_t$. Sliwinski showed that the volumetric losses in the hydraulic motor are nonlinear function of speed n (see formulas (2) and (4)).

Another disadvantage of the Balawender method is, recommended by him, the method of describing the effective absorbency $q_e$ (equations (19) and (20)). These formulas will only describe the absorbency well if the leakage in the motor will increase non-linearly. A non-linear increase in leakage occurs when, as a result of pressure increase, there is an increase in the clearances in the working mechanism. Such a phenomenon occurs in motor that is not equipped with a clearance compensation unit in the working mechanism [40].

### 4.1. Influence of pressure on the working volume

Balawender and Toeth adopted the simplification that the increase in working volume of a motor is a linear function of pressure drop in the motor [2, 3, 44, 45]. In fact, each dimension of the engine working chamber changes linearly under the pressure. Sliwinski in [36] showed that the working volume changes nonlinearly in the function of the pressure drop $\Delta p_i$. Therefore, in loaded motor the theoretical working wolume $q_t$ is not equal the actual working wolume $q_r$. The actual working volume $q_r$ is described by the following formula [36, 37]:

$$q_r = q_t + \underbrace{\left(C_{q1} \cdot \Delta p_i + C_{q2} \cdot \Delta p_i^2 + C_{q3} \cdot \Delta p_i^3\right) m^2 \cdot H}_{\Delta q_p} \quad (22)$$

where $\Delta q_p$ is the change in working volume of the loaded motor. So, the increase in working volume $\Delta q_p$ can not be the reason of the mechanical and volumetric losses in the motor. Therefore the relationship (4) should take the form:

$$\Delta Q_L = \underbrace{\left(\frac{C_{id}}{n^{0.5}}\right) \cdot m^2 \cdot H \cdot n}_{\Delta q_L} \quad (23)$$

where $\Delta q_L$ is the change in unit volumetric losses (that is volumetric losses per one revolution of the motor shaft). It can be assumed that the influence of pressure drop $\Delta p_i$ in motor working mechanism on the $\Delta q_L$ is very small and for father consideration can be neglected. The $\Delta Q_L$ is mainly caused by the movement of liquid in the spaces between the mating teeth during rotation of the working mechanism with speed n [39, 40].

### 4.2. Flow rate in a hydraulic motor

Taking into account the formula (22) the flow rate in a motor can be written with the following equation:

$$Q_1 = \underbrace{q_r \cdot n}_{Q_r} + \underbrace{\Delta q_L \cdot n + Q_C + Q_{Lfg} + Q_{CU} + Q_{Le}}_{Q_L} = Q_2 + Q_{Le} \quad (24)$$

Immediately from the measured datas the flow rate per one revolution of the motor shaft $q_e$ should be calculated according to formulas (8) and (9):

$$q_{e1} = q_r + \cdot q_L + \frac{Q_C + Q_{Lfg} + Q_{CU} + Q_{Le}}{n} \quad (25)$$

$$q_{e2} = q_r + \cdot q_L + \frac{Q_C + Q_{Lfg} + Q_{CU}}{n} \quad (26)$$

where:
- $Q_C$ – the flow rate component depends on the liquid compressibility and can be described as a sum:

$$Q_C = Q_{C1} + Q_{C2} \quad (27)$$

- $Q_{C1}$ – the flow rate component depends on the pressure difference $p_1 - p_H$ ($p_1 > p_H$) and describes as [36, 37]:

$$Q_{C1} = Q_1 \cdot \int_{p_H}^{p_1} \frac{1}{K_{Z(p)}} dp \quad (28)$$

- $Q_{C2}$ – the flow rate component depends on the pressure difference $p_H - p_2$ ($p_H \gg p_2$) and describes as [36, 37]:

$$Q_{C2} = Q_2 \cdot \int_{p_2}^{p_H} \frac{1}{K_{Z(p)}} dp \quad (29)$$

- $K_{Z(p)}$ – the tangential isentropic bulk modulus (Fig. 7) [36, 54, 55].

Because the pressure drop $\Delta p_i$ in the working mechanism influences on the volumetric losses $Q_L$ then the $Q_1$ and $Q_2$ should be related to pressure $p_H$ in high pressure working chambers. Then:

$$Q_{1(p_H)} = Q_1 + Q_{C1} \quad (30)$$

$$Q_{2(p_H)} = Q_2 - Q_{C2} \quad (31)$$

The flow rate $q_{e(pH)}$ related to the pressure $p_H$ should be calculated as:

$$q_{e1(pH)} = \frac{Q_{1(p_H)}}{n} \quad (32)$$

$$q_{e2(pH)} = \frac{Q_{2(p_H)}}{n} \quad (33)$$

If a flow meter is located in the motor outflow line the component $Q_{Le}$ is omitted. It should be noted that the $q_{e(pH)}$ is a nonlinear function of pressure drop $\Delta p_i$ in the working mechanism and is directly proportional to the inverse of the motor shaft speed n.

### 4.3. Theoretical working volume

If the pressure drop $\Delta p_i$ in the motor decreases then the leakages $Q_L$ in motor decreases also. Therefore, the value of $q_{e(pH)}$, described by formulas (25) and (26), tends to certain value $q_n$:

$$\lim_{\Delta p_i \to 0} \left(q_{e1(pH)}\right)_{(n)} = q_{t1} + \Delta q_L = q_{n1} \quad (34)$$

$$\lim_{\Delta p_i \to 0} \left(q_{e2(pH)}\right)_{(n)} = q_{t2} + \Delta q_L = q_{n2} \quad (35)$$

The share of leakages $Q_L$ in general flow rate in the motor decreases if rotational speed n increases. Thus, the theoretical working volume can be axpressed as:

$$\lim_{n \to \infty}(q_{n1}) = q_{t1} \qquad (36)$$

$$\lim_{n \to \infty}(q_{n2}) = q_{t2} \qquad (37)$$

The graphical interpretation of $q_n$ and theoretical working volume $q_t$ is shown in Fig. 3.
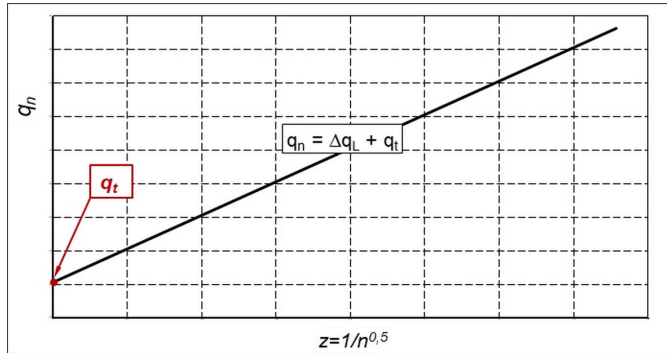


*Fig. 3. The graphical implementation of the new method.*

As a final theoretical working volume is proposed:

$$q_t = 0.5 \cdot (q_{t1} + q_{t2}) \qquad (38)$$

In general, it is advisable to use flow meters of the same type and class to measure flow rates $Q_1$ and $Q_2$. Furthermoref the compressibility of liquid is considered, then:

$$q_{t1} = q_{t2} = q_t \qquad (39)$$

In practice, only one flowmeter is used in the test stand measuring system. In such a case $q_t = q_{t1}$ or $q_t = q_{t2}$.

## 4.4. Assesment of the determined value of actual and theoretical working volume

The correctness of the determination of the value of the actual and theoretical workin volume can be assessed using the definitions of mechanical efficiency $\eta_m$ and volumetric efficiency $\eta_v$. For hydraulic motor [36, 37]:

$$\eta_m = 2 \cdot \frac{M}{q_r \cdot \Delta p_i} \qquad (40)$$

$$\eta_v = \frac{q_r \cdot n}{Q_1} \qquad (41)$$

Threrefore for any rotational speed n of the motor and for $\Delta p_i \to 0$ the volumetric efficiency should be $\eta_v < 1$. Furthermore for $\Delta p_i = 0$ is and $q_r = q_t$. If rotational speed of the tested motor is minimum ($n_{min}$), can be assumed that for $\Delta p_i = 0$ is $\eta_v \approx 1$. Therefore:

$$q_{t(max)} \leq \frac{Q_1}{n_{min}} \qquad (42)$$

For assessment of the minimum theoretical working volume $q_{t(min)}$ and actual working volume $q_{r(min)}$ the maximum mechanical efficiency $\eta_m$ of hydraulic motor should be observed in all range of operating parameters (that is speed n and load M). If $\eta_m \leq 1$, then:

$$q_{t(min)} < q_{r(min)} \leq 2\pi \cdot \frac{M}{\Delta p_i} \qquad (43)$$

To sum up:

$$\frac{Q_1}{n_{min}} \geq q_t < 2\pi \frac{M}{\Delta p_i} \qquad (44)$$

## 4.5. General test procedure

In order to obtain experimental data and determine the $q_t$ and $q_r$ the flow $Q_1$ or $Q_2$ (at $T_1$ = cosnt) should be measured firstly for several values of n and for several constant values of $\Delta p$. Then should be calculate the $\Delta p_{ich}$ according to the method described in [35, 38]. After determining the $\Delta p_{ich}$, it is possible to calculate the $\Delta p_i$. If the flow meter is located in the low-pressure line the $Q_{(pH)}$ should be calculated. Next, the characteristics $q_e = f(_{\Delta pi})_{n=const}$ should be developed and calculate the $q_{r1}$ (or $q_{r2}$). Finally, the characteristics $q_r = f(n)$ should be plotted and the walue of $q_{t1}$ (or $q_{t2}$) should be calculated.

## 5. Results of experiment

### 5.1. Object of research

The object of research was a prototype of a satellite hydraulic motor (Fig. 4 and Fig. 5). The principle of operation of the satellite engine is widely known and is described, inter alia, in [23, 34, 36, 37, 39, 41, 54].
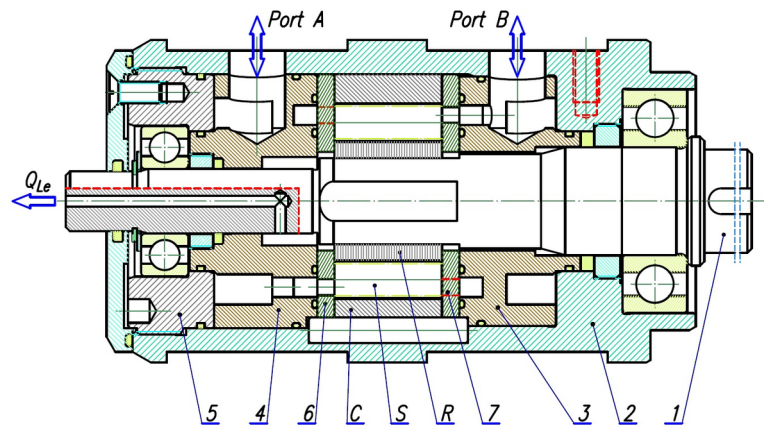


*Fig. 4. Tested motor (satellite motor) [36, 37, 39]: C – curvature, S – satellite, R – rotor, 1 – shaft, 2 – case, 3 – inflow/outflow manifold, 4 – outflow/inflow manifold, 5 – rear body, 6 and 7 – distribution (compensation) plates*

The geometrical working volume $q_g$ of the tested motor is 34.44 cm³/rev. This working volume depends on the height H of the working mechanism, areas of the minimum and maximum working chamber ($A_{min}$ and $A_{max}$ – Fig. 5) and the numbers of the humps of the rotor and the curvature ($n_R$ and $n_C$ respectively). The geometrical working volume $q_g$ was calculated according to [36, 37, 39]:

$$q_g = n_C \cdot n_R \cdot H \cdot (A_{max} - A_{min}) \qquad (45)$$
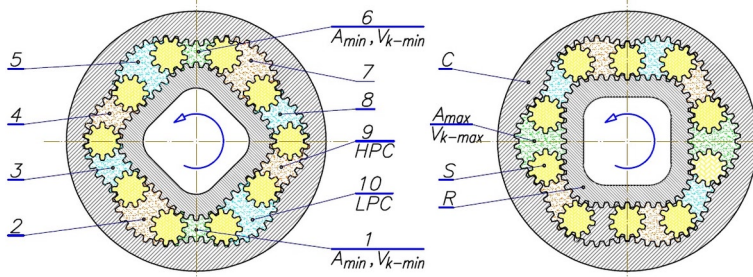
Fig. 5. Working (satellite) mechanism [23, 34, 36, 37, 41, 54]: C – curvature, R – rotor, S – satellite, 1÷10 – working chambers, HPC – high pressure chambers, LPC – low pressure chambers, $V_{k-min}$ – dead chamber, $V_{k-max}$ – working chamber with maximum volume and with maximum area $A_{max}$

where $n_C = 6$, $n_R = 4$, H = 25 mm, $A_{min} = 26.11$ mm$^2$ and $A_{max} = 83.51$ mm$^2$.

## 5.2. The measuring system

The diagram of hydraulic and the measuring system of the test stand is shown in Fig. 6.
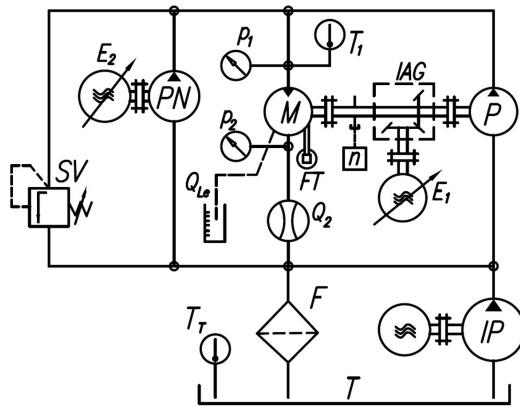


Fig. 6. The test stand measuring system [36, 37, 41]: P – pump, M – tested motor, PN – pump for filling leaks in P and M, IP – impeller pump, SV – safety valve, F – filter, T – tank, IAG – intersecting axis gear, $E_1$ and $E_2$ – electric motors with frequency converters, $T_1$, $T_T$ – temperature sensors, $Q_2$ – flowmeter, $Q_{Le}$ – leakage measurement, FT – force transducer for torque measurement, n – inductive sensor for measure of rotational speed.

The following measuring instruments were installed in the measuring system of the test stand (Fig. 6):
- two strain gauge pressure transducers ($p_1$ and $p_2$) with ranges of 2.5, 10, 40 MPa and class 0.3;
- piston flowmeter ($Q_2$) with volume of measured chamber 0.63dm$^3$, a range of 200 l/min and class 0.2;
- strain gauge force transducer FT for measurement of the torque M with a range of 100 N and class 0.1;
- inductive sensor for measurement of the rotational speed n of the motor shaft with accuracy of ±0.01 rpm);
- RTD temperature sensor with class A and max. measurement error 0.5 °C – for measurement of the liquid temperature $T_1$ in the inflow port of the motor.

Pressure drop $\Delta p$, speed n and oil temperature $T_1$ settings were made very precisely, with as few deviations as possible. Thus, for speed ±0.1 rpm, for pressure drop ±0.05 MPa and for temperature ±1.0 °C. Each recorded measurement result is the average of three repetitions. This regime of parameter settings enabled to determine the working volume with the lowest possible error.

A valuable advantage of the piston flow meter should also be highlighted here. Well, it measured the average flow rate from more than 18 revolutions of the motor shaft (due to the large volume of the flow meter's measuring chamber of as much as 0. 63dm$^3$)!

## 5.3. Working liquid parameters

The working liquid in test stand was the Total Azolla ZS 46 oil. The research was conducted with the temperature in the motor inflow port $T_1 = 43$ °C ($\nu = 40$ cSt, $\rho = 873$ kg/m$^3$). According to the proposed new method, the characteristic of tangential isentropic bulk modulus $K_{Z(p)}$ of mineral oil is needed to correct determine the theoretical and actual working volume. This characteristic was shown in Fig. 7 [55].

## 5.4. Characteristics of flow rate in the outflow line of the motor
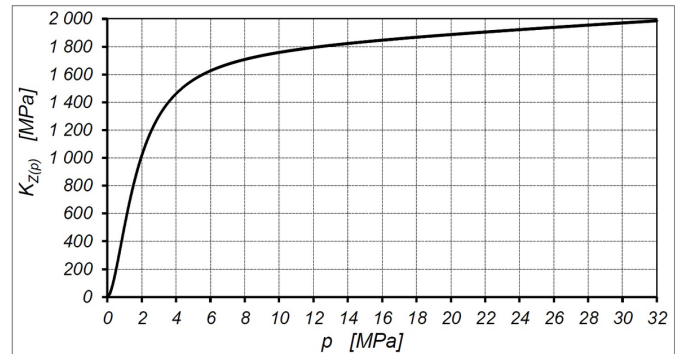


Fig. 7. Bulk modulus $K_{Z(p)}$ of Total Azolla ZS 46 mineral oil [36, 37, 55]

The motor tests were carried out for in the pressure drop range $\Delta p$ up to 32 MPa and in the range of rotational speed n from 50 to 1500 rpm. The flow rate $Q_2$ was measured over the entire pressure drop $\Delta p$ range at n = const. (Fig. 8).
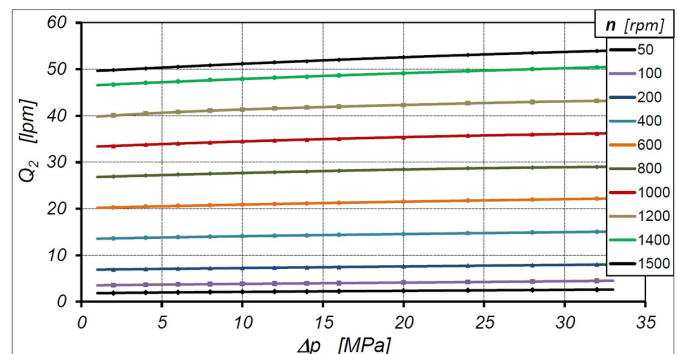


Fig. 8. Flow rate $Q_2$ vs $\Delta p$ at n = const

The external leakage $Q_{Le}$ in the tested motor was exceedingly small regardless of rotational speed n (Fig. 9).
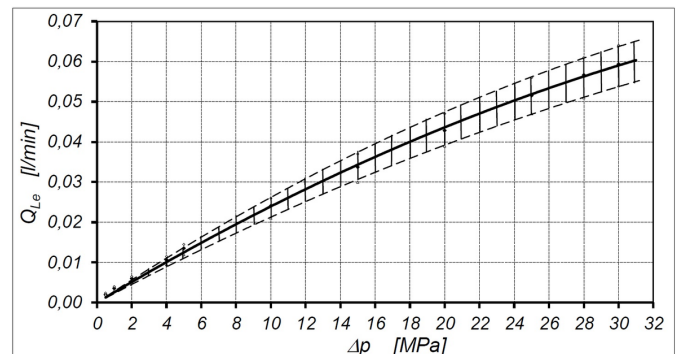


Fig. 9. External leakage $Q_{Le}$ in tested motor [36]

## 5.5. Characteristics of flow rate vs pressure drop in working mechanism

Pressure drop $\Delta p_{ich}$ in internal channels of the tested motor was calculated according to the formula (6):

$$\Delta p_{ich} = 0.003224 \cdot Q_2^2 + 0.02183 \cdot Q_2 \tag{46}$$

where $Q_2$ in [l/min] and $\Delta p_{ich}$ in [MPa] [35, 36, 38].

In Fig. 10 characteristics $Q_2 = f(\Delta p_i)$ at n = const. are shown. The pressure drop $\Delta p_i$ in working mechanism was calculated according to formula (5) and (46). The influence of liquid compressibility was omitted.
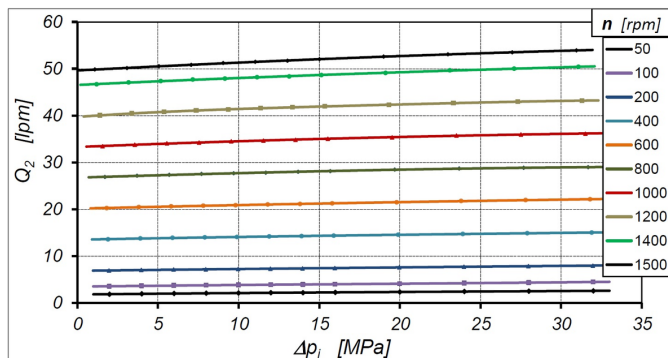


Fig. 10. Flow rate $Q_2$ vs $\Delta p_i$ at n = const. The influence of liquid compressibility was omitted

The flow rate $Q_{2(pH)}$ is related to the pressure $p_H$ in the high-pressure chamber of the motor. Values of $Q_{2(pH)}$ was calculated according to formulas (29) and (31). The characteristics of $Q_{2(pH)} = f(\Delta p_i)$ are shown in Fig. 11.



Fig. 11. Flow rate $Q_{2(pH)}$ vs $\Delta p_i$ at n = const



Fig. 12. Characteristics of $q_{e2}$ vs $\Delta p_i$ at n=const – according to Balawender's method

The characteristics presented in the Fig. 10 and Fig. 11 show that the $Q_{2(pH)}$ is smaller than $Q_2$. In the whole range of pressure drop $\Delta p_i$, the difference does not exceed 2%.

## 5.6. Theoretical working volume according to Balawender method

The characteristics of $q_{e2} = f(\Delta p_i)$ determined according to Balawender's method are shown in Fig. 12.
From the above characteristics it can be seen that:

- $q_{t(50)} = 36.81$ cm³/rev;
- $q_{t(100)} = 35.112$ cm³/rev;
- $q_{t(200)} = 34.315$ cm³/rev;
- $q_{t(1500)} = 33.221$ cm³/rev.

Then, according to formula (18) the theoretical working wolume of the satellite motor is $q_t = 34.757$ cm³/rev.
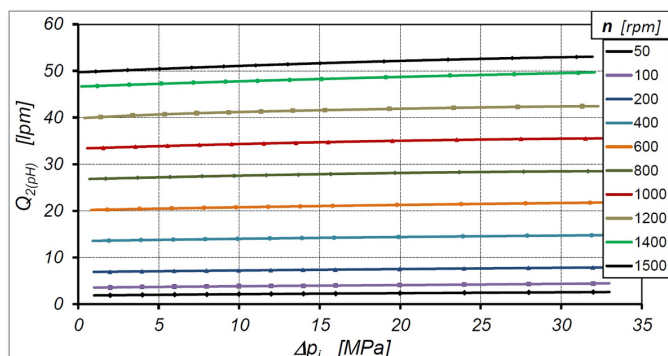
## 5.7. Theoretical working volume according to proposed new method

According to proposed new method the characteristics of $q_{e2(pH)} = f(\Delta p_i)$ (shown in Fig. 13) were calculated taking into consideration flow rate $Q_{2(pH)}$ related to the pressure in the high-pressure working chamber (Fig. 11). That is, the values of $q_e$ were calculated according to formula (33). All characteristics of $q_{e2(pH)} = f(\Delta p_i)$ have been described by equations. These equations are presented in the Table 1. The characteristics of $q_n = f(1/n^{0.5})$ are shown inFig. 14.

The results of research shown that, according to the new method, the working wolume $q_{rn}$ should be described by equation (35) and then (Fig. 14 – the black line):

$$q_n = 24.989 \cdot n^{-0.5} + 32.527 \tag{47}$$

Table 1. Equations $q_{e(pH)} = f(\Delta p_i^3)$ of motor output flow rate at n = const.

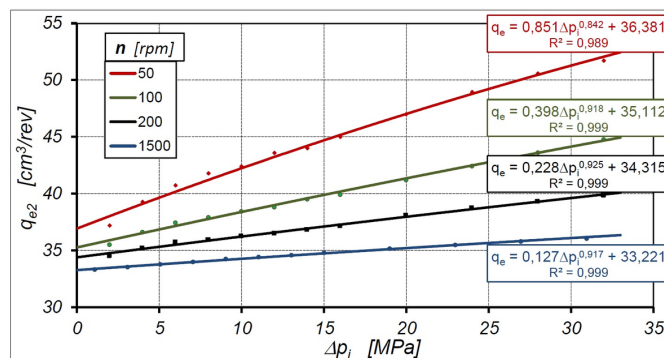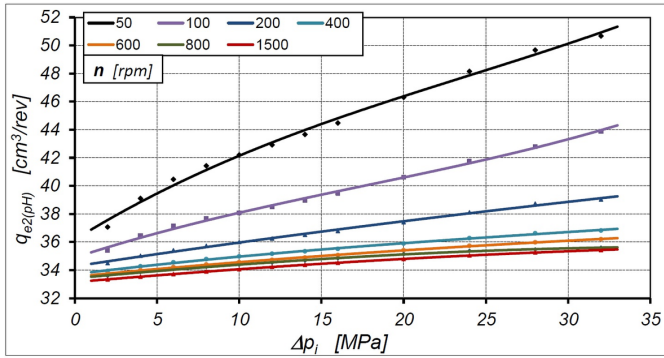| No. | n [rpm] | $q_{e2(pH)} = f(\Delta p_i^3)$ | $q_n$ [cm³/rev.] | $R^2$ |
|---|---|---|---|---|
| 1 | 50 | $q_{e2(pH)} = 0.000207\Delta p_i^3 - 0.014881\Delta p_i^2 + 0.725139\Delta p_i + 36.1738$ | 36.1738 | 0.9944 |
| 2 | 100 | $q_{e2(pH)} = 0.000149\Delta p_i^3 - 0.007912\Delta p_i^2 + 0.383909\Delta p_i + 34.8875$ | 34.8875 | 0.9969 |
| 3 | 200 | $q_{e2(pH)} = 0.000007\Delta p_i^3 - 0.001087\Delta p_i^2 + 0.179735\Delta p_i + 34.2685$ | 34.2685 | 0.9962 |
| 4 | 400 | $q_{e2(pH)} = 0.000023\Delta p_i^3 - 0.002158\Delta p_i^2 + 0.144543\Delta p_i + 33.7095$ | 33.7095 | 0.9979 |
| 5 | 600 | $q_{e2(pH)} = 0.000006\Delta p_i^3 - 0.001128\Delta p_i^2 + 0.114055\Delta p_i^2 + 33.5305$ | 33.5305 | 0.9985 |
| 6 | 800 | $q_{e2(pH)} = -0.000009\Delta p_i^3 - 0.000939\Delta p_i^2 + 0.107023\Delta p_i + 33.4235$ | 33.4235 | 0.9979 |
| 7 | 1000 | $q_{e2(pH)} = -0.000000\Delta p_i^3 - 0.001292\Delta p_i^2 + 0.110816\Delta p_i + 33.2959$ | 33.2959 | 0.9983 |
| 8 | 1100 | $q_{e2(pH)} = 0.000032\Delta p_i^3 - 0.002719\Delta p_i^2 + 0.122427\Delta p_i + 33.2802$ | 33.2802 | 0.9983 |
| 9 | 1200 | $q_{e2(pH)} = 0.000044\Delta p_i^3 - 0.003723\Delta p_i^2 + 0.142850\Delta p_i + 33.2614$ | 33.2614 | 0.9991 |
| 10 | 1300 | $q_{e2(pH)} = 0.000017\Delta p_i^3 - 0.001736\Delta p_i^2 + 0.107760\Delta p_i + 33.2487$ | 33.2487 | 0.9986 |
| 11 | 1400 | $q_{e2(pH)} = 0.000007\Delta p_i^3 - 0.000913\Delta p_i^2 + 0.090225\Delta p_i + 33.2403$ | 33.2403 | 0.9986 |
| 12 | 1500 | $q_{e2(pH)} = -0.000001\Delta p_i^3 - 0.000870\Delta p_i^2 + 0.100618\Delta p_i + 33.2245$ | 33.2245 | 0.9999 |

Fig. 13. Characteristics $q_{e2(pH)}$ vs $\Delta p_i$ at $n = const$ – according to the new method. Equations in Table 1



Fig. 15. Characteristics of $\Delta q_p$ plotted according to equations from Table 1
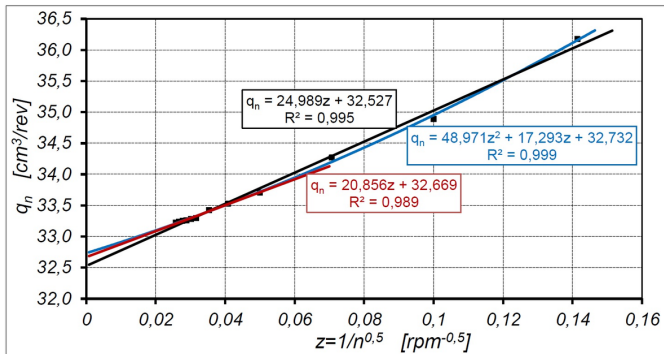


Fig. 14. Characteristics of working volume $q_n$ vs $n$ (according to the new method)



Fig. 16. Characteristics of $\Delta q_p$ vs $\Delta p_i$ for $n > 800$ rpm

That is, the theoretical working volume of satellite motor is $q_t = 32.527$ cm³/rev.

Nevertheless, the results of experiment, shown in Fig. 14, indicate some non-linearity between $q_n$ and $n^{-0.5}$ (the blue characteristic in Fig. 14). Then the $\Delta q_L$ is better to describe by following empirical formula:

$$\Delta q_L = \left( \frac{C_{id1}}{n} + \frac{C_{id2}}{n^{0.5}} \right) \cdot m^2 \cdot H \qquad (48)$$

It has been observed that if results for low speed $n$ of motor will be neglected, then the linear relationship between $q_n$ and $n^{-0.5}$ (the red line in Fig. 14) gives similar result in $q_t$ like nonlinear relationship (the blue line in Fig. 14). It is supposed that, the additional flow rate depends on the speed $n$ can exist in commutation unit of the satellite motor. Furthermore, at the present time, it seems reasonable to ignore low rotational speeds in the search of theoretical working volume $q_t$. For tested motor was ignored the low speed up to 400 rpm (the red line in Fig. 14). Then as the theoretical working volume of satellite motor should be adopted $q_t = 32.669$ cm³/rev. The relative difference between the $q_t = 32.527$ cm³/rev for all range of motor speed $n$ (the black line in Fig. 14) in about 0.44%.

## 6. The assessment of actual working volume

Taking into consideration formulas (22) and (33) and the formulas from Table 1 is possible to assess the actual working volume $q_r$ in tested hydraulic motor. Theoretically the $q_r$ would be independent on rotational speed $n$. But results of experiment show some deviations. In Fig. 15 are shown selected characteristics of $\Delta q_p$ plotted according to equations from Table 1.

It can be seen that for $n > 800$ rpm the $\Delta q_p$ takes a constant value. Characteristics of $\Delta q_p = f(\Delta p_i)$ for $n > 800$ rpm are shown in Fig. 16.

Therefore, taking into account formula (22), the actual working volume $q_r$ is:

$$q_r = 32{,}669 + 0{,}11167414 \cdot \Delta p_i + 0{,}00174171 \cdot \Delta p_i^2 + 0{,}00001286 \cdot \Delta p_i^3$$

$$[cm^3/rev.] \qquad (49)$$

Thus, the results of experiment confirm the theoretical considerations. That is, the pressure drop $\Delta p_i$ in motor's working chambers has a significant influence on the actual working volume $q_r$. Therefore, the actual working volume $q_r$ should be taken into account (instead of theoretical working volume $q_t$) for calculation the volumetric losses and mechanical losses in hydraulic motor (and the same volumetric efficiency and mechanical efficiency).

## 7. Discussion

The characteristics presented in the Fig. 10 and Fig. 11 show that the $Q_{2(pH)}$ is smaller than $Q_2$. In the whole range of pressure drop $\Delta p_i$. The difference does not exceed 1,6%. This difference is typical of mineral oil [54] and has an effect on the characteristics of flow rate per one revolution of the motor shaft (Fig. 13) and finally on the theoretical working volume $q_t$.

The analysis of experimental data allow to confirm the theoretical considerations that the $q_e$ is an nonlinear function of $\Delta p$ for $n=const$ regardless of the used method (Balawender method or the new method).

The value of the correlation coefficient $R^2$ of nonlinear function $q_e = f(\Delta p)_{n=const}$ is close to one ($R^2 > 0.99$, Table 1). Hence, the conclusion is that equations (25) and (26) very well describe fluid flow per one revolution of the motor shaft.

Based on the results of the motor tests, it can be concluded that regardless of the used method (Balavender or the new method):
  a) the rotational speed $n$ has no influence on the theoretical working volume $q_t$ of a hydraulic motor;
  b) pressure drop $\Delta p_i$ in the motor working chambers has no influence on the theoretical working volume $q_t$ of a hydraulic motor;
  c) the flow rate $q_e$ per one revolution of the motor shaft is a nonlinear function of pressure drop $\Delta p_i$ in the motor working chambers (Fig. 12, Fig. 13 and Table 1).

Furthermore the results of the satellite motor test confirm that:

a) is correct to describe the $q_e$ by a third order polynomial (expressed by the formula (25) or (26));

b) the flow rate $q_n$ (and the same $q_e$) is a nonlinear function of the motor rotational speed n. The results of experiment and calculations can be described by equation (34) or (35) with sufficient accuracy ($R^2 = 0.995$ – Fig. 14).

Nonetheless, the theoretical working voume $q_t$ calculated using the new method ($q_t = 32.669$ cm³/rev.) is about 6% smaller than $q_t$ calculated using Balawender's method ($q_t = 34.757$ cm³/rev.). This difference is not small. Designers of hydraulic positive displacement machines and researchers conducting development studies on these machines should therefore follow the method proposed in this article. Incorrectly adopted the value of the theoretical working volume gives an incorrect assessment of mechanical and volumetric losses.

Based on the analysis of the test results, it can be concluded that the motor working volume increases as a function of the pressure drop $\Delta p_i$ in the motor. For tested motor this increase is 2,2 cm³/rev. (Fig. 16). This is 6% of the theoretical working volume. It seems like a lot. But when analyzing the structure of the satellite mechanism in the engine (Fig. 4 and Fig. 5), it is concluded that the stiffness of the curvature and the planet are small. Thus, such an increase in the working volume at a pressure of 32 MPa is real.

Furthermore, characteristics of volumetric efficiency and mechanical efficiency are the test for correctness of the theoretical working volume assessment, because the maximum value of these efficiencies cannot be bigger than 1. If the theoretical working volume, calculated using the Balawender method, is taken, the volumetric efficiency of the motor for $\Delta p < 15$ MPa is larger than 1 ($\eta_v > 1$) (Fig. 17). It has no physical sense and the conclusion should be that the theoretical working volume has been miscalculated. Therefore, it confirms, that the Balawender's method is inaccurate and overestimates the theoretical working volume. Furthermore, for $q_t = 34,757$ cm³/rev. the mechanical efficiency is understated (Fig. 18).
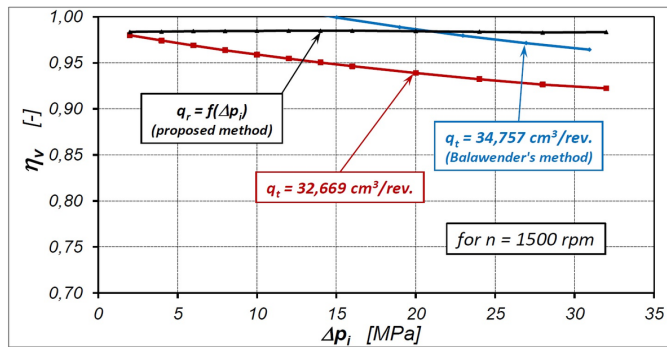


*Fig. 17. Volumetric efficiency $\eta_v$ vs $\Delta p_i$ of tested motor (n = 1500 rpm) – comparison for $q_t$ according to Balawender method and proposed new method.*
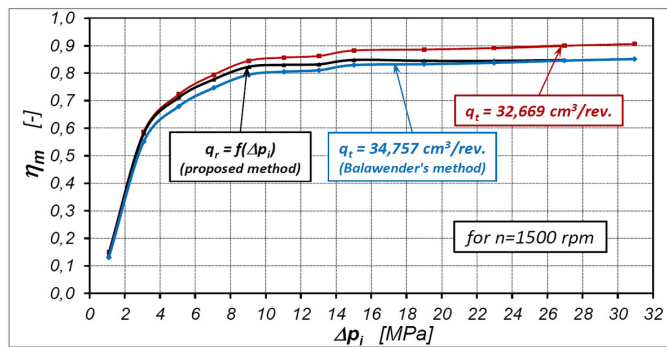


*Fig. 18. Mechanical efficiency $\eta_m$ vs $\Delta p_i$ of tested motor (n = 1500 rpm) – comparison for $q_t$ according to Balawender method and proposed new method.*

On the above figures, characteristics of efficiencies for actual working volume $q_r$ are also shown. Volumetric efficiency is nearly constant

in all range of pressure drop $\Delta p_i$. It shows that axial clearances of satellites and rotor decreasases under the influence of increasing the pressure drop $\Delta p_i$ in the motor. Thus, the axial clearance compensation unit in motor operating correctly and the increasing in actual working volume $q_r$ may be caused by small stiffness of curvature, rotor and teeth in these elements (also in satellites).

The next consideration is the difference between the geometric working volume $q_g$ (34.44 cm³/rev.) and the theoretical working volume $q_t$ (32.669 cm³/rev.) derived from the test results. The relative difference is about 5.5%. Thus, this difference is not small. The geometric working volume $q_g$ results from the CAD drawing documentation. Satellite, curvature and rotor are made by wire electrical discharge machining method (WEDM method). They are made with some technological allowance for finishing treatment (lapping). The size of the allowance and the final geometrical dimensions after lapping are not known – is the secret of the company manufacturing the satellite motors. Furthermore, there are clearances at the tops of the teeth in the satellite mechanism components. In effect is a certain volume included in the geometric working volume. As a result, the geometrical working volume of a satellite motor is smaller than that in the CAD documentation.

## 8. Conclusions

The new methodology for determining the theoretical and actual working volume of a hydraulic motor is presented in this article.

It has been shown that a simplified approach (consisting in accepting for calculations the flow characteristics as a function of the pressure drop $\Delta p$ measured in the motor ports and neglecting the liquid compressibility) to the calculation of the theoretical working volume $q_t$ of the hydraulic motor results in an overestimated value of this volume. For the motor presented in this article, the differences are as high as 6%. Thus, an overestimated value of the theoretical working volume results in overestimated volumetric efficiency and understated mechanical efficiency of the motor (calculated according to commonly known definitions, i.e. based on the theoretical working volume) (Fig. 17 and Fig. 18).

The article shows that for a satellite engine, for a theoretical working volume determined in a simplified manner (taking into account $\Delta p$), the volumetric efficiency is greater than one in the range of $\Delta p < 15$ MPa (Fig. 17). Thus, it proves a considerable inaccuracy of the method known and used so far. To a lesser extent, the overestimated value of the theoretical working volume reduces the value of the mechanical efficiency. This efficiency is underestimated by about 3% (Fig. 18).

Therefore, in order to determine the correct value of the theoretical working volume of a hydraulic motor, an analysis of the flows in the motor as a function of the pressure drop $\Delta p_i$ in its working chambers should be performed. In order to calculate the $\Delta p_i$ (according to the formula (5)) it is necessary to determine the pressure drop $\Delta p_{ich}$ in the internal channels of the motor.

It has also been shown that when determining the theoretical working volume, the compressibility of the liquid should not be neglected. The flow rate $Q_{(pH)}$ corresponding the value of high-pressure $p_H$ in the working chambers should be used for calculations the theoretical working volume.

The test results also confirmed that the working volume increases if the pressure drop $\Delta p_i$ in this motor increases (Fig. 16 and formula (22)). This working volume was called the actual working volume $q_r$. It has been shown that the actual working volume $q_r$ should be taken to calculate the volumetric and mechanical losses and also the volumetric and mechanical efficiency of the motor.

In subsequent publications, the influence of the type of liquid on the value of the theoretical and actual working volume of the hydraulic motor will be presented. Similar analyzes will also be carried out for the pump.

# References

1. Bak M. Torque capacity of multidisc wet clutch with reference to friction occurrence on its spline connections. Scientific Reports 2021; 11: 21305, https://doi.org/10.1038/s41598-021-00786-6.
2. Balawender A. Analiza energetyczna i metodyka badań silników hydraulicznych wolnoobrotowych (Energy analysis and methodics of testing of low-speed hydraulic motors). Scientific book of the Gdansk University of Technology, Mechanika 1988; 54. Gdansk University of Technology Publishing House.
3. Balawender A. Opracowanie metodyki wyznaczania teoretycznej objętości roboczej pomp i silników hydraulicznych wyporowych (The development of the methodology for the determination of the theoretical working volume of positive displacement pumps and hydraulic motors). PhD dissertation. Gdansk University of Technology: 1974.
4. Banaszek A. Methodology of flow rate assessment of submerged hydraulic ballast pumps on modern product and chemical tankers with use of neural network methods. Procedia Computer Science 2021; 192(4): 1894-1903, https://doi.org/10.1016/j.procs.2021.08.195.
5. Banaszek A, Petrovic R. Problem of non proportional flow of hydraulic pumps working with Constant pressure regulators in big power multipump power pack unit in open system. Technicki Vjesnik 2019; 26(2): 294-301, https://doi.org/10.17559/TV-20161119215558.
6. Ding H. Application of non-circular planetary gear mechanism in the gear pump, Advanced Material Research 2012; 591-593: 2139–2142, https://doi.org/10.4028/www.scientific.net/AMR.591-593.2139.
7. Garcia-Bravo J, Nicholson J. What is the real size of that pump? Fluid Power Journal 2018, https://fluidpowerjournal.com/real-size-pump/.
8. Guo S, Chen J, Lu Y, Wang Y, Dong H. Hydraulic piston pump in civil aircraft: current status, future directions and critical technologies. Chinese Journal of Aeronautic 2020; 33(1): 16-30, https://doi.org/10.1016/j.cja.2019.01.013.
9. Guzowski A, Sobczyk A. Reconstruction of hydrostatic drive and control system dedicated for small mobile platform. Proceedings of the 8th FPNI Ph.D Symposium on Fluid Power. 8th FPNI Ph.D Symposium on Fluid Power. Lappeenranta, Finland. June 11–13, 2014. V001T05A012. ASME, https://doi.org/10.1115/FPNI2014-7862.
10. International Organisation for Standardization. ISO 8426:2008. Hydraulic fluid power – Positive displacement pumps and motors – Determination of derived capacity, https://www.iso.org/standard/40351.html.
11. Jasinski R. Analysis of the heating process of hydraulic motors during start-up in thermal shock conditions. Energies 2022; 15(1): 55, https://doi.org/10.3390/en15010055.
12. Jasinski R. Problems of the starting and operating of hydraulic components and systems in low Ambient Temperature (Part IV): modelling the heating process and determining the serviceability of hydraulic components during the starting-up in low ambient temperature. Polish Maritime Research 2017; 24(3): 45–57, https://doi.org/10.1515/pomr-2017-0089.
13. Karpenko M, Bogdevicius M. Review of energy-saving technologies in modern hydraulic drives. Science – Future of Lithuania 2017; 9(5): 553-558, https://doi.org/10.3846/mla.2017.1074.
14. Karpenko M, Prentkovskis O, Sukevicius S. Research on high-pressure hose with repairing fitting and influence on energy parameter of the hydraulic drive. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2022; 24(1): 25–32, http://doi.org/10.17531/ein.2022.1.4.
15. Kim T, Kalbfleisch P, Ivantysynova M. The effect of cross porting on derived displacement volume. International Journal of Fluid Power 2014; 15(2): 77-85, https://doi.org/10.1080/14399776.2014.923605.
16. Klarecki K, Rabsztyn D, Hetmanczyk P. Analysis of pulsation of the sliding-vane pump for selected settings of hydrostatic system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2015; 17(3): 338–344, http://dx.doi.org/10.17531/ein.2015.3.3.
17. Kollek W, Osinski P, Stosiak M, Wilczynski A, Cichon P. Problems relating to high-pressure gear micropump. Archives of Civil and Mechanical Engineering 2014; 14(1): 88-95, https://doi.org/10.1016/j.acme.2013.03.005.
18. Li D, Liu Y, Gong J, Wang T. Design of a noncircular planetary gear mechanism for hydraulic motor. Mathematical Problems in Engineering 2021; https://doi.org/10.1155/2021/5510521.
19. Lisowski E, Filo G, Rajda J. Analysis of the energy efficiency improvement in a load-sensing hydraulic system built on the ISO plate. Energies 2021; 14(20): 6735, https://doi.org/10.3390/en14206735.
20. Luan Z, Ding M. Research on non-circular planetary gear pump. Advanced Material Research 2021; 339: 140-143, https://doi.org/10.4028/www.scientific.net/AMR.339.140.
21. Manring N, Williamson C. The theoretical volumetric displacement of a check-valve type, digital displacement pump. Journal of Dynamic System, Measurement and Control 2019; 141(3), https://doi.org/10.1115/1.4041713.
22. Michael P, Garcia-Bravo J. The determination of hydraulic motor displacement. Proceedings of the 17th Scandinavian International Conference on Fluid Power SICFP'21. Linkoping, Sweden. June 1-2, 2021, https://doi.org/10.3384/ecp182p188.
23. Oshima S, Hirano T, Miyakawa S, Ohbayashi Y. Development of a rotary type water hydraulic pressure intensifier. JFPS International Journal of Fluid Power System 2009; 2(2): 21-26, https://doi.org/10.5739/jfpsij.2.21.
24. Oshima S, Hirano T, Miyakawa S, Ohbayashi, Y. Study on the output torque of a water hydraulic planetary gear motor. Proceedings of the the Twelfth Scandinawian International Conference on Fluid Power. Tampere, Finland. May 18-20, 2011.
25. Osiecki L. Mechanizmy rozrządu hydraulicznych maszyn wielotłoczkowych osiowych (Commutation units of hydraulics axial piston machines). Monografie 2006; 72. Gdansk University of Technology Publishing House.
26. Osinski P, Deptula A, Partyka M. Hydraulic tests of the PZ0 gear micropump and the importance rank of its design and operating parameters. Energies 2022; 15(9): 3068, https://doi.org/10.3390/en15093068.
27. Osinski P, Warzynska U, Kollek W. The influence of gear micropump body asymmetry on stress distribution. Polish Maritime Research 2017; 24(1): 60-65, https://doi.org/10.1515/pomr-2017-0007.
28. Patrosz P. Influence of gaps' geometry change on leakage flow in axial piston pumps. In: Stryczek J., Warzyńska U. (eds) Advances in Hydraulic and Pneumatic Drives and Control 2020. NSHP 2020. Lecture Notes in Mechanical Engineering. Springer, Cham., https://doi.org/10.1007/978-3-030-59509-8_7.
29. Pobedza J, Sobczyk A. Properties of high-pressure water hydraulic components with modern coatings. Advanced Materials Research 2014; 849: 100-107, https://doi.org/10.4028/www.scientific.net/AMR.849.100.
30. Post W. Models for steady-state performance of hydraulic pumps: determination of displacement. Proceedings of the 9th Bath International Fluid Power Workshop. University of Bath, United Kingdom. September 9-11, 1996; 9: 339-352, https://research.tue.nl/en/publications/models-for-steady-state-performance-of-hydraulic-pumps-determinat.

31. Saheban Alahadi M J, Shirneshan A, Kolahdoozan M. Experimental investigation of the effect of grooves cut over the piston surface on the volumetric efficiency of a radial hydraulic piston pump. International Journal of Fluid Power 2017; 18(3): 181-187, https://doi.org/10.1080/14399776.2017.1337440.

32. Schlosser W M J, Hilbrands J W. Das theoretische Hubvolumen von Verdrangerpumpen. Olhydrauli und Pneumatik 1963; 4.

33. Schlosser W M J, Hilbrands J W. Das volumetrische Wirkungsgrad von Verdrongerpumpen. Olhydrauli und Pneumatik 1963; 12.

34. Sliwinski P, Patrosz P. Patent PL218888 Satelitowy mechanizm roboczy hydraulicznej maszyny wyporowej (Satellite operating mechanism of the hydraulic displacement machine). 2015; https://ewyszukiwarka.pue.uprp.gov.pl/search/pwp-details/P.401821.

35. Sliwinski P, Patrosz P. The influence of water and mineral oil on pressure losses in hydraulic motor. In: Stryczek J., Warzyńska U. (eds) Advances in Hydraulic and Pneumatic Drives and Control 2020. NSHP 2020. Lecture Notes in Mechanical Engineering. Springer, Cham. https://doi.org/10.1007/978-3-030-59509-8_10.

36. Sliwinski P. Determination of the theoretical and actual working volume of a hydraulic motor. Energies 2020; 13(22): 5933, https://doi.org/10.3390/en13225933.

37. Sliwinski P. Determination of the theoretical and actual working volume of a hydraulic motor – Part II (The method based on the characteristics of effective absorbency of the motor). Energies 2021; 14(6): 1648, https://doi.org/10.3390/en14061648.

38. Sliwinski P, Patrosz P. Methods of determining pressure drop in internal channels of a hydraulic motor. Energies 2021; 14(18): 5669. https://doi.org/10.3390/en14185669.

39. Sliwinski P. Satelitowe maszyny wyporowe. Podstawy projektowania i analiza strat energetycznych. (Satellite displacement machines. Basic of design and analysis of power loss). Monografie 2016; 155. Gdansk University of Technology Publishing House.

40. Sliwinski P. The influence of water and mineral oil on volumetric losses in hydraulic motor. Polish Maritime Research 2017; 24 (s1): 213–223, https://doi.org/10.1515/pomr-2017-0041.

41. Sliwinski P. The influence of water and mineral oil on mechanical losses in a hydraulic motor for offshore and marine application. Polish Maritime Research 2020; 27(2): 125-135, https://doi.org/10.2478/pomr-2020-0034.

42. Stawinski L, Kosucki A, Cebulak M, Gorniak vel Gorski A, Grala M. Investigation of the influence of hydraulic oil temperature on the variable-speed pump performance. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2022; 24(2): 289–296, http://doi.org/10.17531/ein.2022.2.10.

43. Stryczek S. Napęd hydrostatyczny (Hydrostatic drive). PWN: 2016.

44. Toet G, Johnson J, Montague J, Torres K, Garcia-Bravo J. The determination of the theoretical stroke volume of hydrostatic positive displacement pumps and motors from volumetric measurements. Energies 2019; 12(3): 415, https://doi.org/10.3390/en12030415.

45. Toet G. Die Bestimmung des theoretischen Hubvolumens von hydrostatischen Verdrangerpumen und Motoren aus volumetrischen Messungen. Olhydraulik Pnaumatik 1970; 14.

46. Ulanowicz L, Jastrzebski G, Szczepaniak P. Method for estimating the durability of aviation hydraulic drives. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22(3): 557–564, http://dx.doi.org/10.17531/ein.2020.3.19.

47. Volkov G, Fadyushin D V. Improvement of the method of geometric design of gear segments of a planetary rotary hydraulic machine. Journal of Physics: Conference Series 2021; 1889: 042052, https://iopscience.iop.org/article/10.1088/1742-6596/1889/4/042052.

48. Volkov G, Smirnov V, Mirchuk M. Estimation and ways of mechanical efficiency upgrading of planetary rotary hydraulic machines. IOP Conference Series: Materials Science and Engineering 2020; 709: 022055, https://iopscience.iop.org/article/10.1088/1757-899X/709/2/022055.

49. Volkov G, Smirnov V. Systematization and comparative scheme analysis of mechanisms of planetary rotary hydraulic machines. International Conference on Modern Trends in Manufacturing Technologies and Equipment ICMTMTE 2018; 224: 02083, https://doi.org/10.1051/matecconf/201822402083.

50. Wang C, Luan Z, Gao W. Design of pitch curve of internal-curved planet gear pump strain in type N-G-W based on three order ellipse. Advanced Material Research 2013; 787: 567-571, https://doi.org/10.4028/www.scientific.net/AMR.787.567.

51. Wilson W E. Performance criteria for positive displacement pumps and fluid motors. Transition ASME 1949; 71(2).

52. Wu X, Chen C, Hong C, He Y. Flow ripple analysis and structural parametric design of a piston pump. Journal of Mechanical Science and Technology 2017; 31, 4245–4254, https://doi.org/10.1007/s12206-017-0823-8.

53. Zaluski P. Experimental research of an axial piston pump with displaced swash plate axis of rotation. In: Stryczek J., Warzyńska U. (eds) Advances in Hydraulic and Pneumatic Drives and Control 2020. NSHP 2020. Lecture Notes in Mechanical Engineering. Springer, Cham., https://doi.org/10.1007/978-3-030-59509-8_12.

54. Zaluski P. Influence of fluid compressibility and movements of the swash plate axis of rotation on the volumetric efficiency of axial piston pumps. Energies 2022; 15(1): 298. https://doi.org/10.3390/en15010298.

55. Zaluski P. Wpływ położenia osi obrotu tarczy wychylnej na sprawność objętościową pomp wielotłoczkowych osiowych (Influence of the position of the swash plate rotation axis on the volumetric efficiency of axial piston pumps). Ph.D. dissertation. Gdansk University of Technology: 2017.

56. Zhang B, Song S, Jing C, Xiang D. Displacement prediction and optimization of a non-circular planetary gear hydraulic motor, Advances in Mechanical Engineering 2021; 13(11): 1–13, https://doi.org/10.1177/16878140211062690.

57. Zhao H, Wang B, Chen G. Numerical study on a rotational hydraulic damper with variable damping coefficient. Scientific Reports 2021; 11: 22515, https://doi.org/10.1038/s41598-021-01859-2.
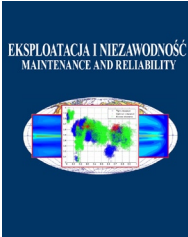
# An anomaly detection method based on random convolutional kernel and isolation forest for equipment state monitoring

**Indexed by:**
Web of Science Group

Xinhao Shu[a], Shigang Zhang[a,*], Yue Li[a], Mengqiao Chen[a]

[a] National University of Defense Technology, Laboratory of Science and Technology on Integrated Logistics Support, College of Intelligence Science and Technology, Deya str., Changsha, 410073, Hunan, China

## Highlights

- Random convolution kernel applied in anomaly detection for automatic feature extraction.

- The establishment of the initialization strategy of the 2-D random kernel.

- The anomaly-sensitivity evaluation is based on time series decomposing method.

- Method effectiveness are verified on varying dataset and with result analysis.

## Abstract

Anomaly detection plays an essential role in health monitoring and reliability assurance of complex system. However, previous researches suffer from distraction by outliers in training and extensively relying on empiric-based feature engineering, leading to many limitations in the practical application of detection methods. In this paper, we propose an unsupervised anomaly detection method that combines random convolution kernels with isolation forest to tackle the above problems in equipment state monitoring. The random convolution kernels are applied to generate cross-dimensional and multi-scale features for multi-dimensional time series, with combining the time series decomposing method to select abnormally sensitive features for automatic feature extraction. Then, anomaly detection is performed on the obtained features using isolation forests with low requirements for purity of training sample. The verification and comparison on different types of datasets show the performance of the proposed method surpass the traditional methods in accuracy and applicability.

## Keywords

anomaly detection, random convolutional kernel, isolation forest, multi-dimensional time series, equipment state monitoring.

## 1. Introduction

Anomaly detection is a fundamental technology to ensure the safety and reliability of systems. By analyzing the massive multi-dimensional data generated during the operation for system status monitoring and assessment can significantly improve the efficiency of system maintenance and reliability. However, there are still challenges in the current anomaly detection, such as the difficulty in distinguishing normal and abnormal samples in historical data, which makes it very complicated to manually split enough normal data for training deep learning models. The extraction of anomaly-sensitive features from these collected multi-dimensional data also strongly depends on empirics, resulting in a time-consuming process and increasing uncertainty. Therefore, it is necessary and valuable to propose a detection method that is capable of detecting under the anomaly-mixed data condition with automatic feature extraction.

Deep neural network (DNN) models are usually able to achieve higher detection precision compared to traditional machine learning (ML) models. Among previous approaches of unsupervised learning DNN models like Autoencoder [8] and its variants like VAE[11], DAE[21] and SAE[1] have made fruitful progress in anomaly detec-

tion field. For example, Zong et al. combined the Autoencoder with gaussian mixture model to jointly consider reconstruction error and the distribution of intermediate hidden layer's variables for anomaly detection [25]. Li et al. combined the intermediate layer of a Variational Autoencoder with the reconstructed error for anomaly detection [13]. DNNs large-scale parameters and iterative optimization enable the models to obtain accurate representations by learning from deep relationships and patterns automatically in training data. However, the existence of abnormal samples in the training set will cause the model to deviate from the potential distribution or low-dimensional representation of the normal samples since the data are not labeled, thus the model failing to precisely characterize the normal state and leading to inaccurate detection. Cheng et al. proposed a solution by using a combination of detection loss and reconstruction loss to optimize the learning process, reducing the model's ability of reconstructing the outliers [4]. Some traditional ML models also have similar problems, such as the OCSVM suffering from the anomalies in training which lead to bias of the support vectors [20]. Guo et al. used LOF clustering methods to preprocess the training data of OCSVM to reduce the influence of anomalous samples [7]. Most of the traditional ML based anomaly detection models are on the assumption that outliers

are the minority and distributed away from the center of dataset, such like the widely used method isolation forest proposed by Fei et al. [14, 15, 24]. In addition, the researches conduct detection from the perspective of distance [12], angle [9] and density [22] also achieve favorable results. However, when the abnormal data differ insignificantly from the normal data or under the complex data conditions, these methods can hardly achieve accurate detection by only using the superficial features on its specific perspective rather than analyzing the underlying patterns of the data like DNN methods. Therefore, this type of methods tends to require feature engineering to obtain suitable features for detection based on different business characteristics, which heavily relying on practitioner experience-based analysis. For example, Calheiros et al. encodes time as an additional feature in isolation forests anomaly detection [2]. Puggini et al. proposed a dimension reduction method based on forward selection group analysis and used the processed features for isolation forest detection, which has superior interpretability compared to the traditional PCA [16].

To summarize, DNN with deep feature extraction capabilities are sensitive to the purity of the training dataset, while some traditional ML models are not constrained by outliers in training but have difficulty in mining deep features and require manual feature engineering [3]. Thus, we attempt to combine the advantages of both, using the DNN approach for feature extraction and the ML for anomaly detection. The rocket series methods are successful examples based on this concept, which have performed well on time series classification tasks in recent years [5, 6, 19]. The method extracts features through massive 1-D random convolution kernels and uses these features as a high-dimensional description of the original data, and then using a linear classifier to classify the descriptions. Overall, the framework can be regarded as a single-layer convolutional neural network without feedback. Benefited from the random setting of the convolutional kernel parameters, the method avoids the back-propagation of the kernels optimization and achieve state of the art accuracy at a significantly reduction in time expense. Previous to the Rocket, feature extraction methods based on random convolution concept have been widely researched [1]. Jimenez et al. proposed a sequence similarity measure by convolving the target series with a random sequence [10]. In the study of Saxe et al., random convolution kernels are used in the feature extraction of images and the obtained features are used as objects for SVM classification [17], which indicates the possibility of applying random convolution methods on multidimensional objects. In the anomaly detection task of multidimensional time series, the fault pattern is complex including the vary of the value and the relation of the features, therefore the extracted fusion feature is more sensitive in representation of the equipment status [3]. By the above studies we can conclude that the random convolution kernel has great potential of extracting features from different dimensions and scales, and the random setting of the convolution kernel parameters provides an automatic approach for feature extraction.

In this regard, we propose an unsupervised anomaly detection method that combines random convolution kernels with isolation forest. To the best of our knowledge, few researches have utilized random kernels in anomaly detection task and combine these two methods. Specifically, we establish an initialization strategy for the parameters of the random convolution kernels, and generate feature series through sliding these kernels with dot production in the multidimensional sequence. To filter the invalid feature series during the generation, we propose a selection method using time series decomposing algorithm. This method evaluates anomaly sensitivity of feature series by analyzing the similarity of its split points, and the several most sensitive feature series are used as the object for isolation forest. The effectiveness of the proposed method is evaluated on a turbine engine simulation dataset and bearing vibration datasets with comparing to other anomaly detection methods.

## 2. Method

The framework of the proposed method is shown in Fig. 1. In the model initialization stage, the random convolution kernels are initialized to generate massive feature series which describes the original data from different perspectives. Then the feature series selection method sorts the generated feature series by anomaly sensitivity and uses several most sensitive features as the detecting objects of the isolation forest. The anomaly threshold is set by the sample's anomaly score given by the isolation forest. In the test phase, the feature series of the input data are generated by the kernels corresponding to the anomaly-sensitive features in the initialization, and anomaly detection is performed by the same isolation forest had been trained.
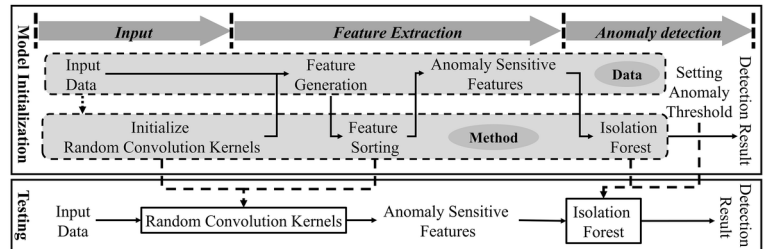


*Fig. 1. Framework of the proposed method*

### 2.1. Feature series generation by random convolution kernels

Random convolution kernel is similar to the convolution layer in the neural network, with the difference that the parameters of the former are randomly generated and the training process is not required. From the perspective of signal analysis, the operation of 1-D convolution kernel is similar to the wavelet analysis process, except that wavelet analysis uses a wavelet function, whereas the function of the random convolution method is a random sampling of specific distributions. Fourier transform is also commonly used in signal analysis, but it can only indicate the frequency characteristics of anomalous signals, while wavelet analysis can indicate the location of anomalies. By using wavelet functions with different parameters, wavelet analysis is able to extract signals of different frequencies [5]. The 1-D random convolution method is extended further on this basis, by randomly setting parameters like kernel weights, kernel length and dilation to extract many types of features at different frequencies and scales.

Intuitively, the occurrence of anomalies will inevitably lead to differences in the time series at different frequencies and scales, which have been the object of previous researches by using traditional signal analysis methods. In multidimensional time series, anomalies also lead to variations in the relationship between different features. As an example, Fig. 2 depict a simple situation.
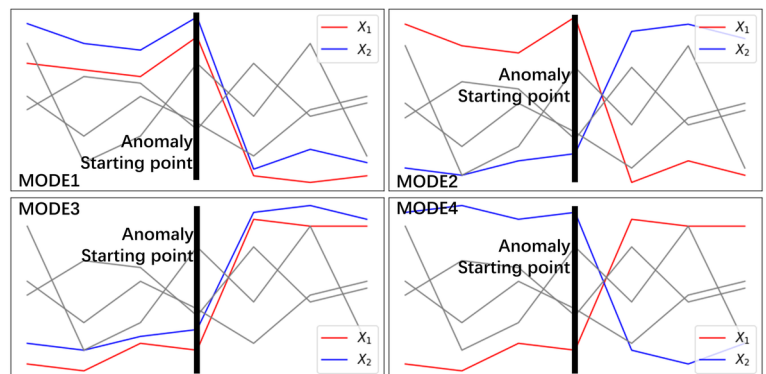


*Fig. 2. Anomalous patterns*

As shown in Fig. 2, the occurrence of an abnormality causes a change in the values of features $x_1, x_2$, with four different combination patterns. If summing the two signals in modes 1 and 3, and differencing the two signals in modes 2 and 4, the result of this kind linear combination can enlarge the amplitude of value change between normal and abnormal state. Based on the above description, the proposed feature extracting method of 2-D convolutional kernel is depicted in Fig. 3, where a large number of kernels with different parameters are successive sliding over the original data with dot production to generate a sequence of features. The dot production over different dimensions can be regarded as a linear combination of the features extracted by multiple 1-D convolution kernels, and the result not only expresses the extracted features variation of individual dimension, but also takes into account the relative variation of features in different dimensions. The random setting of the parameters of the convolution kernel implies the randomization of the weights of features participating in the linear combinations, and by generating massive kernels will enable the method to obtain plentiful features from diverse perspective.

Massive convolutional kernels will be randomly initialized



Fig. 3. Random convolution kernel based feature extraction of multidimension time series

whether the detection object is 1-D or multidimensional time series. In kernels initialization, we referred to the Rocket [5] approach and proposed a 2-D random convolutional kernel initialization strategy for multidimensional time series anomaly detection tasks. Specifically, it can be divided in to shape initialization and weights initialization.
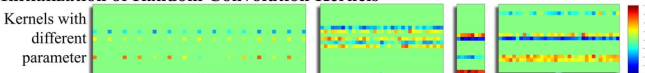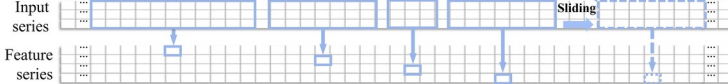
The shape of the kernel is controlled by three parameters: its width is determined by the dimension of the input data, and its length is jointly determined by the original length and the dilation.

- The original length of the convolution kernel is chosen randomly with equal probability within a candidate set, and the candidate value depends on the type of signal and its frequency of the detecting object. For signals with high frequency such as acceleration sensors, a larger candidate value is usually given such like $\{10, 20, 40, 80\}$, while for slowly varying signals with low frequency such as temperature and pressure, a smaller candidate value is usually selected, such like $\{5, 9, 13, 17\}$.
- The dilation is sampled from an exponentially distributed sequence to acquire convolutional kernels with varying sparsity, and to avoid the convolutional kernels from being too sparse, we add a factor $\alpha \in [0,1]$ to control them. Let the length of the input data be $l_{in}$, the original length of the convolution kernel be and $l_r$, the dilation $d$ and the kernel length after dilation be $l_k$.

$$d = \left\lfloor 2^x \right\rfloor, x \sim U\left(0, \log_2\left(\alpha \frac{l_{in}-1}{l_r-1}\right)\right) \quad (1)$$

$$l_k = l_r + (l_r - 1) \times d \quad (2)$$

The number of the dimension of original series participated in the linear combination of the kernel is controlled by the parameter $m$, which is an integer randomly sampled in the range $[1, n]$, where $n$

is the number of the original dimension of input data. Here only the number of selected dimensions is set, while the specific dimensions are randomly selected, so that convolution kernels with the same parameter $m$ will extract features from different combinations of dimensions. Fig. 4 depicts the role played by the above parameters in the convolution kernel.
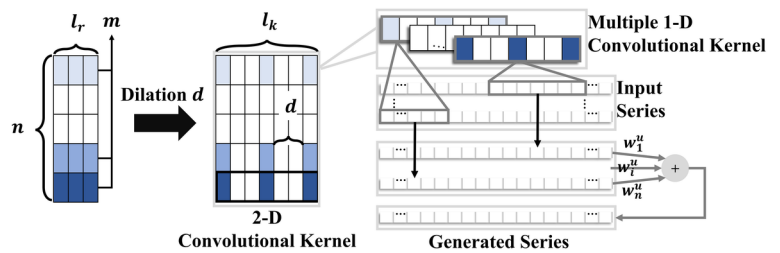


Fig. 4. Convolutional Kernel

The initialization strategy of weight largely determines the quality of the generated feature series. Operationally, the calculating result of the 2-D kernel can be regarded as the linear combination of the results of 1-D kernel, as shown in Fig. 4. Reflecting on the weight matrix, the weight vectors of the 1-D kernel are arranged vertically on the matrix of 2-D kernel, and the diversity of linear combinations is achieved by assigning different weights to each vector's result. Let the length of the original kernel be $l_r$ and the number of dimensions of the original data be $n$, from which $m$ dimensions are randomly selected to participate in the operation. Suppose the weight matrix of 2-D kernel is denoted as $W = \left[w_1, w_2, ..., w_n\right]^T$ ($w \subset R^{n \times l_r}$), where $w_i = \left[w_{i1}, w_{i2}, ..., w_{il_r}\right]$ ($w_i \subset R^{l_r}$) represents 1-D kernel's weight vector. The specific weight initialization process can be divided into two steps, as follows:

(1) Assignment of weights to the 1-D kernels in the linear combination: The corresponding $m$ weights represent by $w_u = \left[w_1^u, w_2^u, ..., w_m^u\right]$, and we analyze the influence of the weights on the generated features series from data distribution. Assuming that the sequence of $m$ dimensions involved in the operation is $X = \left[X_1, X_2, ..., X_m\right], X_i \sim N\left(u_i, \sigma_i\right)$, the distribution of the generated series $X'$ can be described as follows:

$$\begin{aligned} P(X') &= \sum_{i=1}^m w_i^u N(u_i, \sigma_i^2) \\ &= N\left(\sum_{i=1}^m w_i^u u_i, \sum_{i=1}^m (w_i^u)^2 \sigma_i^2\right) \end{aligned} \quad (3)$$

We limit that the weight vector $w_u$ is sampled from the distribution $U\left(-1/\sqrt{m}, 1/\sqrt{m}\right)$, so that the variance of $X'$ is $\sigma'^2 \leq (1/m)\sum_{i=1}^m \sigma_i^2$. This ensures that the variance of the generated series always smaller than the mean value of the variance of all the dimensions involved in the linear combination. Thus, avoiding the abnormal signal being swamped by the increasing variance. Meanwhile, checking whether the condition $\left|\sum w_u\right| < \varepsilon (\varepsilon \approx 0)$ is satisfied, otherwise re-initialize $w_u$. This constraint ensures that $w_u$ is as even as possible over the positive and negative intervals and avoids assigning too much computational weight to single dimension in the linear combination.

(1) Initialization of the weight vector of the 1-D kernel. The weight vector $w_i$ of the 1-D kernel corresponding to the dimensions

involved in the operation is sampled from $N(w_i^u, \sigma^2)$. For the remaining unselected dimensions, the weight vectors are $w_i = \mathbf{0}$. Where the variance $\sigma^2$ is set to 0.01 in this paper. For the special case that the input data is 1-D series, the 2-D weight matrix of the kernel is taken to be compressed into one dimension by $w = \sum w_i$. The initialization of kernel weights is summarized in Algorithm 1.

---

**Algorithm 1** Initialization of Kernel Weights

---

**Input:** Kernel length $l_k$, Dimension of input data $n$, Number of dimensions involved in the operation $m$, Variance $\sigma$, Error $\varepsilon\,(\varepsilon \approx 0)$

   **Output:** Weight Matrix $w$

1:     Initialize $w[1,2,...,n], w_u[1,2,...,m], idx[1,2,...,m]$ (array)

2:     $idx \leftarrow sample([0,1,...,n], m)$

3:     **while** True **do**

4:        $w_u \leftarrow sample\left(\left\lfloor U\left(-1/\sqrt{m}, 1/\sqrt{m}\right)\right\rfloor, m\right)$

5:        **if** $\left|\sum w_u\right| < \varepsilon$ **then**

6:           break

7:        **end if**

8:     **end while**

9:     **for** $i = 1 \rightarrow n$ **do**

10:       **if** $i \in idx$ **then**

11:          $w[i] \leftarrow sample(N(w_u[i], \sigma^2), l_k)$

12:       **else**

13:          $w[i] \leftarrow \mathbf{0}$

14:       **end if**

15:     **end for**

    return $w$

---

The preprocessing includes standardization to avoid the impact of different units and padding to ensure the generated series by different kernels are aligned in time. On the assumption that there are no anomalies in the earlier data, for kernel of length $l_k$ and dilation $d$, repeating the first $l_k + (l_k - 1) \times d$ in front of the input data to pad the sequence. During the convolution, the kernel slides over the preprocessed data and conduct dot production. For a $n$ dimensions, $L$ samples multidimensional time series $X = [X_1, X_2, ..., X_n]^T\,(X \subset R^{n \times L})$, preprocessed as $\overline{X}$, kernel $K(W, l_k)$ produces a feature series $T'$:

$$T' = [t'_1, t'_2, ..., t'_L] \tag{4}$$

$$t'_i = \sum_{j=1}^{n} w_j \cdot \overline{X}_{j,[i,i+l_k]} \tag{5}$$

If $k$ kernels are initialized, generation with $X$ produces a k-dimensional time series $T = [T'_1, T'_2, ..., T'_k]^T\,(T \subset R^{k \times L})$, each $T'_i$ represent a feature series generated by a specific kernel.

## 2.2. Abnormal sensitive feature series selection

The random initialization strategy of the parameters of the random convolution kernel can extract different features of the original data, but it will inevitably generate many irrelevant features at the same time. The isolation forest itself has a certain anti-irrelevant feature capability, but since the proportion of irrelevant features in the generated features is difficult to estimate, preprocessing the generated features can further enhance the stability of the combined method. Since abnormal event cause changes in the content of series segments, we use the time series decomposing method, which split series into segments by its content, to obtain the split points caused by anomalies.

By analyzing the similarity of the series split points distribution with that of anomalous split points, the anomaly-sensitivity of generated series can be evaluated and then sort the generated series with the sensitivity. In this section, we combine the relevant research on time series decomposition methods by Zhao et al. [23]. The feature series selection can be divided into three parts specifically: ①Converting the generated numerical series into symbolic series, which characterize the result of numerical analysis and describe the data in a more abstract way to facilitate the following content-based decomposition. ②Using time series decomposing method to search the split points of each symbolic series and calculating the integrated distribution of all split points. ③Evaluate the anomaly sensitivity of the generated series by the similarity between the distribution of the split points of each sequence and the distribution of the integrated split points. Then rank the generated series based on sensitivity and select several top-ranking series as the object of anomaly detection. The main framework of the method is depicted in Fig. 5, and the specific technical details are described in the following.

### 2.2.1. Symbolic series conversion

Symbolic series describe the numerical series in a more abstract approach. In Zhao et al.'s study [23], he used a 1-D directional gradient histogram (HOG) to describe the shape character of the sampled sequences, and clustered the description vectors into a certain number of categories. The conversion of series from numerical to symbolic is accomplished by replacing the numerical values of subsequences using the labels of each category. In this paper, the description of the original series has been completed within the generation of feature series by random convolution kernel, which not only considers multiple dimensions but also avoid the limitation of specific description methods to describe different types of data. Among the description approaches of anomalies, the simplest form is to directly express them as numerical magnitudes rather than shapes or frequencies. Therefore,
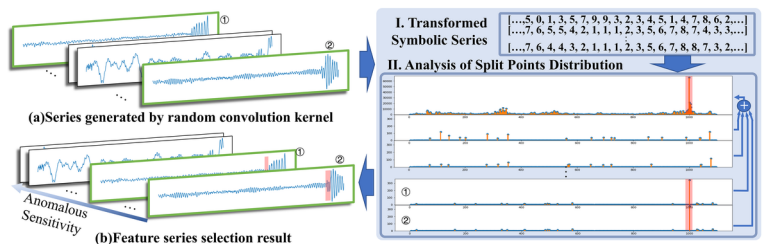
*Fig. 5. Feature series selection framework*

we only research the pattern that describe anomalies in numerical values by using the simplest equal-interval division method for symbolic transformation of the series data. This method evenly divides the interval between maximum and minimum of the object sequence into $n$ segments and uses the symbols $c_i \in \{1, 2, ..., n\}$ to describe the value falling in, thereby realizing the transition from the time series $T = t_1, t_2, ..., t_L\,(t_i \in R)$ to the symbolic series $C = c_1, c_2, ..., c_L$. The conversion process of $T \rightarrow C$ is shown in Fig. 6.
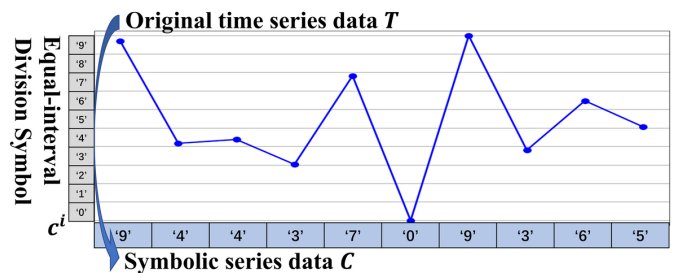
*Fig. 6. Equal-interval symbolic series conversion*

## 2.2.2. Split points search

The purpose of the split points searching is to obtain the location of possible abnormality on the time axis. As above analysis, anomalous events can cause changes in the content of the series segments, which can be reflected as an increase in the disorganization of the symbol series $C$. And this disorganization can be quantified by using information entropy. In the searching process, the location of anomalous events is obtained by the target that maximizing the information gain of each split point in series $C$, so that the character descriptions within each segment are homogeneous and between segments are heterogeneous. This part refers to the search method for split points in the study of Zhao et al. which is briefly outlined below [23]. Specifically, the symbolic sequence $C$ is recursively decomposed until all its subsequences reach the minimum length $l_s$ or the number of split points reaches the limit $\beta$. Each split point $s$ cuts the parent sequence $C_p$ into two subsequences $\{C_l, C_r\}$. Using the information gain as the criterion for the selection of the segmentation points $\Delta E(s, C_p)$:

$$\Delta E(s, C_p) = E(C_p) - \left[ \frac{|C_l|}{|C_p|} E(C_l) + \frac{|C_r|}{|C_p|} E(C_r) \right] \quad (6)$$

$$E(C) = -\sum_{i=1}^{n} p_i^C \log p_i^C \quad (7)$$

where $p_i^C$ is the frequency of symbol $i$ in the sequence $C$. The split point with the maximum information gain is denoted by $s^*$. The searching for best split points can be transformed into the maximization problem:

$$\arg\max_{s \in C_p} \{\Delta E(s, C_p)\} \quad (8)$$

And the weight $v_{s^*}^{C_p}$ for the split point $s^*$ is:

$$v_{s^*}^{C_p} = |C_p| \cdot \Delta E(s^*, C_p) \quad (9)$$

The split point search is summarized in Algorithm 2 [23].

---

**Algorithm 2** Searching for split points

**Input:** Input series $T = t_1, t_2, ..., t_L$, Limitation number of split points $\beta$, Minimum length of the subsequence $l_s$

**Output:** Split points set $S$, Weights of the split points $V$

1:     Symbolic series conversion $T \to C$

2:     Initialization set $Q = \{C\}, S = \varnothing, V = \varnothing$

3:     **while** $Q \neq \varnothing$ **do**

4:         $C_p \leftarrow Q.dequeue()$

5:         **if** $|C_p| < l_s$ **then**

6:            continue

7:         **else**

8:            Obtain $s^*$ of $C_p$ according to Eq.(8), $S.add(s^*)$

9:            Obtain $w_{s^*}^{C_p}$ of $s^*$ according to Eq.(9), $V.add(v_{s^*}^{C_p})$

10:       **if** $|S| > \beta$ **then**

11:           break

12:       **end if**

13:       $Q \leftarrow Q.enqueue(\{C_l, C_r\})$

14:       **end if**

15:     **end while**

16:     **return** $S, V$

---

Fig. 7 depicts the split points obtained by the above searching algorithm, such as the distribution of split points (c), (e), (g) for feature series (b), (d), (f).
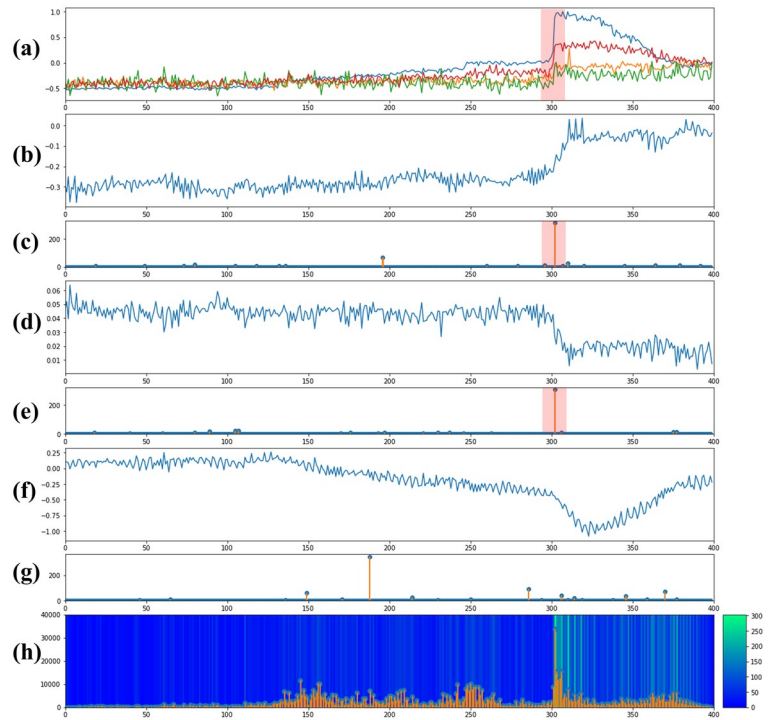


*Fig. 7. Distribution of the split points location and its weight and density of part of the generated feature series. (a) is the input data. (b), (d), (f) are the feature series generated by random convolution kernel. (c), (e), (g) are the corresponding split points' location and weights. (h) is the density (depict by background color) and weight (depict by vertical axis) and location (depict by horizontal axis) of integrated split points of all generated series*

### 2.2.3. Feature series ranking by anomalous sensitivity

As the generated feature series explain the anomalies in different perspective, the searched split points of each series are dominated by anomalies with varying strength. But the distribution of anomalous split points of different features series has similarity in location on the time axis since the anomaly is prevalent in all features. This similarity will be expressed as the density of split points caused by anomalies is higher than that caused by other reasons in the cumulative distribution among split points of all feature series, which is depict in Fig. 7(h) by background color. As it shown that the high-density location of Fig. 7(h) is consistent with the position of the anomaly start in (a). On the other hand, when the part of the split points of a series are dominated by anomalies, its weights is also significantly higher than that of the points caused by other reasons within the series. And this can be found from the generated feature series in Fig. 7(b)(d), which can significantly distinguish normal data from abnormal data. In the distribution of its split points (c) and (e), the weight of the anomalous split points are significantly higher than that of the other points, which coincides with the peak of the cumulative weight and the high-density position in the integrated distribution (h). Comparing to the feature series (f), there is also a relative high weight of its point distribution in (g) might be dominated by other reasons, thus the distribution is not consistent with the integrated distribution, and the difference between normal and abnormal in (f) is relatively obscure. Together with

the above two factors, the features with strong expression ability for anomalies have two characteristics: ① Its split points are distributed at the higher density locations in the integrated distribution. ② Its split points with high weight are also near the peak of the cumulative weight in the integrated distribution. Based on the above analysis, the task of evaluating the anomaly sensitivity of the feature series transformed into measuring the similarity of the split point locations and weights to the integrated distribution.

Specifically, for a multidimensional time series $MT = \{T_1, T_2, ..., T_k\}$ with $k$ features series, the corresponding split points set is $MS = \{S_1, S_2, ..., S_k\}$, and by computing the concatenation of $MS$ to obtain the integrated split point set $\mathbb{S} = S_1 \cup S_2 \cup ... \cup S_k = \{\mathbb{s}_1 \cup \mathbb{s}_2 \cup ... \cup \mathbb{s}_\gamma\}$. The weight of the points in $\mathbb{S}$ is $\mathbb{V} = \{\mathbb{v}_1, \mathbb{v}_2, ..., \mathbb{v}_\gamma\}$, where $\mathbb{v}_i$ is the sum of the weights of split point $\mathbb{s}_i$ in each set of $MS$:

$$\mathbb{v}_i = \sum_{\mathbb{s}_i \in \mathbb{S} \cap \mathbb{s}_i \in S_j} V_{\mathbb{s}_i}^j \tag{10}$$

where $V_{\mathbb{s}_i}^j$ is denoted as the weight value corresponding to the split point $\mathbb{s}_i$ of $\mathbb{s}_j$, and $V_{\mathbb{s}_i}^j = 0$ when $\mathbb{s}_i$ does not exist in the $S_j$. The frequency of each split point in $\mathbb{S}$ is denoted as $\mathbb{P} = \{\mathbb{P}_1, \mathbb{P}_2, ..., \mathbb{P}_\gamma\}$, where $\mathbb{P}_i$ is the frequency of $\mathbb{S}_i$ occurring in all split points set $MS$:

$$\mathbb{P}_i = \left( \sum_{\mathbb{s}_i \in \mathbb{S} \cap \mathbb{s}_i \in S_j} 1 \right) \Big/ \sum_{j=1}^{k} |S_j| \tag{11}$$

Kullback-Leibler (KL) divergence can measure the extent of one distribution explains the other, which can be interpreted as the similarity of one distribution to the other, but it is asymmetric. In this study, we modify the discrete KL formula to jointly consider the split point weights. Using the factor $\rho_i$ to represent the similarity between the distribution of split point set $S_i$ of feature series $T_i (T_i \in R^L)$ and the integrated distribution $\mathbb{S}$:

$$\rho_i = \sum_{j=1}^{\gamma} \mathbb{v}_{\mathbb{S}_j} \frac{v_{\mathbb{S}_j}^i}{\sum v_i} \times p_{\mathbb{S}_j} \log \frac{p_{\mathbb{S}_j}}{\mathbb{P}_{\mathbb{S}_j}} \tag{12}$$

where $V_i = \{v_{\mathbb{S}_1}^i, v_{\mathbb{S}_2}^i, ..., v_{\mathbb{S}_\beta}^i\}$ is the weight of the split point in $S_i$, and $p_{\mathbb{S}_j} = 1 / L$. Intuitively, the second part of the equation is the traditional discrete KL. The explanation for the value of $p_{\mathbb{S}_j}$: For a time series with $L$ samples, the probability that any point is selected as a splitting point is $1 / L$. Calculating the KL can be understood as the process of sampling each point in the point set $\mathbb{S}$, which explain the second part of the Eq.(12).

The first part of the Eq.(12) considers the weights of the split points. The higher the ratio of the weight $v_{\mathbb{S}_j}^i$ contributes among the sum of the weights $\sum V_i$ of the series $T_i$, the greater the difference between the data segments before and after the point $\mathbb{S}_j$. This ratio is then multiplied with the accumulated weight $\mathbb{v}_{\mathbb{S}_j}$, indicating that only if other features also have high weights at split point $\mathbb{S}_j$, then the ratio can contribute more similarity to integrated distribution. Thus, the similarity factor $\rho_i$ accounting more for the prevalence of different feature series that all occur numerical changes at point $\mathbb{S}_j$. The traditional KL is numerically non-negative, but in Eq.(12), the meaning of its value is modified so that the first part of the equation is always greater than 0 and the larger the value, the more similar the weights of the two distributions are, while the second part is always smaller than 0. Therefore, factor $\rho$ is a negative value. And the smaller the value, the more similar the two distributions are, signifying that the split point distribution of the series is similar to the anomaly distribution,

which also means that the series has stronger anomaly sensitivity. Algorithm 3 summarizes the above description.

---

**Algorithm 3** Feature selection method

**Input:** Multivariate time series $MT = \{T_1, T_2, ..., T_k\}$

**Output:** Sorted multivariate time series $\overline{MT} = \{\overline{T}_1, \overline{T}_2, ..., \overline{T}_k\}$

1:  Initialize the set $\mathbb{S} = \varnothing, \mathbb{V} = \varnothing, \mathbb{P} = \varnothing, S = \varnothing, V = \varnothing, \rho = \varnothing$

2:  **For** $i = 1 \to k$ **do**

3:      Searching split point of $T_i$ according to Algorithm2,

4:      $S_i, V_i \leftarrow$ Algroithm2($T_i$)

5:      $S.add(S_i)$, $V.add(V_i)$, $\mathbb{S} \leftarrow \mathbb{S} \cup S_i$

6:  **end for**

7:  Obtain $\mathbb{V}$ from $\mathbb{S}, S, V$ according to Eq.(10)

8:  Obtain $\mathbb{P}$ from $\mathbb{S}, S$ according to Eq.(11)

9:  **For** $i = 1 \to k$ **do**

10:     Obtain $\rho_i$ from $\mathbb{S}, \mathbb{V}, \mathbb{P}, S, V$ according to Eq.(11), $\rho.add(\rho_i)$

11: **end for**

12: Sort $MT$ according to $\rho$, obtain $\overline{MT} = \{\overline{T}_1, \overline{T}_2, ..., \overline{T}_k\}$

13: **return** $\overline{MT}$

---

The above analysis demonstrates that the impact of anomalous events on all generated features is prevalent, so the feature selection method can capture this factor and rank feature series according to anomalous sensitivity, which signifies that the method is based on the assumption that the anomaly is exist. When there are no anomalies in the data, the feature series will be sorted with other factors that dominate split points, but this kind of obvious consistency is practically impossible in the condition, since these series descript the input data on different perspective. Therefore, the ranking results have limited impact to subsequent anomaly detection in this situation. The number of finally selected features series is not specified in this section, since the proportion of irrelevant features generated by the random convolution kernel is not controllable, and the purpose of the feature series selection aims to further reduce the risk of isolation forests being affected by irrelevant feature series. The several top rank series, obtained by ranking according to the similarity of the distribution of split points and integrated split points, consider the significant variation in all series comprehensively, which means these series contain the most information embedded in the rest features series. Therefore, a wide range of the specific number of the selected feature series can be chosen. In this paper, a precise number is obtained by the balance between the detection result and number of redundant feature series in the Section 3.3.2.

### 2.3. Isolation forest anomaly detection

Isolation forest assumes that the anomalies locate far from the center of the dataset with sparsely distribution, and using tree method by randomly select the feature and split point in leaf node to isolate the anomalies which usually have a short split path, as in Fig. 8 [14]. Without the feedback from the training set to update the parameters, isolation forest has limited sensitivity to the outliers in the training set. The method has linear time complexity which is suitable for multidimensional data with large volumes and with its excellent performance it is popularly used in industrial application.

The random process of tree node attributes equipped the method robustness to noisy features in the high-dimensional data. However, the number of interfering features in multi-dimensional series generated by random convolution kernel is difficult to estimate, and the excessive amount nonsense features will result unstable in detection result. By the propose feature series selection method to preprocess the mas-
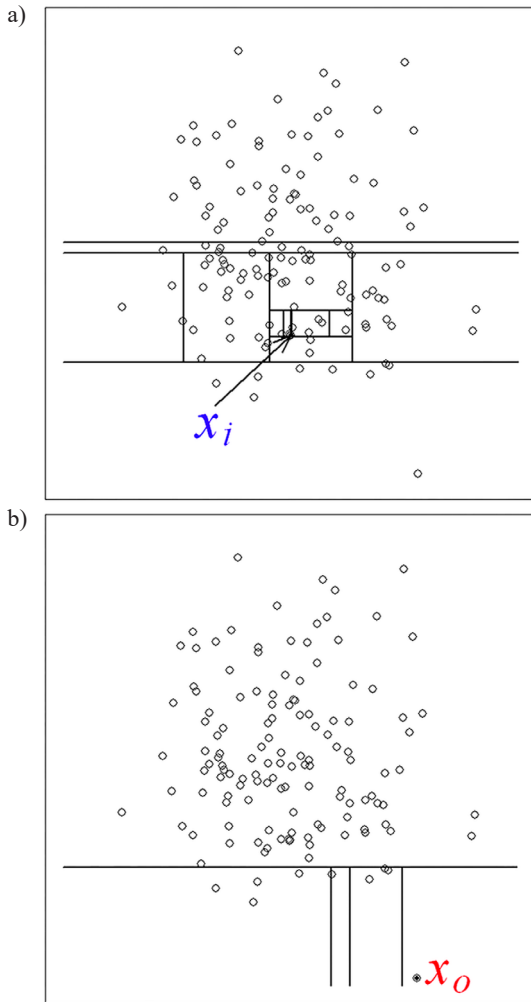
a)

b)

Fig. 8. Isolation forest anomalies detection: a) Isolation of normal point $x_i$, b) Isolation of outliner $x_o$

sive series can enhance the robustness of the isolation forest in the input side. Let the $m$ be the number of feature series which are selected in the rank in Section2.2.3, the $L$ samples selected feature series are represented by $\bar{M} = \left[ \bar{T}_1, \bar{T}_2, ..., \bar{T}_m \right]$, $\bar{T}_i = \left[ t_{i1}, t_{i2}, ..., t_{iL}, \right]^T$, then the detection object can be represented by $X_j = \left[ t_{1j}, t_{2j}, ..., t_{mj} \right]$.

## 3. Experimental results

In this paper, we conduct evaluation in three datasets, which include two different data types and varying dimensions to test the effectiveness of the method. And the proposed method is compared with traditional methods under the same data conditions.

### 3.1. Dataset

The three datasets used in this paper are the turbine engine simulation dataset and the public bearing vibration dataset. Among them, the simulation data of the aero-engine is generated by the GSP software, which is sampled at a low frequency (HZ). The bearing vibration dataset includes bearing failure data from Case Western Reserve University (CWRU) (https://engineering.case.edu/bearingdatacenter/welcome) [18] and bearing degradation data provided by NASA (http: //ti.arc.nasa.gov/project/prognostic-data-repository) which are sampled in high frequency (kHz). The details of the dataset are shown in Table 1.

The detail of the dataset is as follow:
- Gas Turbine Simulation Dataset: Gas turbine Simulation Program (GSP) is a software that can simulate various parameters of turbines engine and is often used to assist in structural design and operating condition analysis. Since it is difficult to consistently collect operating data under fault conditions in practical, this study uses efficiency degradation at the inlet fan as an abnormal event and simulates abnormal data at five different degradation rates by GSP. The generated dataset includes a total of 24 dimensions (sensors) such as temperature, pressure, air flow and thrust. The simulation model and generated data of Dataset1 is shown in

Table 1. Statistics of the datasets

|  | Dimension | Name | Type | Length | Anomaly ratio | Degradation rate |
|---|---|---|---|---|---|---|
| Gas Turbine Simulation Dataset (GSP) | 24 | #Dataset1 | train | 500 | 10% | 4% |
|  |  |  | test | 120 | 16.67% |  |
|  |  | #Dataset2 | train | 500 | 10% | 2% |
|  |  |  | test | 120 | 16.67% |  |
|  |  | #Dataset3 | train | 500 | 10% | 1% |
|  |  |  | test | 120 | 16.67% |  |
|  |  | #Dataset4 | train | 500 | 10% | 0.50% |
|  |  |  | test | 120 | 16.67% |  |
|  |  | #Dataset5 | train | 500 | 10% | 0.25% |
|  |  |  | test | 120 | 16.67% |  |
|  |  | #Dataset6 | train | 500 | 15% | 0.50% |
|  |  |  | test | 400 | 25% |  |
|  |  | #Dataset7 | train | 500 | 15% | 0.25% |
|  |  |  | test | 400 | 25% |  |
| Bearing Failure Dataset (CWRU) | 1 | #Dataset8 | train |  |  |  |
|  |  |  | test |  |  |  |
|  |  | #Dataset9 | train | 5000 | 20% |  |
|  |  |  | test |  |  |  |
|  |  | #Dataset10 | train |  |  |  |
|  |  |  | test |  |  |  |
| Bearing Degradation Dataset (NASA) | 4 | #Dataset11 | - | 984 | 46.14% |  |
|  |  |  | - |  |  |  |

Fig. 9. GSP model and data generation: a) GSP simulation model, b) GSP simulation data

normal is not known in advance, so we also use receiver operating characteristic (ROC) curve and the area under curve (AUC) to avoid the influence of the threshold setting. In addition, we repeat each test for several times to obtain the average performance to avoid the effect of randomness.

### 3.3. Analysis and Test

Initialization parameters of random convolution kernels: The number of convolution kernels is set to 1000. For the gas turbine simulation dataset, which is sampled by a low frequency, the set of the original kernel length candidates is set to $[2,3,4,5,8]$, for the CWRU bearing data is set to $[8,15,30,50,80,100]$, and for the NASA bearing degradation data set to . These two sets of vibration data are sampled at a higher frequency, so a larger original length of the kernel is considered. The number of features series participate in each kernel, that is parameter in section 2.1, is chosen randomly among . The dilation controller in Section 2.1 is set to 0.2. The top 10 features in the anomaly sensitivity rank are selected for all datasets to isolation forest according to the analysis in Section 3.3.2. The number of random trees in the isolation forest is set to 100.

#### 3.3.1. Effectiveness of feature series selection

In the first part of the experiment, we validate the proposed feature filtering method. Taking the Dataset5 as an example, the filtered features obtained by the proposed method in section 2.2 are shown in Fig. 10.

As shown in Fig. 10(b), the nonsense feature series among the massive generated series are ranked in the bottom and anomaly sensitive feature series are ranked in the top, which proves that

Dataset1 has the largest degradation rate among all the simulation datasets, and it can be seen from Fig. 9(b) that the difference between normal and abnormal is relatively obvious. The normal data are sampled from the engine operating under no degradation, and the abnormal data are sampled under the efficiency is degraded by 4% in the fan inlet. The simulation dataset is used to test the effectiveness of the proposed method on multi-dimensional data sampled at a low frequency.

- Bearing Failure Dataset: Consists by concatenation of normal and abnormal data from the same operating conditions in the original dataset. Abnormal events including inner race, outer race and ball damage. The dataset contains only one dimension, which is the sensor at the drive-side of the equipment. The dataset is used to validate the effectiveness on one-dimensional high-frequency signals.
- Bearing Degradation Dataset: In this dataset, four vibration sensors were used to measure the vibration of four bearings respectively and recorded the degradation process collectively, where the degradation of each bearing was inconsistent. We analyzed the abnormal starting points of the dataset and labeled all the succeeding data as abnormal. The dataset is used to test the effectiveness on multidimensional high-frequency signals.

### 3.2. Evaluation Metrics

This paper considered *precision* , *recall* and *F1score* for anomalous samples to evaluate the performance of the methods. The anomaly detection algorithm usually provides the anomaly score for each sample. Thus, the threshold setting determines the result of the detection. In order to focus on the performance of algorithm itself, we set the anomaly threshold directly by the proportion of anomalous samples in each training dataset. For example, the percentage of outliers in the training set of Dataset1 is 10%, then the top 10% of samples with the highest abnormal scores given by the detection method will be marked as abnormal and the smallest score among them is set as the threshold for test set. However, in practical the proportion of ab-



Fig. 10. Features series generated by random kernels and anomaly-sensitivity rank

the method is capable of evaluating anomaly sensitivity. Although there might be a few series ranked in a wrong position, since the object of subsequent anomaly detection is a pack of several feature series with high ranks, which impact on the final result can be ignored.

Dataset5 and Dataset7 has the lowest degradation rate in the simulation thus the outliers are similar to the normal data. We assess the performance in separating abnormal and normal data of the proposed method by using the t-SNE to visualize the low-dimensional distribution of the original data and the top 5 generated anomaly-sensitive data in these two datasets, and the results are shown in Fig. 11.

As depicted in Fig. 11(a)(c), the distributions of normal and abnormal data in the original data are overlapped and hard to separate. After processing of random convolution kernel and anomaly sensitivity selection, the normal and abnormal data are clearly separated, as shown in Fig. 11(b)(d), indicating that feature series obtained are with high sensitivity to abnormalities.

#### 3.3.2. Parameters impact on the detection result

In the second part of the experiment, we analyze the effect of the key parameters of the proposed method on the detection result, as shown
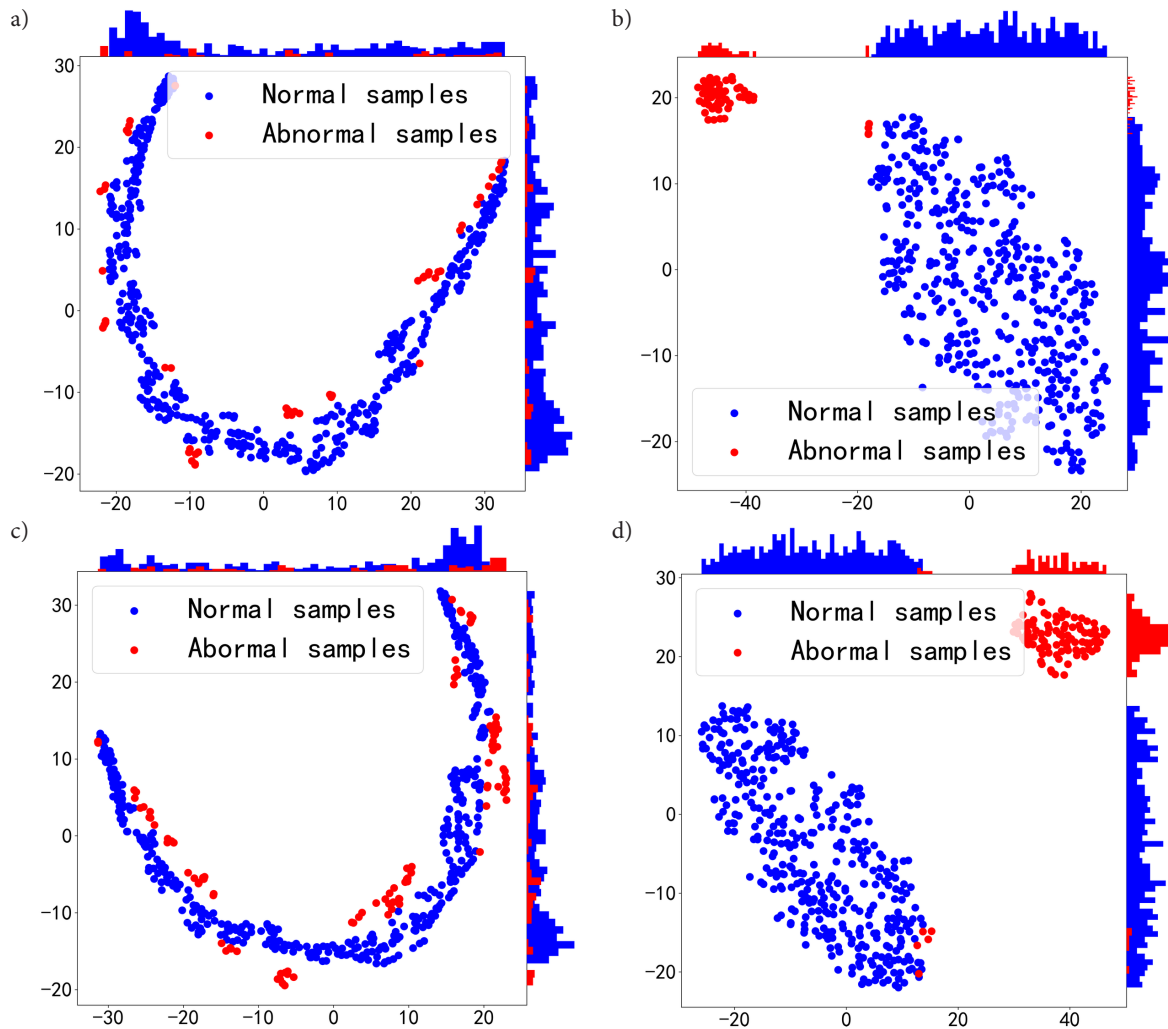
*Fig. 11. Comparison on low dimensional distribution of original data with selected generated series data: a) Dataset5 original distribution, b) Dataset5 selected series distribution, c) Dataset7 original distribution, d) Dataset7 selected series distribution*

in Fig. 12. In the test to analyze the impact of the number of convolutional kernels on the detection results, only the top 5 feature series are selected as the detection object and the reason is that using fewer features makes it more evident that the change in the detection result comes from the anomalous expressing ability of the generated series. In Fig. 12(a), as the number of convolutional kernels increases, the values of  and  gradually raise and the variance gradually decreases until reaching a steady state after kernels exceed 100. From this, it

can be concluded that increasing the number of kernels can increase the probability of obtaining anomaly-sensitive series and improve the stability of detection result, but the boosting effect will touch a ceiling after a specific number of kernels. Fig. 12(b)(c) provide the analysis of the impact of the number of selected series used in isolation forest on the results and the impact of the proportion of invalid series among the selected series on the results, respectively.

*Table 2. Comparison between proposed method and baselines (AUC)*

| | ECOD | COPOD | ABOD | CBLOF | HBOS | KNN | PCA | OCSVM | Auto Encoder | Ablation Experiment | |
| | | | | | | | | | | Isolation forest | Proposed method |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset1 | 0.9265 | 0.9830 | 0.9119 | 0.7881 | 0.9400 | 0.9875 | 0.9850 | 1.0 | 0.7865 | 0.9751 | 1.0 |
| Dataset2 | 0.7625 | 0.8465 | 0.8525 | 0.9890 | 0.8015 | 0.9395 | 0.8285 | 0.9450 | 0.7780 | 0.9159 | 0.9991 |
| Dataset3 | 0.6715 | 0.6930 | 0.8950 | 0.9919 | 0.6825 | 0.9875 | 0.7160 | 0.9625 | 0.8007 | 0.9239 | 0.9999 |
| Dataset4 | 0.5480 | 0.5300 | 0.9500 | 0.9329 | 0.5675 | 0.9565 | 0.5935 | 0.8830 | 0.8513 | 0.8620 | 0.9807 |
| Dataset5 | 0.5235 | 0.5540 | 0.8985 | 0.7510 | 0.5370 | 0.9420 | 0.5435 | 0.7975 | 0.9768 | 0.7792 | 0.9914 |
| Dataset6 | 0.5369 | 0.5609 | 0.8215 | 0.8880 | 0.5400 | 0.8887 | 0.5683 | 0.7314 | 0.6004 | 0.7884 | 0.9630 |
| Dataset7 | 0.5226 | 0.5078 | 0.8179 | 0.7532 | 0.5167 | 0.9010 | 0.5374 | 0.7375 | 0.8422 | 0.7150 | 0.9687 |
| Dataset8 | 0.9137 | 0.8898 | - | 0.8005 | 0.8041 | 0.9063 | 0.6039 | 0.9156 | - | 0.9140 | 0.9957 |
| Dataset9 | 0.7262 | 0.7206 | - | 0.6566 | 0.7210 | 0.7056 | 0.6634 | 0.7171 | - | 0.7246 | 0.9339 |
| Dataset10 | 0.9056 | 0.8532 | - | 0.8820 | 0.6502 | 0.8966 | 0.6404 | 0.9052 | - | 0.9053 | 0.9983 |
| Dataset11 | 0.6274 | 0.7764 | 0.8862 | 0.8005 | 0.7597 | 0.9178 | 0.5027 | 0.5334 | 0.5768 | 0.9036 | 0.9683 |

*Note: The ABOD and Autoencoder can only be used for multi-dimensional data, thus, there is no result in the Dataset8~ Dataset10 which are one-dimensional dataset*
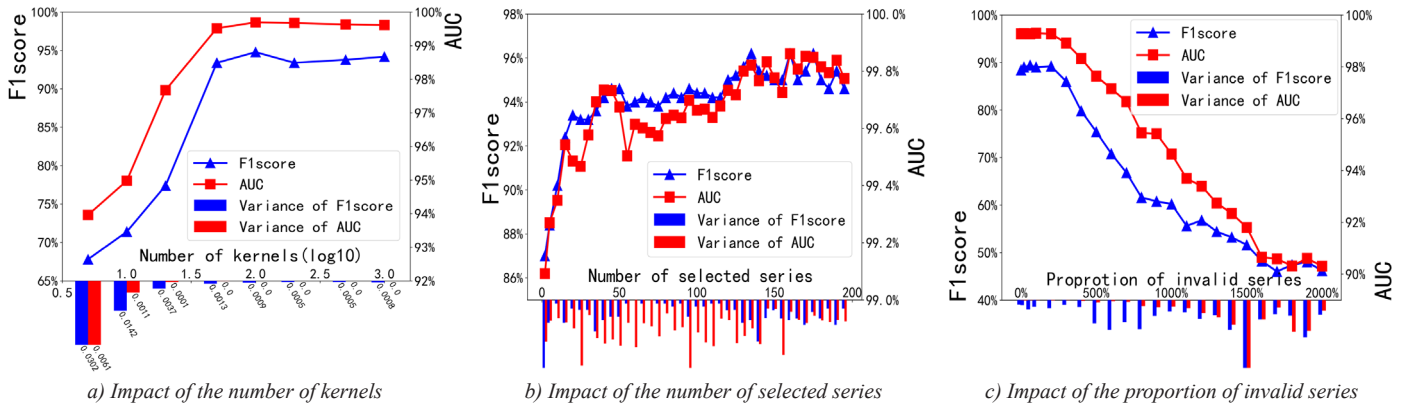
*a) Impact of the number of kernels*     *b) Impact of the number of selected series*     *c) Impact of the proportion of invalid series*

*Fig. 12. Impact of key parameters on the result*

From the Fig. 12(b), it can be found that the detection accuracy raises with the increase of the number of selected series. When the number of selected series exceeds 5, the change rate of and are less than 3.17% and 0.4% respectively, and the variance of both are slightly reduced. It indicates that using only a few series in the top of the anomaly-sensitive rank can significantly distinguish normal data from abnormal, meanwhile the increase in the number of selected series can improve the stabilization of the detection. Therefore, the number of the selected feature series in the study is set to 10 to avoid excessive information loss and feature series redundancy in detection object. Fig. 12(c) record the impact of proportion of invalid series, and the experiment is conducted by combining top 5 series and some invalid series at the bottom of anomaly sensitivity rank to be the detecting object of the isolation forest. As it depicted, the detection result is relatively stable when the proportion of invalid series is in the range of 0%-100%, which is mainly contributed by the robustness of the isolation forest. As the proportion continually raise, drops to the unacceptable points. The main reason is that the massive invalid series cover the abnormally sensitive series, distract the tree-building process of the isolation forest, leading to the almost useless result. Combining the above analysis can conclude that the proposed method has strong robustness.

### 3.3.3. Comparing with other methods

We compare the proposed method with the commonly used anomaly detection algorithms, and the ablation experiment is also conducted to verify the effect of the random convolution kernels. The programs of common anomaly detection algorithms are provided by the anomaly detection package named pyod (https://github.com/yzhao062/pyod). Among them, we fine-tune the structure and parameters of the Autoencoder to ensure the net can fit the training data and reach a steady state.

- Table 2 records the of methods applied in each dataset. The best records are marked in bold.
- Fig. 13 records ROC curve of each dataset, which are plotted base on the anomaly scores provided by each method.
- Fig. 14 record the impact of the rate of the degradation on the result. The degradation rate of Dataset1~Dataset5 varying from 4% to 0.25% with the difficulty of detecting abnormal gradually increase.

As shown in Table 2, the ABOD, CBLOF and KNN achieve relatively high on Datasets1~7, exceeding the isolation forest, while on Dataset8~11, the result of ECOD, OCSVM and isolation forest are close. The performance gap indicates that different methods have different applicability on different data types. But the proposed method benefits from the cross-dimensional and multi-scale feature extraction ability of the random convolution kernel, which can effectively process data of different frequencies, and achieve superior detection results on both data types. It can also be found in the ablation experiment that the combination of the random convolution kernel greatly boosts the detection effect of the isolation forest.

The ROC curves depicted in Fig. 13 visualizes the performance in separating the abnormal from the normal by using the anomaly scores of each method. It can be found that under almost all of the false positive rates the accuracy of proposed method surpasses the others. (a) Impact on precision

Fig. 14 shows that most anomaly detection algorithms gradually lose its ability to identify outlies as the degradation rate decreases. Among them, the precision of KNN on some datasets is close to the proposed method but the recall is inferior, indicating that using Euclidean distance to measure the difference between normal and abnormal of these datasets is not accurate enough. As the dimension increased, the KNN has the risk of distance failure, which will lead to a further decline in detection effect. The performance of ABOD suggests that the perspective of angle is also inappropriate to distinguish outliers. The CBLOF detects abnormalities from data density, and its performance is close to the proposed method when abnormalities are more apparent in Dataset2 and Dataset3, but declines significantly as the degradation rate decreases. The stability of precision and recall of the proposed method outperforms others, and the decline is much smaller than isolation forest used alone, which can prove that the feature series calculated by random convolution kernel and feature selection are more stable in anomaly expression.

## 4. Discussion

The methods performance by training using normal samples only is discussed in this section. It can be found in Table 2 that Autoencoder of DNN method performs poorly among these datasets, which probably caused by the mixture of anomalies in the training set. Therefore, we choose the Autoencoder to the comparison with the proposed method. The validation strategy is as follow: Firstly, training the both methods with pure normal data in training set of Dataset1~7, then testing the model using the complete training dataset to set the anomaly threshold according to its percentage of anomalies, and finally evaluating it on the test dataset. The of the Autoencoder and the proposed method are shown in Table 3, and the best records are marked in bold.

As shown in Table 3, the Autoencoder obtains the priority on six datasets, meanwhile the proposed method is close to its with a maximum fallback of 0.09%. Comparing to the condition that the models trained by abnormal mixed datasets, the performance of Autoencoder decreases significantly while the proposed method only decreases by 3.61% at the maximum. This result confirms the aforementioned analysis that when the training samples are mixed with abnormal data, the DNN models will indiscriminately learn and reconstruct the normal and abnormal data due to its powerful fitting ability, thus causing deviation in representation of the normal samples and leading to the poor performance. From the comparison, we can conclude that the proposed method is less sensitive to the purity of the training data. Considering that the desirable pure training dataset is difficult to ob-
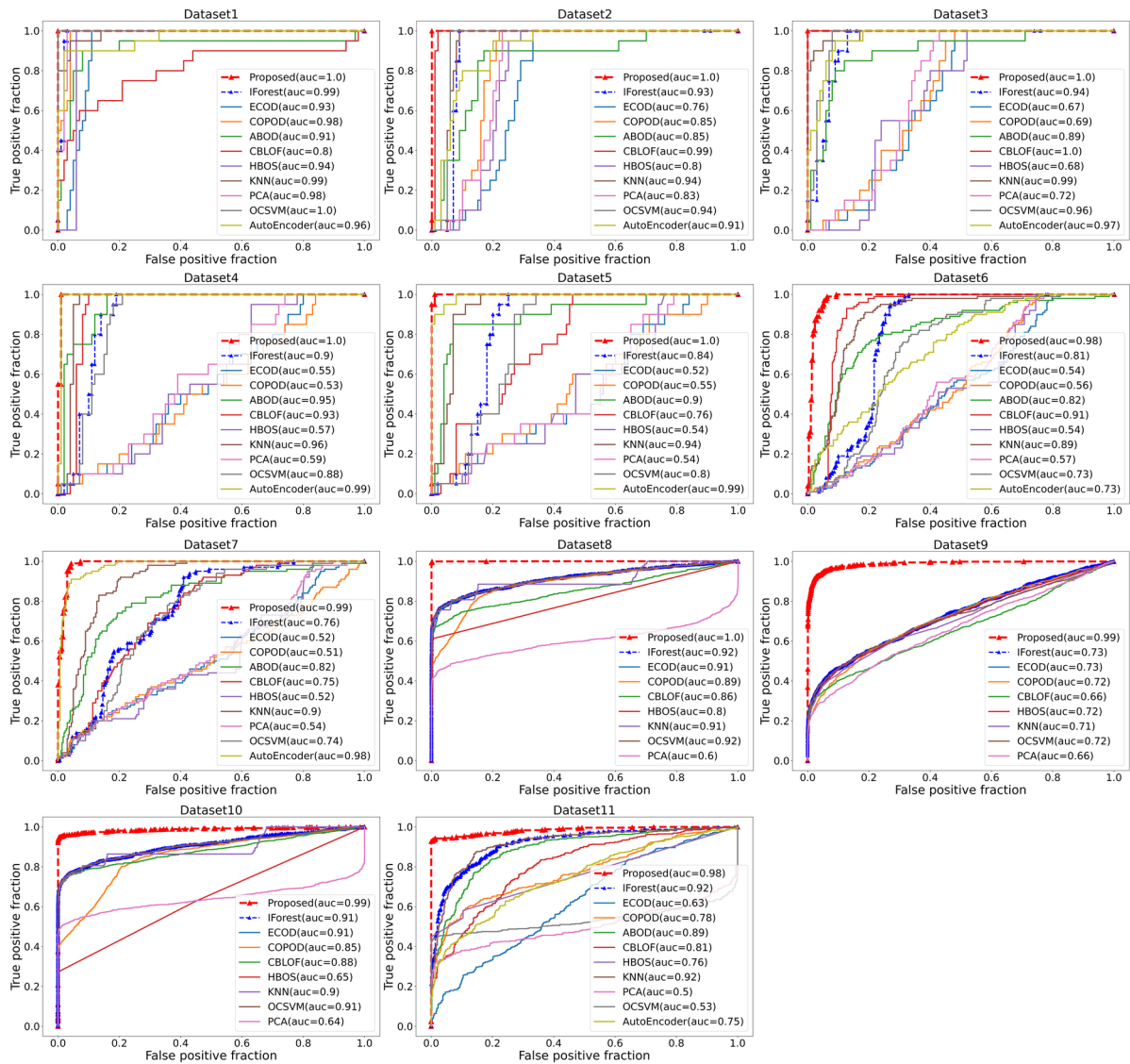
Fig. 13. ROC curve of all methos in each dataset
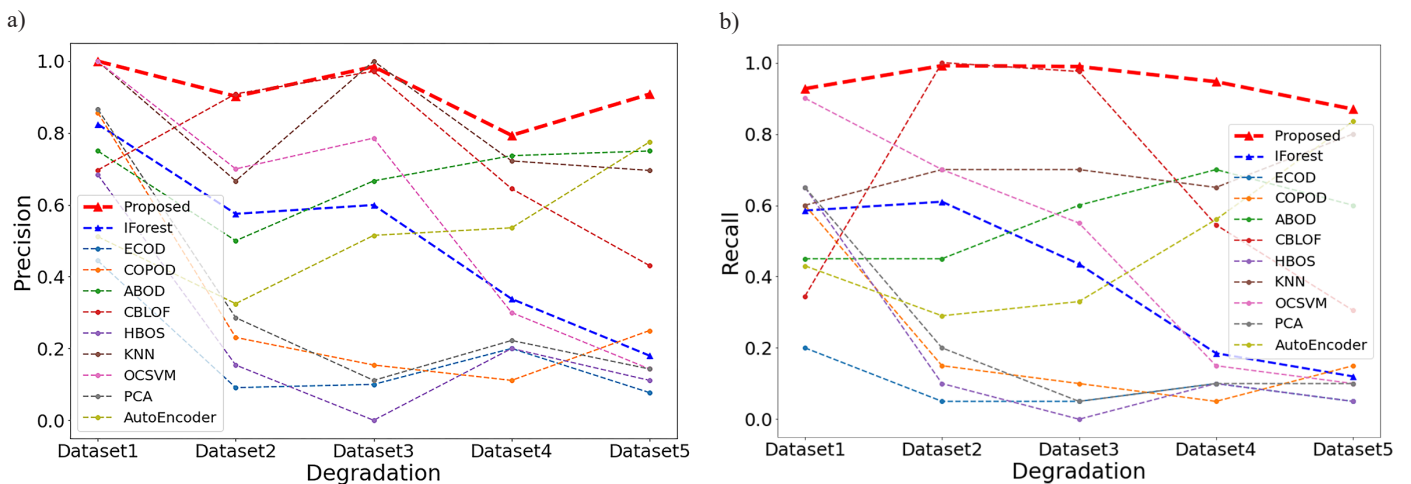


Fig. 14 Impact of the rate of the degradation: a) Impact on precision, b) Impact on recall

tain in practical, the proposed method is more suitable for realistic scenario.

## 4. Conclusions

In this paper, we propose an anomaly detection method combining random convolution kernel and isolation forest. The method consists of three parts: feature series generation with random convolution kernel, feature filtering based on time series decomposing and anomaly

*Table 3. Performance comparison on pure and mixed training dataset (AUC)*

| Datasets | Pure normal training dataset | | Abnormal mixed training dataset | |
|---|---|---|---|---|
| | Autoencoder | Proposed method | Autoencoder | Proposed method |
| Dataset1 | 1.0 | 1.0 | 0.7865 | 1.0 |
| Dataset2 | 1.0 | 1.0 | 0.7780 | 0.9980 |
| Dataset3 | 1.0 | 1.0 | 0.8007 | 0.9997 |
| Dataset4 | 0.9994 | 0.9939 | 0.8513 | 0.9782 |
| Dataset5 | 0.9995 | 0.9993 | 0.9768 | 0.9895 |
| Dataset6 | 1..0 | 0.9991 | 0.6004 | 0.9630 |
| Dataset7 | 0.9973 | 0.9988 | 0.8422 | 0.9687 |

data conditions. The main contributions of the research are that: (1) We apply the concept of random convolution kernel to the anomaly detection task and established the initialization strategy of kernel parameters. (2) We propose a feature series selection method based on time series decomposing, and achieve automatic anomaly-sensitive feature series generation by combining it with random convolution kernel. In the experiment the proposed method outperforms the other commonly used methods on different types of data, providing a new solution to the unsupervised anomaly detection problem.

detection by isolation forests. The first two part combined as an automatic feature generation method alleviates the reliance of manual feature engineering, and the generated anomaly sensitive feature series enhances the isolation forest performance the on anomaly-mixed
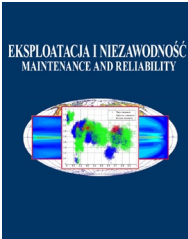
# References

1. Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013; 35(8): 1798–1828, https://doi.org/10.1109/TPAMI.2013.50.
2. Calheiros R N, Ramamohanarao K, Buyya R et al. On the effectiveness of isolation-based anomaly detection in cloud data centers: On the effectiveness of isolation-based anomaly detection in cloud data centers. Concurrency and Computation: Practice and Experience 2017; 29(18): e4169, https://doi.org/10.1002/cpe.4169.
3. Chalapathy R, Chawla S. Deep Learning for Anomaly Detection: A Survey. 2019. http://arxiv.org/abs/1901.03407
4. Cheng Z, Wang S, Zhang P et al. Improved autoencoder for unsupervised anomaly detection. International Journal of Intelligent Systems 2021; 36(12): 7103–7125, https://doi.org/10.1002/int.22582.
5. Dempster A, Petitjean F, Webb G I. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. Data Mining and Knowledge Discovery 2020; 34(5): 1454–1495, https://doi.org/10.1007/s10618-020-00701-z.
6. Dempster A, Schmidt D F, Webb G I. MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Virtual Event Singapore, ACM: 2021: 248–257, https://doi.org/10.1145/3447548.3467231.
7. Guo K, Liu D, Peng Y, Peng X. Data-Driven Anomaly Detection Using OCSVM with Boundary Optimzation. 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, IEEE: 2018: 244–248, https://doi.org/10.1109/PHM-Chongqing.2018.00048.
8. Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks. Science 2006; 313(5786): 504–507, https://doi.org/10.1126/science.1127647.
9. Jahromi A F, Hajiloei M, Dehghani Y, Lahoninezhad S. Improved subspace-based and angle-based outlier detections for fuzzy datasets with a real case study. Journal of Intelligent & Fuzzy Systems 2022; 42(6): 5471–5481, https://doi.org/10.3233/JIFS-211955.
10. Jimenez A, Raj B. Time Signal Classification Using Random Convolutional Features. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, IEEE: 2019: 3592–3596, https://doi.org/10.1109/ICASSP.2019.8682489.
11. Kingma D P, Welling M. Auto-Encoding Variational Bayes. 2014. http://arxiv.org/abs/1312.6114
12. Lei Z, Zhu L, Fang Y et al. Anomaly detection of bridge health monitoring data based on KNN algorithm. Journal of Intelligent & Fuzzy Systems 2020; 39(4): 5243–5252, https://doi.org/10.3233/JIFS-189009.
13. Li Y, Wang Y, Ma X. Variational autoencoder-based outlier detection for high-dimensional data. Intelligent Data Analysis 2019; 23(5): 991–1002, https://doi.org/10.3233/IDA-184240.
14. Liu F T, Ting K M, Zhou Z-H. Isolation Forest. 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, IEEE: 2008: 413–422, https://doi.org/10.1109/ICDM.2008.17.
15. Mensi A, Bicego M. Enhanced anomaly scores for isolation forests. Pattern Recognition 2021; 120: 108115, https://doi.org/10.1016/j.patcog.2021.108115.
16. Puggini L, McLoone S. An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data. Engineering Applications of Artificial Intelligence 2018; 67: 126–135, https://doi.org/10.1016/j.engappai.2017.09.021.
17. Saxe A M, Koh P W, Chen Z et al. On Random Weights and Unsupervised Feature Learning. International Conference on Machine Learning (ICML 2011), Bellevue, Washington, USA, 2011.
18. Smith W A, Randall R B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mechanical Systems and Signal Processing 2015; 64–65: 100–131, https://doi.org/10.1016/j.ymssp.2015.04.021.
19. Tan C W, Dempster A, Bergmeir C, Webb G I. MultiRocket: multiple pooling operators and transformations for fast and effective time series classification. Data Mining and Knowledge Discovery 2022. doi:10.1007/s10618-022-00844-1, https://doi.org/10.1007/s10618-022-00844-1.
20. Tian H D, Khoa N, Anaissi A et al. Concept Drift Adaption for Online Anomaly Detection in Structural Health Monitoring. PROCEEDINGS

OF THE 28TH ACM INTERNATIONAL CONFERENCE ON INFORMATION & KNOWLEDGE MANAGEMENT (CIKM '19) 2019: 2813–2821, https://doi.org/10.1145/3357384.3357816.

21. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th international conference on Machine learning - ICML '08, Helsinki, Finland, ACM Press: 2008: 1096–1103, https://doi.org/10.1145/1390156.1390294.

22. Zhang L, Lin J, Karim R. Adaptive kernel density-based anomaly detection for nonlinear systems. Knowledge-Based Systems 2018; 139: 50–63, https://doi.org/10.1016/j.knosys.2017.10.009.

23. Zhao J, Itti L. Decomposing time series with application to temporal segmentation. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, IEEE: 2016: 1–9, https://doi.org/10.1109/WACV.2016.7477722.

24. Zhong S, Fu S, Lin L et al. A novel unsupervised anomaly detection for gas turbine using Isolation Forest. 2019 IEEE International Conference on Prognostics and Health Management (ICPHM), San Francisco, CA, USA, IEEE: 2019: 1–6, https://doi.org/10.1109/ICPHM.2019.8819409.

25. Zong B, Song Q, Min M R et al. Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection. ICLR, 2018.

# Selective maintenance optimization with stochastic break duration based on reinforcement learning

Indexed by:
Web of Science Group

Yilai Liu[a,b], Xinbo Qian[a,c,*]

[a] Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan, China
[b] Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan , China
[c] Precision Manufacturing Institute, Wuhan University of Science and Technology, Wuhan, China

| Highlights | Abstract |
|---|---|
| • Selective maintenance model with stochastic break duration is proposed.<br><br>• Reinforcement learning(RL) method is applied to selective maintenance model.<br><br>• The advantages of considering stochastic break duration and RL are analysed. | For industrial and military applications, a sequence of missions would be performed with a limited break between two adjacent missions. To improve the system reliability, selective maintenance may be performed on components during the break. Most studies on selective maintenance generally use minimal repair and replacement as maintenance actions while break duration is assumed to be deterministic. However, in practical engineering, many maintenance actions are imperfect maintenance, and the break duration is stochastic due to environmental and other factors. Therefore, a selective maintenance optimization model is proposed with imperfect maintenance for stochastic break duration. The model is aimed to maximize the reliability of system successfully completing the next mission. The reinforcement learning(RL) method is applied to optimally select maintenance actions for selected components. The proposed model and the advantages of the RL are verified by three case studies verify. |

**Keywords**

selective maintenance; stochastic break duration; imperfect maintenance; reinforcement learning.

## 1. Introduction

Maintenance can restore aging systems to better condition and extend the system's life and is a crucial factor affecting industrial, military, and aerospace development. In many industrial and military applications, systems usually perform a sequence of missions with a finite break between two adjacent missions. Maintenance of the equipment is essential [39]. Maintenance actions can be performed during the break to guarantee the reliability of system successfully completing the next mission during subsequent production or missions. However, due to limited maintenance resources (time, manpower, spare parts, etc.), it may be impossible to perform maintenance on all components. Therefore, only some of the system components can be maintained during the limited break so that the reliability of the system meets the requirements or is maximized to complete subsequent production or missions successfully. In this case, managers need to decide which components to maintain based on the actual situation, rather than always following a fixed schedule for all components [4]. This maintenance strategy is known as selective maintenance.

Selective maintenance is vital in balancing limited maintenance resources with system performance. Rice et al [37] first introduced the

selective maintenance problem by considering only one maintenance action to replace the failed components, assuming that all components are identical and that the lifetime follows an exponential distribution. Since 1998, many researchers have studied selective maintenance. Cassady et al [7] extended the model in Rice et al [37], assuming that the component life obeys Weibull distribution and considers three maintenance actions: minimal repair, preventive replacement and corrective replacement, and takes the total maintenance time as the constraint to maximize the reliability of the system successfully completing the next mission. Rajagopalan et al [36], an improved enumeration method was used to solve the selective maintenance problem with the constraints of total maintenance time and cost and the objective function of maximizing the next mission reliability of the system, which improves computational efficiency. Xu et al [44] further improved the enumeration method based on Rajagopalan et al [36], significantly reducing the number of candidate solutions and improving computational efficiency. When the scale of the system is large, the number of different components of the system and the number of maintenance actions increase. The enumeration method does not apply to selective maintenance problems with large and complex solution spaces when the number of feasible solutions grows exponentially. Lust et al [27]

---

(*) Corresponding author.
E-mail addresses: Y. Liu (ORCID: 0000-0003-2370-9250): 1454786628@qq.com, X. Qian: xinboqian@wust.edu.cn

studied a multi-component system with a series-parallel general structure and proposed a selective maintenance optimization method based on a heuristic algorithm, which has a better solution efficiency and promotes the optimization of the selective maintenance model. For the time being, only three maintenance actions were considered in the above study, and imperfect maintenance was not considered. However, in reality, imperfect maintenance is more realistic in engineering. Therefore, some researchers gradually considered imperfect maintenance [11, 19, 33]. Pandey et al [33], it was proposed that introducing imperfect maintenance can describe the decision problem more accurately and was more in line with practical applications. Among other works, Diallo et al [11] was first to propose a selective maintenance model for large k-out-of-n systems and an improved two-stage approach to improve computational efficiency, and Khatab et al [19] considered the stochastic of maintenance action quality.

Various uncertainties are inevitable in maintenance decisions of engineering systems, and ignoring these potential uncertainties may lead to inefficient optimization decisions, and the system may face the risk of not completing the mission [49]. Current studies on selective maintenance problems assume mainly deterministic values for break duration. In practice, unexpected events may lead to early termination or continuation of the mission, resulting in an increase or decrease in the break duration. For example, delays in flight departures or ship departures due to weather can lead to increased break duration. In the military, the time of the next mission start cannot be accurately determined, so the break between two adjacent missions is also uncertain. In similar situations, the break duration should be a random variable that obeys an appropriate distribution. Other literature [17, 18, 20, 25] considered the stochastic break duration with the decision goal of reducing maintenance resources. Zhao et al [48] considered stochastic mission time and multiple maintenance workers with different capacities. However, in many engineering practices, when maintenance resources cost, time, and manpower are limited, selective maintenance problems are often aimed at maximizing the reliability of system successfully completing the next mission rather than minimizing maintenance resources [6].

In recent years, selective maintenance optimization problems have been intensively studied. With the increasing complexity of selective maintenance optimization models, some advanced intelligent optimization algorithms, such as particle swarm algorithm [28], artificial bee colony [10], ant colony algorithm [25, 40], and genetic algorithm [5, 13, 43] have been widely adopted. As the scale of the system becomes larger, the factors considered become more comprehensive. Therefore the solution of large-scale selective maintenance decision-making problems poses new challenges, and the efficiency of optimization algorithms and global optimization capabilities need to be further improved [8]. Reinforcement learning belongs to machine learning methods, which have attracted more and more attention from researchers in solving decision problems [22]. Some reinforcement learning algorithms can be explored to obtain immediate payoffs and then select appropriate strategies to obtain the optimal solution of the model [14]. In recent years RL is effective in decision performance and computational efficiency. Other heuristic solution methods continuously iterate the algorithm randomly on the feasible solution space until the best solution is obtained or the number of iterations reaches the maximum. It may lead to problems such as complex model solving and limited computational efficiency [39]. In contrast, in RL, the agent continuously learns from each iteration and, in return, improves the result of the next iteration based on the previous one, and the optimal solution converges faster, thus improving the computational efficiency [31]. Although RL methods have been successfully applied to different problems and have significant advantages, they have not yet attracted sufficient attention in selective maintenance optimization.

In summary, this paper proposes a new selective maintenance model that considers the stochastic break duration. To maximize the reliability of the system successfully completing the next mission, each component has multiple optional maintenance actions, including

minimal repair, imperfect maintenance, and replacement. The selective maintenance decision problem is modeled as a Markov decision process (MDP), and a RL approach is proposed to solve the model.

The rest of this paper is presented as follows. Section 2 is the related work about RL in other maintenance areas. Section 3 is the problem description and basic assumptions and describes the evaluation of imperfect maintenance and system reliability based on the Kijima type II model. Section 4 presents the selective maintenance model and the solution method of this paper. Three case studies are given in Section 5 to verify the accuracy of the model and the validity of the method. Finally, a summary and an outlook for future works are given in Section 6.

## 2. Related work

The main objective of selective maintenance optimization is to maximize the reliability of the system successfully completing the next mission. As the number of components and optional maintenance actions increases, traditional solution methods may have the problems of difficult model solving and limited solving efficiency. In recent years RL has become an effective method for solving complex decision problems. RL has been applied to solve various decision problems such as scheduling, manufacturing and maintenance. In this section, we briefly review the work of RL in other maintenance areas and selective maintenance.

Nooshin et al [47] proposed a dynamic condition-based maintenance (CBM) model that considers components subject to degradation and random shocks. Instead of discretizing the degradation state, the exact degradation level was considered as the state of the system, and finally deep reinforcement learning (DRL) was used to derive the optimal maintenance action for each degradation level. Mahmoodzadeh et al [29] studied CBM of dry gas pipeline and proposed a test bench to simulate pipeline corrosion while interacting with the RL to adjust the maintenance action and minimize maintenance costs. Peng et al [35] considered that RL can be effective in solving MDP problems with large state spaces, and models the CBM problem as a discrete-time continuous-state MDP rather than a discrete system with deterioration conditions. An RL algorithm was proposed to minimize the long-run average cost, and a Gaussian process regression function was used to model the state transfer and the value functions of the states in RL. Stephane et al [2] used MDP to model preventive maintenance for equipment consisting of multi-non-identical components with different probability distributions of failure times, which has the advantage of not requiring to estimate the main parameters of the model. Finally, the optimal strategy was solved using Monte Carlo reinforcement learning, which was not restricted by mathematical formulas. Huang et al [15] formulated the preventive maintenance(PM) decision for serial production lines as an MDP framework, considered the system production loss, and used DRL to solve the optimization model.

In addition to the above maintenance optimization, there are also some applications of RL for decision optimization problems. Andriotis et al [1] considered that in engineering systems management decisions can be made with MDPs or partially observable MDPs. For large multi-component systems, the number of system states and actions grows exponentially with the number of components, and it is difficult to characterize the environmental dynamics of the whole system, which can only be obtained by expensive numerical simulators. Therefore, a DRL algorithm was proposed to obtain an effective life cycle strategy. Ruan et al [38] studied the aircraft maintenance routing problem, where the objective was to generate maintenance feasible optimal routes for each aircraft under the constraints of maximum flight time, limitation on the number of takeoffs between two consecutive maintenance checks, and labor capacity maintenance. An RL approach was developed to solve the problem, by comparing with common optimization software, RL can solve the problem quickly and efficiently. Panagiotis et al [34] studied the maintenance

problem of a stochastic production/inventory system producing a single type of product, maximizing the total profit of the system when maintenance and repair duration as random variables. The commonly used dynamic programming methods were not suitable for solving the problem discussed in this paper, so a RL approach was proposed. Hu et al [16] proposed an RL framework with extreme learning machine optimization algorithm for aircraft life cycle maintenance, considering engine lifetime, performance degradation and random failures. It was found that the RL-driven maintenance strategy have a advantage compare to the CM, schedule Maintenance and prognostics and health driven strategies.

According to the above reviews, RL is effective in decision performance and computational efficiency. To the best of our knowledge, the proposed approach is novel in dealing with the single-mission selective maintenance problem.

## 3. Problem statements

### 3.1. Selective maintenance problem description for multi-component systems

In many military and industrial environments, systems are scheduled to perform multiple sequential missions with a finite break between two adjacent missions. Maintenance actions can be performed during the break to restore the aging system to a better condition for subsequent missions. However, due to the constraints of maintenance resources such as time and manpower, it may not be possible to perform maintenance on all components and select only some for maintenance depending on the situation. The basic process of selective maintenance decisions with stochastic break duration is shown in Fig. 1. As shown in Fig. 1, scenario 2 has a longer break compared to scenario 1, and only maintenance action 3 is not completed. And in scenario 1, both maintenance actions 2 and 3 are not completed.
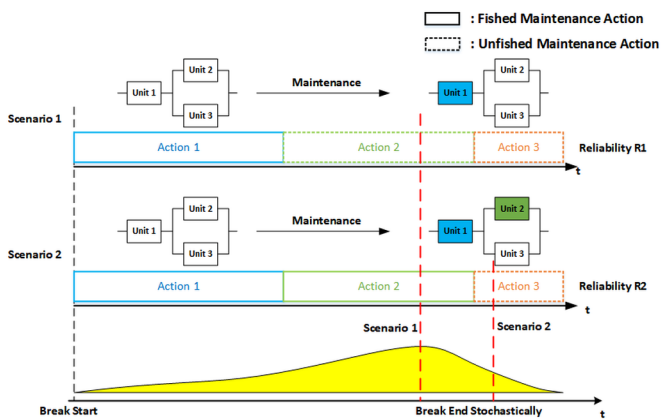


*Fig. 1. Schematic diagram of selective maintenance decisions with stochastic break duration*

To describe the selective maintenance problem, the basic assumptions are as follows:

(1) Assume a series-parallel system, and the system consists of $i$ ($i$ = 1, 2, … , m) independent subsystems in series, and each subsystem i consists of $j$ ($j$ = 1, 2, … , n) independent components $C_{ij}$ in parallel, $i$ and $j$ denote the location of the components in the system. It is assumed that the components have only one failure mode, and the component's states are either failure or functioning. Here the variables $X_{\text{break,s}}(k)$ and $X_{\text{break,e}}(k)$ are used to denote the state of component $C_{ij}$ at the beginning of $k$th break and the end of $k$th break, respectively, i.e., the state of component $C_{ij}$ at the beginning of $k$th break can be expressed as:

$$X_{\text{break,s}}(k) = \begin{cases} 1, & \text{if } C_{ij} \text{ functioning at the beginning of } k\text{th break} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The state of the component $C_{ij}$ at the end of the kth break can be expressed as:

$$X_{\text{break,e}}(k) = \begin{cases} 1, & \text{if } C_{ij} \text{ functioning at the end of } k\text{th break} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

(1) Assume that during the break, the set of optional maintenance actions for the component is {do nothing(DN), minimal repair(MR), imperfect maintenance(IM), preventive replacement(PR), corrective replacement(CR)}, and the corresponding codes of maintenance actions are shown in Table 1. No maintenance means doing nothing, and no maintenance resources are consumed. The minimal repair can only be performed on failed components, consumes fewer resources, and can restore the failed components to functioning, but it does not change the reliability. The imperfect maintenance effect is between minimal repair and replacement. Preventive replacement can only be performed on functioning components, and corrective replacement can only be performed on failed components. When $X_{\text{break,s}}(k)=0$, the $C_{ij}$ optional maintenance actions are minimal repair, imperfect maintenance, and corrective replacement. When $X_{\text{break,s}}(k)=1$, the $C_{ij}$ op-

*Table 1. Codes of different maintenance action l*

| Maintenance Action | Do Nothing | Minimal Repair | Imperfect Maintenance | Preventive Replacement | Corrective Replacement |
|---|---|---|---|---|---|
| Corresponding code $l$ | 0 | 1 | $2,…, L_{ij}-2$ | $L_{ij}-1$ | $L_{ij}$ |

tional maintenance actions are imperfect maintenance and preventive replacement. Fig. 2 shows the correspondence between maintenance action and component state.
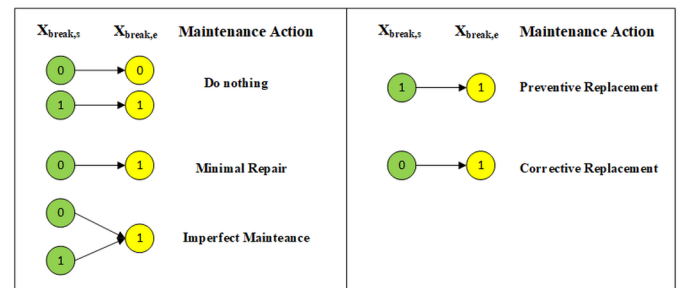


*Fig. 2. Component state changes under different maintenance actions of components*

(2) It is assumed that all maintenance actions can only be performed during the break. If the current maintenance action is not completed by the beginning of next mission, then it is assumed that the maintenance action has no repair effect on the component.

(3) Assume that only two types of maintenance resource constraints, maintenance time and manpower are considered in this paper.

(4) Assume that failure time of the component $C_{ij}$ in the system obeys a two-parameter Weibull distribution.

### 3.2. Stochastic break duration

The break duration is stochastic because unexpected events may lead to early termination or continuation of production or mission such that the break duration decreases or increases randomly. In this study, the break duration $Z_k$ is a random variable that obeys $f(Z_k)$. Therefore, the number of maintenance actions that can be completed during the

break is also uncertain. A binary decision variable $W_{ij}(l)$ is used to indicate whether the component $C_{ij}$ is maintained during the break, which is defined as follows:

$$W_{ij}(l)=\begin{cases}1, \text{ if the maintenance action } l \text{ for component } C_{ij} \\ 0, \text{otherwise}\end{cases} \quad (3)$$

The maintenance time consumed during the break can be expressed as:

$$T = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=0}^{L} t_{ij}(l)w_{ij}(l) \quad (4)$$

where $t_{ij}(l)$ is the maintenance time of completing maintenance action $l$.

The break duration $Z_k$ as a random variable obeying $f(Z_k)$, it is required that the probability of completing the maintenance action during the break should be greater than or equal to a predetermined critical value $\tau$, the range of $\tau$ values is (0,1], which is expressed as follows:

$$\Pr\left(T \leq Z_k\right)\geq\tau \quad (5)$$

### 3.3. Evaluating the reliability of system successfully completing the next mission

There are many imperfect maintenance models about imperfect maintenance action [3, 23, 32, 30, 41, 42]. In this paper, we use the Kijima type II model to represent the maintenance effect of maintenance action by age reduction. The effective age of the component can be expressed as:

$$A_{ij}(k+1) = b_{ij}(l)B_{ij}(k) \quad (6)$$

where $A_{ij}(k+1)$ is the effective age of component $C_{ij}$ after taking maintenance action $l$ during the $k$th break. $B_{ij}(k)$ is the effective age of component $C_{ij}$ at the beginning of the $k$th break. $b_{ij}(l)$ ($0 \leq b_{ij}(l) \leq 1$) is the age reduction factor, which is influenced by the number of maintenance resources invested, the more maintenance time required for the executed maintenance actions, the smaller $b_{ij}(l)$ is, the better the maintenance effect.

Fig. 3 shows the relationship between the maintenance time of the component and effective age after the component is maintained during the break. The age reduction factor $b_{ij}(l)$ can be expressed as:

$$b_{ij}(l) = 1-\left(\frac{t_{ij}(l)}{t_{ij}(L)}\right)^{1/\zeta_l} \quad (7)$$

where $t_{ij}(l)$ is the maintenance time for component $C_{ij}$ to complete maintenance action $l$ within the break. When the state of component $C_{ij}$ is 1, $t_{ij}(L)$ is the time consumed for the preventive replacement of component $C_{ij}$. When the state of component $C_{ij}$ is 0, $t_{ij}(L)$ is the time consumed for corrective replacement of component $C_{ij}$. $\zeta_l$ is a characteristic constant reflecting the relationship between maintenance time and age reduction factor function. When the maintenance action consumes the same time, the larger the $\zeta_l$, the more obvious the maintenance effect.

According to the above effective age model, the conditional survival probability of a component after maintenance can be expressed as [21]:

$$r_{ij}(x)=1-\Pr\left\{Y - A_{ij} \leq x|Y > A_{ij}\right\}=\frac{\Pr\left\{Y > x + A_{ij}\right\}}{\Pr\left\{Y > A_{ij}\right\}} \quad (8)$$

where the random variable $Y$ represents the failure time. If the component is functional at the beginning of the $k$th mission and has an effective age of $A_{ij}$, then $r_{ij}(x)$ represents the probability that the component does not fail at any moment $x$. Since the failure time of component $C_{ij}$ obeys the Weibull distribution, it is functioning at the beginning of the $k$th mission and has an effective age of $A_{ij}(k)$. The conditional survival probability of component $C_{ij}$ at the end of $k$th mission is:
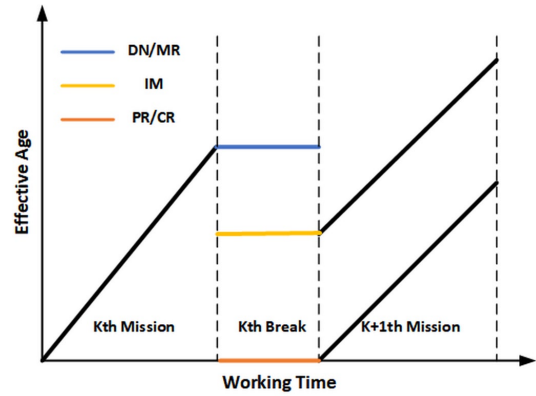


*Fig. 3.  Relationship between component working time and effective age in the Kijima II model*

$$r_{ij}(k) = \exp\left[-\left(\frac{U(k) + A_{ij}(k)}{\eta_{ij}}\right)^{\beta_{ij}} + \left(\frac{A_{ij}(k)}{\eta_{ij}}\right)^{\beta_{ij}}\right] \quad (9)$$

where $U(k)$ is the duration of the $k$th mission. $\eta_{ij}$ is the scale parameter in the Weibull distribution of component $C_{ij}$, $\beta_{ij}$ is the shape parameter in the Weibull distribution of component $C_{ij}$. The reliability $R_{ij}(k)$ of component $C_{ij}$ in the $k$th mission depends on the conditional survival probability $r_{ij}(k)$ and the component state $X_{break,e}(k-1)$ at the end of the $k$th break. The expression for the reliability $R_{ij}(k)$ of component $C_{ij}$ as:

$$R_{ij}(k) = r_{ij}(k)X_{break,e}(k-1) \quad (10)$$

The study in this paper is a complex series-parallel system, i.e. the system consists of subsystems in series and subsystems comprised of components in parallel. The reliability $R_i(k)$ of subsystem $i$ in $k$th mission can be expressed as:

$$R_i(k) = 1-\prod_{j=1}^{n}\left(1 - R_{ij}(k)\right) \quad (11)$$

The reliability $R_{sys}(k)$ of the system in $k$th mission as:

$$R_{sys}(k) = \prod_{i=1}^{m}R_i(k) = \prod_{i=1}^{m}\left(1 - \prod_{j=1}^{n}(1 - R_{ij}(k))\right) \quad (12)$$

where $m$ is the number of subsystems in the system and $n$ is the number of components in the subsystem.

## 4. Selective maintenance model and optimization based on the stochastic break duration

### 4.1. Selective maintenance optimization model

For a system performing sequential missions, using limited maintenance resources in a finite break to maximize the reliability of the system to complete the next mission is the key to maintenance decisions. Assume that the states $X_{\text{break},s}(k)$ and effective age $B_{ij}(k)$ of each component in the system are known at the beginning of the $k$th break. Given the optional maintenance actions of each component, the selective maintenance problem can be described as follows: with limited maintenance time and manpower, select the components to be maintained and their corresponding maintenance action so that the reliability of the system to complete the next mission is maximized. When the break duration $Z_k$ is a random variable and the probability distribution function is known, the selective maintenance decision model can be expressed as:

$$\max_{[w_1, w_2, \cdots, w_n] \in A} R_{sys} = \prod_{i=1}^{m}\left(1 - \prod_{j=1}^{n}\left(1 - R_{ij}\left(k+1\right)\right)\right) \qquad (13)$$

Subject to:

$$p\left(T = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=0}^{L} t_{ij}(l)W_{ij}(l) \le Z_k\right) \ge \tau \qquad (14)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=0}^{L} W_{ij}(l) \le 1 \qquad (15)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=0}^{L} X_{break,e}(k)W_{ij}(l) \le 1 \qquad (16)$$

$$W_{ij}(l) \le 1 - Y_{break,s}(k) \qquad (17)$$

$$X_{break,e}(k) = Y_{break,s}(k) + (1 - Y_{break,s}(k)) \cdot W_{ij}(l) \qquad (18)$$

$$A_{ij}(k+1) = [b_{ij}(l) \cdot W_{ij}(l) + (1 - W_{ij}(l))] \cdot B_{ij}(k) \qquad (19)$$

$$W_{ij}(l), X_{break,e}(k), Y_{break,s}(k) \in \{0,1\}; b_{ij}(l) \in [0,1] \qquad (20)$$

In the above selective maintenance decision model, Eq. (13) is the decision objective to maximize the reliability of system successfully completing the next mission. Eq. (14) is the chance constraint, when the break duration is a random variable, the probability of completing the selected maintenance action is required to be greater than or equal to $\tau$, the range of $\tau$ values is [0,1]. Eq. (15) illustrates that in each break, each maintenance action is selected at most once and can only be performed on one component. Eq. (16) shows that in each break, a component that is selected for maintenance can perform at most one maintenance action. Eq. (17) shows that minimal repair can only be performed on the failed component. Eq. (18) is used to update the state of the component $C_{ij}$, for example, when the component $C_{ij}$ state $X_{\text{break},s}(k)=0$ at the beginning of the $k$th break, after maintenance i.e. $W_{ij}(l)=1$, the component $C_{ij}$ state $X_{\text{break},e}(k)=1$. Eq. (19) is used to update the effective age of the component $C_{ij}$, for example, when the component $C_{ij}$ after maintenance i.e. $W_{ij}(l)=1$, then $A_{ij}(k+1)=b_{ij}(l) \cdot B_{ij}(k)$. When the component $C_{ij}$ does not maintenance $W_{ij}(l)=0$, then $A_{ij}(k+1)=B_{ij}(k)$.

### 4.2. The reinforcement learning solution method for selective maintenance optimization
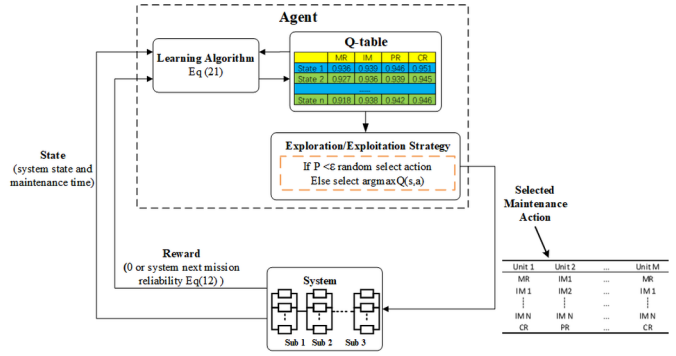


*Fig. 4. Reinforcement learning framework for selective maintenance optimization*

In this study, a reinforcement learning(RL) based framework is used to describe the selective maintenance decision process using MDP and solved using the Q-learning algorithm. According to this framework, the decision agent interacts with the system and selects a maintenance action at a specific time (decision period) to maximize the decision goal. The described framework is shown in Fig. 4. In MDP, the main factors that determine the decision process include the transfer law of states in the system and the maintenance action scheme. The interaction of these two factors leads to a particular reward for the decision-maker, usually represented by an objective function. MDP is an extension of the Markov chain, and its state space, action space, and reward are described as follows:

State space $S$: It defines a finite two-dimensional state space, each state represents the state of the system at a decision moment and the total maintenance time. The state space can be expressed as $S=\{X_{ij}; T\}$, where $X_{ij}$ consist of the states of all the components in the system, the component state is binary variables, and $T$ is the total maintenance time. If the system consists of 5 components, the state space at a decision moment can be expressed as $S= \{0_{1,1}, 1_{1,2}, 0_{2,1}, 0_{2,2}, 1_{2,3}; 0.5\}$, where $0_{1,1}$ represents component $C_{1,1}$ in failed, $1_{1,2}$ represents component $C_{1,2}$ in functioning, and 0.5 represents the total maintenance time. The RL agent moves from the initial state to the terminated state and assigns an ordinal number to each state.

Action space $A$: The action space consists of optional maintenance actions for all components, which can be expressed as A=$\{l_{ij}\}$, $l=\{DN_{ij}, MR_{ij}, IM1_{ij},..., IMn_{ij}, PR_{ij}, CR_{ij}\}$. Given the current state, the agent can select an action from the action space. By judging whether the selected action meets the constraint, the punishment or reward is obtained in turn. It indicates which actions the agent can choose for each observed state. Given the current state of the system, if the agent is not terminated state, any action in action space can be selected.

Reward $R$: Rewards reflect the aptness of the RL agent for the current maintenance action, so here the reward function is defined as the objective function. The objective function of this paper is to maximize the reliability of the system successfully completing the next mission. In this paper, a negative reward is used when the maintenance action selected by the agent does not meet the constraints. When the maintenance action selected by the agent satisfies the constraints and is not the terminated state, 0 is used as a reward. When the maintenance action selected by the agent satisfies the constraints and is the terminated state, the Eq. (13) is used as a reward.

RL is a simulation-based dynamic programming algorithm mainly used to solve Markov decision problems and is an intelligent agent learning optimal control strategy. Compared with traditional dynamic programming, the RL approach does not require a state transfer probability matrix and avoids dynamic programming modeling dimen-

sional disaster [46]. The state space size of this problem is $2^N$, and the action space size is $L^N$, where $N$ is the total number of components in the system. The Q-learning algorithm is one of the more commonly used RL algorithms. The optimal policy is derived by constructing a table of state-maintenance action $Q$. The Q-learning algorithms have been shown to eventually reach a convergence condition for each state through continuous learning in a stochastic environment [45]. In this study, the selective maintenance decision optimization problem is modeled as an MDP. The Q-learning algorithm in RL is used to solve it to obtain the optimal maintenance policy, as follows:

Step1: Initialize $Q(s,a)=0$, $Q$ value table is a list of rows, the value of the $n$th row $m$th column represents the value of the action of m maintenance action in the state of $S_n$, set the maximum number of cycles max_episode.

Step2: Initialize the state $\mathbf{S}$ at the beginning of each cycle, the state of each component after the end of the $k$th mission and the current total maintenance time $\mathbf{T}$=0.

Step3: Select the maintenance action $w_n$ according to the $\varepsilon$-greedy policy and get the reward r. Update the Q-value table using the above reward according to Eq. (21) .

Step4: Update the state $\mathbf{S}$. Use Eq. (14) to determine whether the state reaches the termination state. If not, repeat the above steps from step 2.

Step5: When the number of cycles equals max_episode, stop the cycle to get the final Q-value table.

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left( r + \gamma \max Q(s',a') - Q(s,a) \right) \qquad (21)$$

where $\alpha$ $(0 < \alpha < 1)$ is the learning rate and $\gamma$ $(0 < \gamma < 1)$ is the discount factor. The flow chart of the algorithm is shown in Fig. 5.
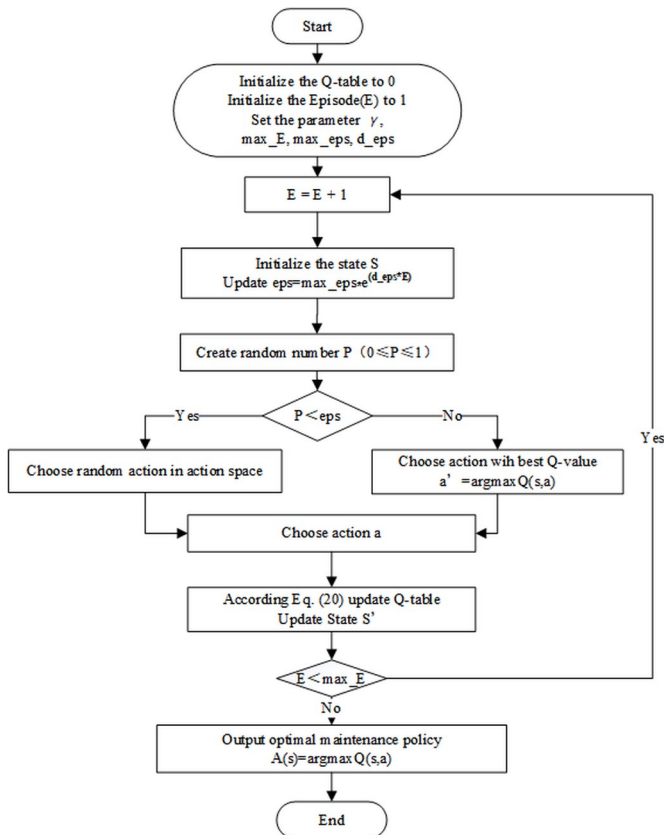


*Fig. 5. Q-learning algorithm flowchart*

In RL, exploration and exploitation are the two core problems. The decaying $\varepsilon$-greedy policy is used here for the agent to learn a better policy. The $\varepsilon$-greedy policy is to select the current optimal action(a'=argmaxQ(s,a)) with probability 1 - $\varepsilon$ , and randomly select the action among all available actions with probability $\varepsilon$. The conventional $\varepsilon$-greedy strategy constantly explores the action space with the same probability of $\varepsilon$. When $\varepsilon$ is small, the exploration is not thorough enough and may obtain the optimal local strategy. When $\varepsilon$ is large, the agent may have long explored the optimal strategy but will continue to explore it, resulting in slow convergence. Therefore, the decaying $\varepsilon$-greedy strategy is used here, respectively, the agent starts exploring with a larger $\varepsilon$ and gradually decreases $\varepsilon$ as the number of iterations increases, and the iteration formula is:

$$\varepsilon = \min\_eps + (\max\_eps - \min\_eps) \cdot e^{d\_eps \cdot E} \qquad (22)$$

## 5. Case study

Three cases are given to test performance of the model and proposed method. The first case is a hydraulic system that is more typical of a real application, in which the key components are analyzed. The superiority of RL and the difference between the stochastic and deterministic break duration are analyzed. The second case is a two-stage 5-component system in which the superiority of RL is verified by comparing the results with other literature. Then, the difference between the stochastic and deterministic break duration is analyzed to illustrate the impact of the stochastic break duration on the system reliability. Due to the redundancy of this case system compared to the first case, the sensitivity analysis of the component parameters is performed here. The third example is a five-stage 14-component coal transportation system, where the performance of the algorithm is compared and the difference between stochastic and deterministic break duration is analyzed. The impact of stochastic on system reliability and the effectiveness of RL for larger scale complex systems are further verified.

### 5.1. Case 1: Hydraulic tension systems

A hydraulic tension system is known to consist of 16 components, which can be divided into two categories of components. The first category is critical components, and the second category is non-critical components. In this paper, the key components pump, solenoid valve, accumulator and cylinder are analyzed, and these four components are connected in series. The parameters of each component are shown in Table 2, where the Weibull distribution shape and scale parameters are derived from the literature [12]. In Table 2, $\zeta$ denotes the characteristic constant of the age regression factor. $\beta$, $\eta$ denote the shape and scale parameters of the Weibull distribution. B($k$) denotes the effective age of the component at the beginning of the kth break. X($k$) denotes the state of each component of the system at the beginning of the kth break. The maintenance actions that can be adopted for each component and their corresponding maintenance times are shown in Table 3, where 0~4 represents the codes of different maintenance actions in order, where the fix is the fixed maintenance time.

#### 5.1.1. Algorithm performance analysis

To further verify the effectiveness of the RL algorithm, a comparison with the genetic algorithm(GA) algorithm used in most of the literature is conducted. Assuming that the $k$th mission is just completed now, the duration $Z_k$ of the break obeys a normal distribution of $N$(0.5, 0.04) with a range of [0.35, 0.65], $\tau$ =0.8, and the duration of the $k$+1th mission $U$=1500 days, all other component parameters are shown in Table 2. Among them, the parameters related to the GA algorithm, the number of populations NP=80, the crossover rate pc=0.8, the variation rate pm=0.05, and the maximum number of iterations iter=1000. The parameters related to Q-learning, the learning rate $\alpha$=0.02, the discount rate $\gamma$=0.5, and the maximum number of iterations iter=10000. Due to the stochastic of the algorithm, 10 sets of simulations were performed for each method to find its optimal

Table 2. Component parameters

| ID | Characteristic constant of age reduction factor | Shape and scale parameters of Weibull distribution | | Effective age of component | Initial state of component |
|---|---|---|---|---|---|
| | $\zeta$ | $\beta$ | $\eta$ | $B(k)$ | $X(k)$ |
| 1 | 2.5 | 2.36 | 1850 | 3500 | 1 |
| 2 | 2.0 | 1.853 | 3657 | 2400 | 1 |
| 3 | 3.0 | 1.46 | 3304 | 4500 | 1 |
| 4 | 3.2 | 2.023 | 3501 | 3500 | 1 |

Table 3. The maintenance time $t_{ij}(l)$ of different maintenance actions $l$ for components (time is in days)

| ID | Maintenance actions $l$ code | | | | | Fix |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | |
| 1 | 0 | 0.0186 | 0.0371 | 0.0557 | 0.0743 | 0.03 |
| 2 | 0 | 0.0443 | 0.0886 | 0.1330 | 0.1770 | 0.03 |
| 3 | 0 | 0.0471 | 0.0943 | 0.1410 | 0.1890 | 0.03 |
| 4 | 0 | 0.0457 | 0.0914 | 0.1370 | 0.1830 | 0.04 |

strategy $A_{best}$, the maintenance time for the optimal strategy $T_{best}$, the average reliability $R_{mean}$, the maximum reliability $R_{best}$, the variance $R_{std}$ and the average running time $\bar{S}$ as the comparison results. To compare the quality of RL and GA solutions, a parameter %QOS is introduced here as a performance metric, $\%QOS=(R_{best}-R_{mean})/R_{best}$. The comparison results and performance metric results are shown in Table 4.

Table 4. Comparison results of the two algorithms(time is in days)

| Method | $A_{best}$ | $T_{best}$ | $R_{best}$ | $R_{mean}$ | $R_{std}$ | $\bar{S}$ | %QOS |
|---|---|---|---|---|---|---|---|
| Q-learning | [4,2,0,4] | 0.446 | 0.9878 | 0.9869 | 0.0016 | 1.15 | 0.09 |
| GA | [3,2,2,2] | 0.460 | 0.9792 | 0.9752 | 0.0032 | 10 | 0.41 |

From Table 4, we can see that the maximum reliability $R_{best}=0.9878$ solved by Q-learning and the total maintenance time $T_{best}=0.446$ days. The the maximum reliability $R_{best}=0.9792$ solved by GA and the total maintenance time $T_{best}=0.46$ days. From the analysis of the results, we can see that the optimal strategy solved by Q learning is better than GA and the reliability is 0.86% higher. The mean and variance of Q-learning results are better than GA in 10 solving results, which indicates that the stability of the Q-learning algorithm is better than GA. Regarding the computation time, running on a computer configured with Intel (R) Core (TM) i5 -6200U CPU @ 2.30GHz, 12G RAM. Although RL has more iterations than GA, the average time spent by GA is 8.85s more than that of RL. Regarding the quality of the obtained solutions, the optimal solution of RL is better than GA, and the average solution of RL deviates from the optimal solution by only 0.09%, while the average solution of GA deviates from the optimal solution by 0.41%. Therefore, the RL algorithm can find higher quality solutions, which verifies the superiority of RL. In order to verify the superiority of this RL, tests on small-scale systems are not sufficient. Section 5.2.1 will further verify the superiority of RL by making comparisons with other literature, and Section 5.3.1 is a comparison of RL with GA in large-scale complex systems.

### 5.1.2. Comparison between stochastic and deterministic of break duration

The difference between the stochastic and deterministic break duration is clarified by substituting the strategy derived from the RL-based deterministic model into the uncertainty model to obtain the reliabil-

ity $R1$. Then comparing the analysis with the reliability $R2$ obtained from the strategy derived from the RL-based uncertainty model, $\Delta R$ calculation schematic is shown in Fig. 6. When the break duration is stochastic, the optimal maintenance policy $A1=[4,2,0,4]$ solved by RL is known from section 5.1.1, and the reliability $R2=0.9878$, and the maintenance time is 0.446 days. When the break duration $Z=0.5$ is a fixed value with all other parameters held constant, the optimal maintenance
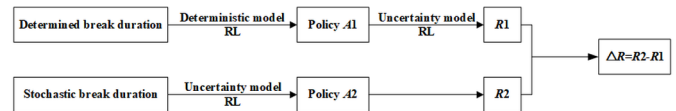


Fig. 6. Calculation schematic for system next mission reliability improvement $\Delta R$ of the proposed stochastic model compared to the determined model

policy $A2=[4,3,0,4]$ solved by RL, the reliability $R=0.9903$ and maintenance time is 0.49 days. In order to compare the difference between the stochastic and deterministic break duration, the strategy $A2$ solved for the deterministic case is substituted into the uncertainty model to find the reliability $R1=0.9795$. Therefore, the difference between the deterministic strategy and the strategy substituted into the uncertainty model is 1.08%. As seen in Table 5 the maintenance policy considering uncertainty is better and system next mission reliability improvement $\Delta R=0.83\%$. Based on the above observations, the reliability of the system successfully complete the next mission in the deterministic case will be overestimated if the uncertainty of the break duration is ignored.

Table 5. Difference between stochastic and determined break duration(time is in days)

| Case | Policy | Reliability | Maintenance time |
|---|---|---|---|
| Stochastic | [4,2,0,4] | 0.9878 | 0.446 |
| Deterministic strategy substitution in the uncertainty model | [4,3,0,4] | 0.9795 | 0.327 |

### 5.2. Case 2:Two-stage 5-component system

The two-stage 5-component system is studied with the structure diagram shown in Fig. 7. The relevant parameters of each component are derived from the Chen et al [9], as shown in Table 6. In Table 6, $\zeta$ denotes the characteristic constant of the age reduction factor. $\beta$ and $\eta$ denote the shape and scale parameters of the Weibull distribution. $B(k)$ denotes the effective age of component at the beginning of the $k$th break. $X(k)$ denotes the state of each system component at the beginning of the $k$th break. The different maintenance actions of various components consume different time, as shown in Table 7, and 0~5 represent different codes of maintenance actions in order.
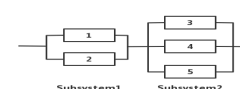


Fig. 7. Two-stage 5-component system structure diagram

*Table 6. Component parameters*

| ID | Characteristic constant of age reduction factor | Shape and scale parameters of Weibull distribution | | Effective age of component | Initial state of component |
|---|---|---|---|---|---|
| | ζ | β | η | B(k) | X(k) |
| 1 | 2.2 | 2.0 | 20 | 20 | 1 |
| 2 | 2.3 | 2.1 | 19 | 25 | 0 |
| 3 | 2.1 | 2.0 | 21 | 25 | 0 |
| 4 | 2.4 | 2.2 | 22 | 25 | 1 |
| 5 | 2.0 | 1.9 | 21 | 20 | 0 |

*Table 7. The maintenance time $t_{ij}(l)$ of different maintenance actions l for components (time is in days)*

| ID | Maintenance actions l code | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | 0 | 0.12 | 0.21 | 0.35 | 0.43 | 0.51 |
| 2 | 0 | 0.15 | 0.25 | 0.30 | 0.42 | 0.58 |
| 3 | 0 | 0.14 | 0.24 | 0.32 | 0.41 | 0.53 |
| 4 | 0 | 0.16 | 0.23 | 0.38 | 0.42 | 0.56 |
| 5 | 0 | 0.13 | 0.18 | 0.35 | 0.43 | 0.48 |

*Table 8. RL vs. GA result(time is in days)*

| Method | Maintenance policy | $R_{sys}$ | Maintenance time |
|---|---|---|---|
| RL | [2,5,5,0,2] | 0.983 | 1.5 |
| GA | [3,5,1,2,2] | 0.982 | 1.48 |

### 5.2.1. Comparison of Q-learning and GA

The studied case is introduced by the Chen et al [9], where the break duration $Z_k$ is a fixed 1.5 days and the $k$+1th mission duration $U$=5 days, which is solved using the GA. Under the condition of the same other parameters, the Q-learning algorithm is used to solve the problem, which is compared with the results of Chen et al [9], where the related Q-learning parameters $\alpha$=0.02 and $\gamma$=0.5. Using the PYTHON programming solution and obtained the maintenance policy $A$(s)=[2,5,5,0,2] . The reliability $R$=0.983 compared to the result solved by the Chen et al [9] using GA is 0.1% larger, and both results are shown in Table 8.



*Fig. 8. The training process of the proposed Q-learning algorithm*

### 5.2.2. RL Results with stochastic break duration

Assume that the $k$th mission is just completed now and the duration of the $k$th break $Z_k$ obeys a truncated normal distribution of $N$(1.5, 0.0225) with the range of [1.35, 1.65], $\tau$ =0.8, and the duration of the $k$+1th mission $U$=5 days. The maintenance policy $A$(s)=[2,5,2,2,2] solved using Q-learning, the reliability $R$=0.98 after maintenance. It can be seen that considering the determined break duration leads to an overestimation of the reliability of the system successfully complete next mission. The Q-learning parameters are the same as section 5.2.1, and the Q-learning process is shown in Fig. 8. In the first 7000 iterations, the agent randomly explores the possible maintenance actions, and the $Q$ matrix converges relatively slowly, after which the value of the $Q$ matrix gradually converges and eventually reaches the convergence state.

### 5.2.3. Comparison between stochastic and deterministic of break duration

When the break duration is a deterministic value of 1.5 days, the strategy solved by RL is $A$1=[2, 5, 5, 0, 2]. Bringing this strategy into the uncertainty model, i.e., the break duration $Z_k$ is a truncated normal distribution $N$(1.5, 0.0225) with the range of [1.35, 1.65], the optimal reliability $R$1=0.969 under the constraint $P(\text{T} \leq Z) \geq \tau$ ($\tau$= 0.8). The optimal maintenance policy $A$2=[2, 5, 2, 2, 2] solved by RL under the above uncertainty model has a reliability $R$2=0.98. Therefore, system next mission reliability improvement $\Delta R$=0.011 shows that the maintenance policy considering stochastic is better than the deterministic one with 1.1% higher reliability.

*Table 9. System next mission reliability improvement ΔR (%) at the different mean and standard deviation of the distribution of break duration*

| Distribution mean | Distribution standard deviation | | | | |
|---|---|---|---|---|---|
| | 0.01 | 0.05 | 0.1 | 0.15 | 0.2 |
| | △R (%) between stochastic and deterministic break duration | | | | |
| 1 | 0.0 | 0.0 | 0.0 | 0.5 | 0.5 |
| 1.2 | 0.0 | 0.6 | 0.5 | 0.5 | 0.5 |
| 1.4 | 0.1 | 1.2 | 1.2 | 0.6 | 0.5 |
| 1.6 | 0.3 | 1.2 | 1.2 | 0.9 | 0.8 |
| 1.8 | 0.3 | 1.4 | 1.4 | 1.2 | 1.2 |
| 2 | 0.1 | 2.3 | 2.2 | 2.1 | 2.1 |
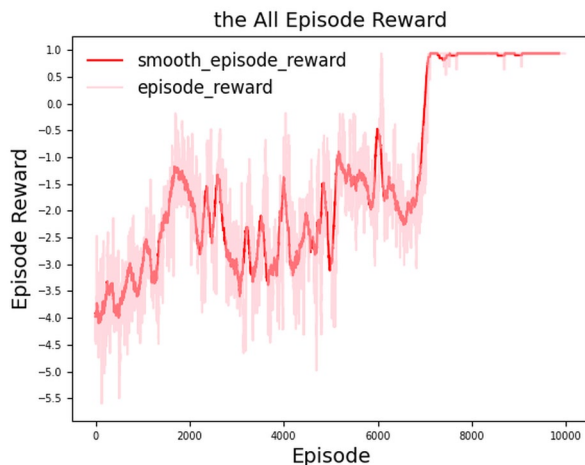
The impact of stochastic break duration on the maintenance strategy is illustrated by comparing the system reliability between stochastic and determined break duration. Under the same chance constraint and other parameters, the sequential simulations obtain the system successfully completing the next mission reliability $R2$ with different mean and standard deviation by varying the break duration obeying distribution in the uncertainty model. Mean $M$={1, 1.2, 1.4, 1.6, 1.8, 2}, standard deviation $STD$={0.01, 0.05, 0.1, 0.15, 0.2}, 30 combinations exist, and 30 sets of simulation experiments were implemented. The determined break duration $T$={1, 1.2, 1.4, 1.6, 1.8, 2}, respectively, are derived from the corresponding maintenance policy $A_T$ in the deterministic model by RL, and the system reliability $R$1 is derived by substituting the maintenance policy $A_T$ into the uncertain model with M=T. The results of system next mission reliability improvement $\Delta R$ are shown in Table 9 and Fig. 9 below.

As seen in Fig. 9, the overall trend of system next mission reliability improvement $\Delta R$ increases with the mean value, indicating
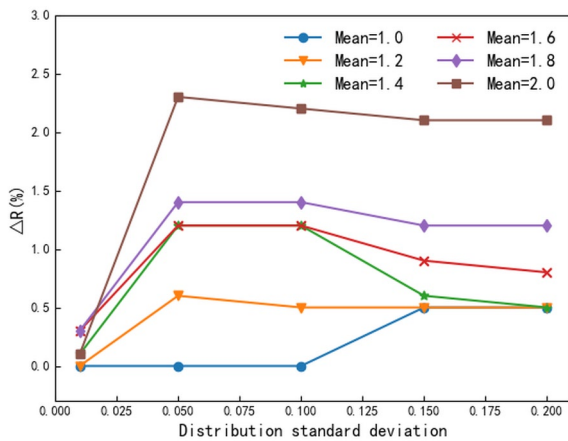
*Fig. 9. System next mission reliability improvement ΔR at the different mean and standard deviation of the distribution of break duration*

that the difference between uncertainty and certainty is more evident with larger mean values. It is because the increase of the mean value leads to a relatively long break time, allowing to select some maintenance actions with higher code. And the higher-code maintenance action requires longer maintenance time and better maintenance effect. Since the strategy is derived when the break is deterministic, there is sufficient time to complete all maintenance actions. However, in the case of uncertainty, there may not be enough time to complete all maintenance actions due to the constraint of insufficient time. As a result, a maintenance action is performed only partially and not fully completed, in which case the component reliability is unchanged. The relatively high code of the selected maintenance action when the mean value is large leads to lower system reliability under the inability to complete all maintenance actions, resulting in a larger ΔR.

In addition, when the mean value is 1, and the standard deviation is less than 0.1, the ΔR is equal to 0 in the first period and increases with the standard deviation. When the mean value is 1, it is at the left end of the distribution range, and the break duration does not change much with the standard deviation increase in the first period. Then it increases more obviously so that there are more optional maintenance actions, and the final system reliability increases. When the mean value is other values, ΔR increases with the standard deviation increase and gradually becomes smaller. It is because, in the beginning, the standard deviation is small, the uncertainty case is close to the deterministic case, and ΔR is small and close to 0. As the standard deviation increases, the duration of the break decreases relatively gradually, and the gap is the largest at the initial stage, leading to the largest ΔR, and then ΔR gradually decreases. The decrease in the break duration causes it as the standard deviation increases. It can be seen from the above figure that ΔR is greater than or equal to 0, and the maximum difference value reaches 2.3%. It shows that the model considering uncertainty is significantly better than the deterministic model. Considering a deterministic break duration can lead to an overestimation of reliability. In case of uncertainty encountered, it may lead to the inability of the system to complete subsequent mission.

### 5.2.4. Sensitivity analysis of component parameters

The optimization objective of this paper is to maximize the reliability of the system successfully completing the next mission. The mission duration $U$, the characteristic parameter $\zeta$, and the Weibull distribution parameter $\beta$, $\eta$ directly affect the optimization results. Sensitivity analysis is performed on the above parameters to verify the validity of the model, the feasibility of the method, and the influence of stochastic on the maintenance policy. For the selective maintenance decision model, the parameter $U$ determines the mission duration, and the larger $U$ is, the lower the reliability $R$. The characteristic

parameter $\zeta$ reflects the relationship between the maintenance time and the age reduction factor. The larger $\zeta$ is, the more pronounced the maintenance effect of the same maintenance time is, i.e., the larger reliability $R$ is. The shape and scale parameters $\beta$ and $\eta$ of the Weibull distribution obeyed by the component failure time, respectively, and the larger $\beta$ and $\eta$ are, the larger reliability $R$ is. The following experiments were conducted to verify the effects of $U$, $\zeta$, $\beta$, and $\eta$ on maintenance decisions.

Simulation tests are performed in three categories, $U$ and $\zeta$, $U$ and $\beta$, and $U$ and $\eta$. 25 combinations exist in each category, respectively. $U=\{5, 6, 7, 8, 9\}$, $\zeta=\{1.8, 2, \text{baseline}(2.2, 2.3, 2.1, 2.4, 2.0), 2.4, 2.6\}$, $\beta=\{1.7, 1.9, \text{baseline}(2.0, 2.1, 2.0, 2.2, 1.9), 2.2, 2.4\}$, and $\eta=\{17, 19, \text{baseline}(20, 19, 21, 22, 21), 22, 24\}$. Except for the baseline parameter value in the table 6, the parameters of the remaining components are taken as shown in the above set and are the same, and all other model parameters and algorithm parameters are the same as in section 5.2.2. Firstly, maintenance policy $A$ is derived in the deterministic case. Then the reliability $R1$ is obtained by substituting maintenance policy $A$ from the deterministic model into the uncertainty model. The reliability $R1$ is compared with the reliability $R2$ obtained in the uncertainty case. The results of system next mission reliability improvement $\Delta R$ for each type of experiment are shown in the following Table 10-12 and Figs. 10-12.

*Table 10. System next mission reliability improvement ΔR(%) for different mission duration U and component characteristic constants ζ*

| characteristic constant $\zeta$ | Mission duration $U$ | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 |
| 1.8 | 0.8 | 1.6 | 2.8 | 4.2 | 6.2 |
| 2 | 0.9 | 1.7 | 2.8 | 4.4 | 6.5 |
| baseline | 1.1 | 1.8 | 3.0 | 4.6 | 6.5 |
| 2.4 | 1.2 | 1.9 | 3.0 | 4.5 | 6.5 |
| 2.6 | 1.3 | 2.1 | 3.1 | 4.7 | 6.6 |

*Table 11. System next mission reliability improvement ΔR(%) for different mission duration U and Weibull distribution shape parameter β*

| Shape parameter $\beta$ | Mission duration $U$ | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 |
| 1.7 | 1.3 | 1.8 | 2.5 | 3.7 | 5.2 |
| 1.9 | 1.3 | 1.9 | 2.9 | 5.8 | 7.9 |
| baseline | 1.1 | 1.8 | 3.0 | 4.6 | 9.0 |
| 2.2 | 1.1 | 2.0 | 3.2 | 6.5 | 9.0 |
| 2.4 | 1.0 | 1.9 | 3.2 | 6.6 | 9.2 |

*Table 12. System next mission reliability improvement ΔR(%) for different mission duration U and Weibull distribution scale parameter η*

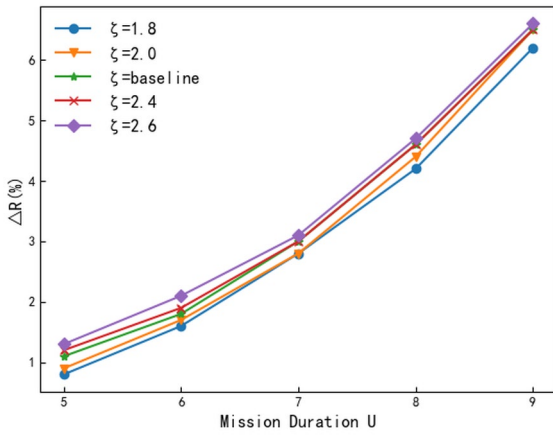| Scale parameter $\eta$ | Mission duration $U$ | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 |
| 17 | 2.1 | 3.9 | 6.2 | 10.6 | 14.8 |
| 19 | 1.2 | 2.4 | 4.0 | 6.0 | 8.8 |
| baseline | 1.1 | 1.8 | 3.0 | 4.6 | 9.5 |
| 22 | 0.7 | 1.2 | 2.0 | 3.2 | 4.7 |
| 24 | 0.6 | 0.9 | 1.3 | 2.0 | 3.2 |

*Fig. 10. System next mission reliability improvement ΔR with different component characteristic constants ζ*
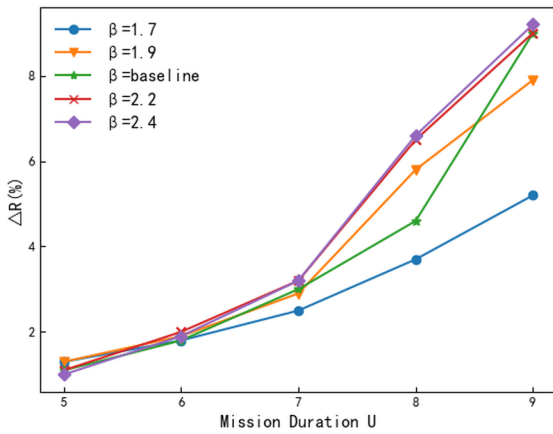


*Fig. 11. System next mission reliability improvement ΔR with different weibull distribution shape parameter β*
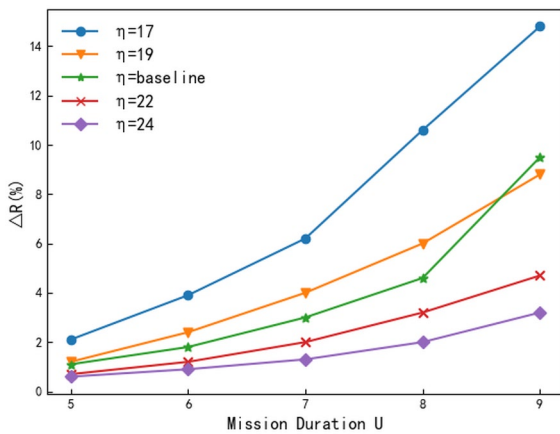


*Fig. 12. System next mission reliability improvement ΔR with different weibull distribution scale parameter η*

From Figs. 10-12, it can see that next mission reliability improvement ΔR changes less with parameter $\zeta$, and the overall ΔR gradually increases with the increase of parameter $\zeta$. When parameter $U$ is less than 7, ΔR is less affected by parameter $\beta$ and almost unchanged, and ΔR is gradually increased by parameter $\beta$ with the increase of parameter $U$. The larger parameter $\beta$ is, the larger ΔR is. ΔR changes

more obviously with the increase of parameter $\eta$, and ΔR gradually decreases with the increase of parameter $\eta$. The above figure shows that ΔR is influenced by parameter $U$ the most, followed by parameter $\eta$, and parameter $\zeta$ has the least influence on ΔR. Among them, ΔR reaches a maximum of 14.8% when analyzing the effect of parameter $\eta$. Therefore, the superiority of uncertainty is mainly influenced by the component parameters $\beta$, $\eta$, and the mission duration $U$, relative to the deterministic break. And in the case of larger mission duration $U$, considering the superiority of stochastic break duration is more prominent. Indicating that the larger the parameter $U$, the greater the uncertainty influence is also.

In summary, the model and algorithm accurately reflect the difference between uncertainty and certainty under each parameter, verifying the validity of the model and the feasibility of the method. This analysis also shows that the model and method apply to other systems. Through the above analysis, ignoring the uncertainty of the break duration can significantly impact the reliability of system to complete the next mission. In the case of large relevant parameters, ignoring the uncertainty of the mission can lead to an overestimation of the system reliability. It can result in a high risk of not being able to complete the next mission.

### 5.3. Case 3: A complex multi-component coal transportation system

To further verify the the validity of the model and the method, which is also valid for large-scale systems, the coal transmission system of literature [24] is used here as an example. The system consists of 5 subsystems connected in series and 14 components connected in parallels, and its structural sketch is shown in Fig. 13. The relevant parameters of each component are shown in Table 13, derived from the literature [24, 26]. In table 13, $m_l^p$, $m_l^f$ denotes the characteristic constants of the age reduction factor for preventive maintenance action and corrective maintenance action, respectively. $t_l^0$, $t_l^p$, $t_l^f$ denotes the fixed maintenance time, preventive maintenance time, and corrective maintenance time, respectively. $\beta_l$ and $\eta_l$ denote the shape and scale parameters of the Weibull distribution. $B(k)$ denotes the effective age of the component at the beginning of the $k$th break. $X(k)$ denotes the state of each system component at the beginning of the $k$th break.
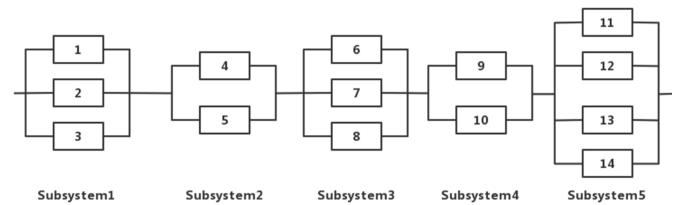


*Fig. 13. Five-stage 14-element system structure sketch*

Each component has 8 different of maintenance action $l$. $L(l=7)$ represents the highest maintenance level, where $l=0$ and $l=7$ denote no maintenance and replacement, respectively. For functioning components, $l=1 \sim 6$ indicates imperfect maintenance. For failed components, $l=1$ indicates minimal repair, and $l=2\sim6$ indicates imperfect maintenance. When $l>1$, the time for maintenance action $l$ is $t=t_{ij,l}+t_l^0$, where $t_{ij,l}$ is expressed as follows:

$$t_{ij,l} = \begin{cases} \dfrac{(l_{ij}-1)t_l^f}{L-1} & X_{\text{break,s}} = 0 \\[4mm] \dfrac{l_{ij} \cdot t_l^f}{L} & X_{\text{break,s}} = 1 \end{cases} \tag{23}$$

Table 13. Relevant parameter values for each component (time is in days)

| ID | $\beta_l$ | $\eta_l$ | $m_l^p$ | $m_l^f$ | $t_l^p$ | $t_l^f$ | $t_l^0$ | $B(k)$ | $X(k)$ |
|----|-----------|----------|---------|---------|---------|---------|---------|--------|--------|
| 1 | 1.5 | 25 | 2.5 | 2.5 | 0.13 | 0.25 | 0.03 | 35 | 1 |
| 2 | 2.4 | 38 | 2.2 | 2.0 | 0.2 | 0.31 | 0.03 | 24 | 0 |
| 3 | 1.6 | 28 | 2.6 | 3.0 | 0.2 | 0.33 | 0.03 | 45 | 0 |
| 4 | 2.6 | 40 | 2.2 | 3.2 | 0.12 | 0.32 | 0.04 | 35 | 0 |
| 5 | 1.8 | 28 | 1.8 | 4.0 | 0.21 | 0.34 | 0.02 | 28 | 1 |
| 6 | 2.4 | 34 | 2.4 | 3.2 | 0.14 | 0.19 | 0.03 | 36 | 1 |
| 7 | 2.5 | 26 | 2.8 | 3.0 | 0.2 | 0.27 | 0.05 | 44 | 0 |
| 8 | 2.0 | 28 | 2.3 | 2.8 | 0.17 | 0.31 | 0.05 | 28 | 0 |
| 9 | 1.2 | 26 | 2.0 | 2.5 | 0.18 | 0.26 | 0.04 | 38 | 1 |
| 10 | 1.4 | 35 | 2.5 | 2.8 | 0.2 | 0.32 | 0.05 | 15 | 0 |
| 11 | 2.8 | 40 | 3.2 | 3.0 | 0.21 | 0.31 | 0.07 | 30 | 0 |
| 12 | 1.5 | 35 | 2.6 | 2.2 | 0.23 | 0.33 | 0.04 | 22 | 1 |
| 13 | 2.4 | 30 | 2.8 | 2.8 | 0.16 | 0.35 | 0.06 | 38 | 1 |
| 14 | 2.2 | 45 | 2.2 | 2.6 | 0.14 | 0.35 | 0.05 | 35 | 0 |

where $t_{ij,l}$ denotes the maintenance time to perform action $l$ on component $C_{ij}$. $l_{ij}$ denotes the selected maintenance action for component $C_{ij}$.

The age reduction factor $b_{ij,l}$ is calculated as follows:

$$b_{ij,l} = \begin{cases} 1-(\dfrac{t_{ij,l}}{t_l^f})^{\frac{1}{m_l^f}} & X_{\text{break,s}} = 0 \\[3mm] 1-(\dfrac{t_{ij,l}}{t_l^p})^{\frac{1}{m_l^p}} & X_{\text{break,s}} = 1 \end{cases} \qquad (24)$$

### 5.3.1. Algorithm performance analysis

Assuming that the $k$th mission has just been completed now, the duration $Z_k$ of the $k$th break obeys a truncated normal distribution of $N(3, 0.0625)$ with range of [2.5, 3.5], $\tau = 0.8$, and the duration of the $k+1$th mission $U$=10 days. All other component parameters are shown in Table 13. Among the parameters related to the GA algorithm, the number of populations NP=150, the crossover rate pc=0.8, the variation rate pm=0.05, and the maximum number of iterations equal to 4000. The parameters related to Q-learning, the learning rate $\alpha$=0.02, the discount rate $\gamma$=0.5, and the maximum number iterations equal to 25000. Due to the stochastic of the algorithm, 10 sets of simulations are performed for each method to find its optimal maintenance policy $A_{\text{best}}$, the maintenance time for the optimal maintenance policy $T_{\text{best}}$,

the average reliability $R_{\text{mean}}$, the maximum reliability $R_{\text{best}}$, the variance $R_{\text{std}}$ and the average running time $\overline{S}$ as the comparison results. In addition, a parameter %QOS is introduced here as a performance metric to compare the quality of RL and GA solution, $\%QOS=(R_{\text{best}}-R_{\text{mean}})/R_{\text{best}}$. The comparison results and performance metric results are shown in Table 14.

From Table 14, we can see that the optimal maintenance policy solved by Q-learning is better than GA, and the maximum reliability $R_{\text{best}}$ is 1.23% higher. Furthermore, the mean and variance of Q-learning results are better than GA in 10 solving results, indicating that the Q-learning algorithm's stability is better than GA. Combined with the experimental results in previous section, the Q-learning algorithm effectively solves the selective maintenance problem and can obtain better values than the GA algorithm. Regarding the computation time, the average time taken by GA is more than twice of RL. Regarding the quality of the obtained solutions, the optimal solution of RL is better than that of GA, and the average solution of RL deviates from the optimal solution by only 1.1%, while the average solution of GA deviates from the optimal solution by 1.38%. Therefore, the RL algorithm can find higher quality solutions and further verifies the effectiveness of the algorithm.This case also illustrates that the advantages of RL are more pronounced for more complex systems.

The iterative evolution of the proposed RL algorithm is shown in Fig. 14. During the initial 15000 iterations, the agent randomly explores all possible maintenance actions, and the $Q$ matrix's value converges slowly. After the first 15,000 iterations of random exploration learning, the $Q$ matrix gradually converges and can eventually reach the convergence state.

Table 14. Comparison results of the two algorithms(time is in days)

| Method | $A_{\text{best}}$ | $T_{\text{best}}$ | $R_{\text{best}}$ | $R_{\text{mean}}$ | $R_{\text{std}}$ | $\overline{S}$ | %QOS |
|--------|-------------------|-------------------|-------------------|-------------------|------------------|----------------|------|
| Q-learning | [0,7,3,7,6,7,6,4,7,7,3,0,0,6] | 2.795 | 0.9414 | 0.9314 | 0.0031 | 71.3 | 1.1 |
| GA | [0,7,3,7,4,5,6,3,5,7,3,7,5,1] | 2.745 | 0.9291 | 0.9171 | 0.0061 | 148.3 | 1.38 |

Table 15. Difference between stochastic and determined break duration(time is in days)

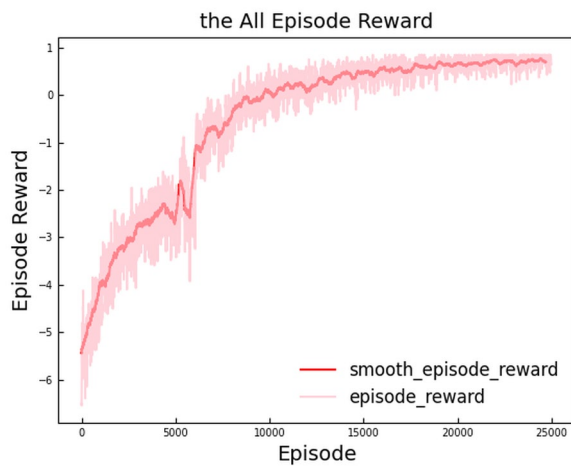| Case | Maintenance policy | Reliability | Maintenance time |
|------|--------------------|-------------|------------------|
| Stochastic | [0,7,3,7,6,7,6,4,7,7,3,0,0,6] | 0.9414 | 2.795 |
| Deterministic strategy substitution in the uncertainty model | [0,6,7,7,6,6,7,4,0,7,2,0,7,2] | 0.924 | 2.703 |

*Fig. 14. The training process of the proposed Q-learning algorithm*

### 5.3.2. Comparison between stochastic and determinism of break duration

When the duration of the break is stochastic, it can be seen from 5.3.1 that the optimal maintenance policy $A1$=[0,7,3,7,6,7,6,4,7,7,3,0,0,6], the reliability of the system successfully completing next mission $R2$=0.9414, and the maintenance time is 2.795 days. When the break duration $Z_k$=3 is a fixed value and other parameters remain unchanged, the optimal maintenance policy $A2$=[0,6,7,7,6,6,7,4, 7,7,2,0,7,2], the system reliability $R$=0.944, and the maintenance time is 2.923 days. To compare the difference between stochastic and determinism, the maintenance policy $A2$ solved for the deterministic case is substituted into the uncertainty model $R1$=0.924. As seen in Table 15, the uncertainty is not negligible, and the difference between the deterministic maintenance policy and the policy substituted into the uncertainty model is 2%. It can also be seen that the strategy considering uncertainty is better than the deterministic one with a reliability 1.74% higher. Based on the above observations, it can be concluded that the optimal solution in the deterministic case does not guarantee the max-imum reliability for successful completion of the next mission in the uncertainty case if the uncertainty in the break duration is ignored.

## 6. Conclusions and future works

This paper presents a new selective maintenance model for a multi-component system with the decision to maximize the system's reliability to complete the next mission. The components can be maintained during the break between two adjacent missions, each with several optional maintenance actions from minimal repair and imperfect maintenance to replacement. At the same time, this selective maintenance optimization model considers the break duration stochastic, represented by an appropriate probability distribution. The selective maintenance optimization problem is modeled as a Markov Decision Process (MDP). Based on the framework of the MDP, a RL approach is proposed to overcome the problems of complexity and low computational efficiency in solving the model by traditional methods. By analyzing three cases, the accuracy of the model and the RL method are demonstrated to be effective in finding the optimal maintenance strategy. By comparing with the GA method, the more complex the system the more obvious the advantage of RL. The RL can obtain a better maintenance policy making the system more reliable to successfully complete the next mission, and the computation takes much less time than GA. It is also demonstrated that the stochastic of the break duration affects the maintenance policy and the reliability of the system successfully complete the next mission. Ignoring the stochastic of the break duration the reliability of the system successfully complete the next mission will be overestimated and may prevent the system from completing the next mission. Therefore, it is necessary to investigate the optimization of selective maintenance under uncertainty.

In future works, we will explore several questions. Here we study systems consisting of two-state multiple components, where the break is the only uncertain maintenance resource. The process has several intermediate states in practical engineering from function to failure. In addition, the maintenance time required for different maintenance actions may be stochastic due to the different skill levels of different technicians. In the future, we will conduct research for multi-state multi-component systems and other uncertain maintenance resources.

## References

1. Andriotis C P, Papakonstantinou K G. Managing engineering systems with large state and action spaces through deep reinforcement learning. Reliability Engineering & System Safety, 2019; 191: 106483, https://doi.org/10.1016/j.ress.2019.04.036.
2. Barde S R A, Yacout S, Shin H. Optimal preventive maintenance policy based on reinforcement learning of a fleet of military trucks. Journal of Intelligent Manufacturing, 2019; 30(1): 147-161, https://doi.org/10.1007/s10845-016-1237-7.
3. Baxter L A, Kijima M, Tortorella M. A point process model for the reliability of a maintained system subject to general repair. Stochastic models, 1996; 12(1): 12-1, https://doi.org/10.1080/15326349608807372.
4. Cao W, Jia X, Hu Q, et al. A literature review on selective maintenance for multi-unit systems. Quality and Reliability Engineering International 2018; 34(5): 824-845, https://doi.org/10.1002/qre.2293.
5. Cao W, Jia X, Hu Q, et al. Selective maintenance for maximising system availability: a simulation approach. International Journal of Innovative Computing and Applications 2017; 8(1): 12-20, https://doi.org/10.1504/ijica.2017.082493.
6. Cao W, Li F, Ran Q. Study on selective maintenance optimization for multi-State systems confronting random missions. Journal of Ordnance Engineering College 2017; 29(2): 17-22.
7. Cassady C R, Murdock Jr W P, Pohl E A. Selective maintenance for support equipment involving multiple maintenance actions. European Journal of Operational Research 2001; 129(2): 252-258, https://doi.org/10.1016/s0377-2217(00)00222-8.
8. Chen Y, Jiang T, Liu Y. Selective maintenance optimization: research advances and challenges. Operations Research Transactions, 2019, 23(3): 27-46, 10.15960/j.cnki.issn.1007-6093.2019.03.003.
9. Chen Y, Ma Y, Liu Q, et al. Research on selective maintenance decision-making of equipment considering imperfect maintenance under sequential mission. AERO WEAPONRY 2019.
10. Chen Z, Zhang L, Tian G, et al. Economic maintenance planning of complex systems based on discrete artificial bee colony algorithm. IEEE Access 2020; 8: 108062-108071, https://doi.org/10.1109/ACCESS.2020.2999601.

11. Diallo C, Venkatadri U, Khatab A, et al. Optimal selective maintenance decisions for large serial k-out-of-n: G systems under imperfect maintenance. Reliability Engineering & System Safety 2018; 175: 234-245, https://doi.org/10.1016/j.ress.2018.03.023.

12. Dui H, Zheng X, Zhao QQ, Fang Y. Preventive maintenance of multiple components for hydraulic tension systems. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2021; 23 (3): 489–497, https://doi.org/10.17531/ein.2021.3.9.

13. Gao H, Zhang X, Yang X, et al. Optimal selective maintenance decision-making for consecutive-mission systems with variable durations and limited maintenance Time. Mathematical Problems in Engineering, 2021; (2021), https://doi.org/10.1155/2021/5534659.

14. Hu M. Research on maintenance decision of wind turbine components based on reinforcement learning. School of Energy Power and Mechanical Engineer.

15. Huang J, Chang Q, Arinez J. Deep reinforcement learning based preventive maintenance policy for serial production lines. Expert Systems with Applications 2020; 160: 113701, https://doi.org/10.1016/j.eswa.2020.113701.

16. Hu Y, Miao X, Zhang J, et al. Reinforcement learning-driven maintenance strategy: A novel solution for long-term aircraft maintenance decision optimization. Computers & Industrial Engineering, 2021; 153: 107056, https://doi.org/10.1016/j.cie.2020.107056.

17. Khatab A, Aghezzaf E H, Djelloul I, et al. Selective maintenance for series-parallel systems when durations of missions and planned breaks are stochastic. IFAC-PapersOnLine 2016; 49(12): 1222-1227, https://doi.org/10.1016/j.ifacol.2016.07.677.

18. Khatab A, Aghezzaf E H, Djelloul I, et al. Selective maintenance optimization for systems operating missions and scheduled breaks with stochastic durations. Journal of manufacturing systems 2017; 43: 168-177, https://doi.org/10.1016/j.jmsy.2017.03.005.

19. Khatab A, Aghezzaf E H. Selective maintenance optimization when quality of imperfect maintenance actions are stochastic. Reliability engineering & system safety 2016; 150: 182-189, https://doi.org/10.1016/j.ress.2016.01.026.

20. Khatab A, Aghezzaf E L H, Diallo C, et al. Selective maintenance optimisation for series-parallel systems alternating missions and scheduled breaks with stochastic durations. International Journal of Production Research 2017; 55(10): 3008-3024, https://doi.org/10.1080/00207543.2017.1290295.

21. Li Z, Xu Y, Gao S, et al. Reliability modeling for repairable multi-state Elements based on Markov process. AERO WEAPONRY, 2018, 10.19297/j.cnki.41-1228/tj.2018.05.012.

22. Li Z, Zhong S, Lin L. An aero-engine life-cycle maintenance policy optimization algorithm: Reinforcement learning approach. Chinese Journal of Aeronautics 2019; 32(9): 2133-2150.

23. Lin D, Zuo M J, Yam R C M. General sequential imperfect preventive maintenance models. International Journal of reliability, Quality and safety Engineering 2000; 7(03): 253-266, https://doi.org/10.1142/S0218539300000213.

24. Liu Y, Chen Y, Jiang T. Dynamic selective maintenance optimization for multi-state systems over a finite horizon: A deep reinforcement learning approach. European Journal of Operational Research 2020; 283(1): 166-181, https://doi.org/10.1016/j.ejor.2019.10.049.

25. Liu Y, Chen Y, Jiang T. On sequence planning for selective maintenance of multi-state systems under stochastic maintenance durations. European Journal of Operational Research 2018; 268(1): 113-127, https://doi.org/10.1016/j.ejor.2017.12.036.

26. Liu Y, Huang H Z. Optimal selective maintenance strategy for multi-state systems under imperfect maintenance. IEEE Transactions on Reliability 2010; 59(2): 356-367, https://doi.org/10.1109/TR.2010.2046798.

27. Lust T, Roux O, Riane F. Exact and heuristic methods for the selective maintenance problem. European journal of operational research 2009; 197(3): 1166-1177, https://doi.org/10.1016/j.ejor.2008.03.047.

28. Lv X Z, Yu Y L, Zhang L, et al. Stochastic program for selective maintenance decision considering diagnostics uncertainty of built-in test equipment. IEEE 2011; 584-589, https://doi.org/10.1109/ICQR2MSE.2011.5976681.

29. Mahmoodzadeh Z, Wu K Y, Lopez Droguett E, et al. Condition-based maintenance with reinforcement learning for dry gas pipeline subject to internal corrosion. Sensors 2020; 20(19): 5708, https://doi.org/10.3390/s20195708.

30. Malik M A K. Reliable preventive maintenance scheduling.AIIE transactions, 1979; 11(3):221-228, https://doi.org/10.1080/05695557908974463.

31. Martínez-Tenor A, Fernández-Madrigal J A, Cruz-Martín A, et al. Towards a common implementation of reinforcement learning for multiple robotic tasks. Expert Systems with Applications 2018; 100: 246-259, https://doi.org/10.1016/j.eswa.2017.11.011.

32. Nakagawa T. Optimum policies when preventive maintenance is imperfect. IEEE Transactions on Reliability 1979; 28(4): 331-332, https://doi.org/10.1109/TR.1979.5220624.

33. Pandey M, Zuo M J, Moghaddass R, et al. Selective maintenance for binary systems under imperfect repair. Reliability Engineering & System Safety 2013; 113: 42-51, https://doi.org/10.1016/j.ress.2012.12.009.

34. Paraschos P D, Koulinas G K, Koulouriotis D E. Reinforcement learning for combined production-maintenance and quality control of a manufacturing system with deterioration failures. Journal of Manufacturing Systems, 2020; 56: 470-483, https://doi.org/10.1016/j.jmsy.2020.07.004.

35. Peng S. Reinforcement learning with Gaussian processes for condition-based maintenance. Computers & Industrial Engineering 2021; 158: 107321, https://doi.org/10.1016/j.cie.2021.107321.

36. Rajagopalan R, Cassady C R. An improved selective maintenance solution approach. Journal of Quality in Maintenance Engineering 2006; 12(2):172-185, https://doi.org/10.1108/13552510610667183.

37. Rice W F, Cassady C R, Nachlas J A. Optimal maintenance plans under limited maintenance time. Proceedings of the seventh industrial engineering research conference: 1998: 1-3.

38. Ruan J H, Wang Z X, Chan F T S, et al. A reinforcement learning-based algorithm for the aircraft maintenance routing problem. Expert Systems with Applications 2021; 169: 114399, https://doi.org/10.1016/j.eswa.2020.114399.

39. Su Y, Meng L, Kong X, et al. Generative adversarial networks for gearbox of wind turbine with unbalanced data sets in fault diagnosis. IEEE Sensors Journal, 2022; https://doi.org/10.1109/JSEN.2022.3178137.

40. Sun Y, Sun Z. Selective Maintenance on a Multi-State Transportation System Considering Maintenance Sequence Arrangement. IEEE Access 2021; 9: 70048-70060, https://doi.org/10.1109/ACCESS.2021.3078140.

41. Tanwar M, Rai R N, Bolia N. Imperfect repair modeling using Kijima type generalized renewal process. Reliability Engineering & System Safety 2014; 124: 24-31, https://doi.org/10.1016/j.ress.2013.10.007.

42. Wang H, Pham H. A quasi renewal process and its applications in imperfect maintenance. International journal of systems science 1996; 27(10): 1055-1062, https://doi.org/10.1080/00207729608929311.

43. Wang S, Zhang S, Li Y, et al. Selective maintenance decision-making of complex systems considering imperfect maintenance. International

Journal of Performability Engineering 2018; 14(12): 2960, https://doi.org/10.23940/IJPE.18.12.P6.29602970.

44. Xu Q Z, LM Guo. Method for solving the selective maintenance problem for series-parallel system. Machinery Design & Manufacture 2016; 0(1): 61-65, https://doi.org/10.3969/j.issn.1001-3997.2016.01.017.

45. Yan J, Zhang Q, Hu X. Review of path planning techniques based on reinforcement learning. Computer Engineering 2021; 47(10):10, 10.19678/j.issn.1000-3428.0060683.

46. Yang Z, Qi C. Preventive maintenance of a multi-yield deteriorating machine:Using reinforcement learning . Systems Engineering-Theory & Practice, 2013, 33(7): 1647-1653.

47. Yousefi N, Tsianikas S, Coit D W. Dynamic maintenance model for a repairable multi-component system using deep reinforcement learning. Quality Engineering, 2022; 34(1):16-35, https://doi.org/10.1080/08982112.2021.1977950.

48. Zhao J, Liu J, Zhao Z, et al. A high-performance maintenance strategy for stochastic selective maintenance. Concurr Comp-Pract E 2019; 31(12): e4840, https://doi.org/10.1002/cpe.4840.

49. Zhao X, Al-Khalifa K N, Hamouda A M, et al. Age replacement models: A summary with new perspectives and methods. Reliability Engineering & System Safety 2017; 161: 95-105, https://doi.org/10.1016/j.ress.2017.01.011.
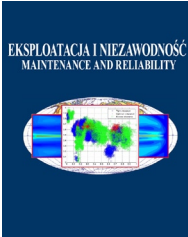
# Comprehensive importance analysis for repairable system components based on the GO method

**Indexed by:**
Web of Science Group

Xiuhong Jiang[a,b,*], Yuying Wang[a], Jiaxin Li[c], Linlin Ye[a]

[a] Shenyang Aerospace University, Department of Electronic Information Engineering, Shenyang 110000, Liaoning, China
[b] Liaoning General Aviation Academy, Shenyang 110000, Liaoning, China
[c] Aero Engine Corporation of China, Shenyang 110000, Liaoning, China

## Highlights

- A method for evaluating the importance of components is proposed for repairable systems.
- Use CRITIC to determine weights of steady-state availability and failure importance.
- The method is verified using a PMSM drive system based on the GO method.

## Abstract

In order to effectively improve the reliability level of the permanent magnet synchronous motor (PMSM) drive system of electric aircraft, a component importance analysis based on the GO method for the repairable systems is proposed. Firstly, the system reliability model GO diagram is established according to the hardware schematic diagram of the PMSM drive system. Secondly, the steady-state availability and failure importance of the components are calculated. In addition, the criteria importance through intercriteria correlation (CRITIC) is adopted to determine the objective weights of steady-state availability and failure importance. The combined weighting is employed to obtain the importance of key components. Meanwhile, a system redundancy design based on the importance of components is proposed to provide data support for the design of the system. Finally, the feasibility and effectiveness of the proposed method are evaluated by an example of an electric aircraft PMSM drive system. This method provides a supporting basis for the optimization design of the entire system.

## Keywords

component importance, electric aircraft, GO method, reliability analysis, permanent magnet synchronous motor.

## 1. Introduction

Electric drive system provides power for the electric aircraft, which is generally composed of the power source, controller, driver board, permanent magnet synchronous motor (PMSM), etc. [7, 8, 18]. The performance of the electric drive system directly affects flight quality and flight safety, especially its reliability. If a failure is not eliminated in time, it can lead to serious accidents such as air distress or even a crash. Hence, in order to avoid such catastrophic events, it is of utmost importance to conduct appropriate importance analysis of the components of the drive system.

Component importance is an elementary part of the system and is determined by system structure, quality of manufacturing, and environmental conditions, etc. Assessment of component importance is one of the key tasks in system reliability analysis [12, 28]. Importance analysis combines the knowledge of sensitivity, risk, hazard and importance, and is a powerful tool for determining system weaknesses and improving system reliability design. Component importance evaluation is the influence of the change of component reliability parameters on the success probability of the system output. By improving

the reliability of components that have a greater impact on the success state of the system, the purpose is to considerably improve the reliability of the system in a simple way and at a lower cost and achieve maximum benefits. The importance analysis of each component in the electric drive system can provide strong data support for the improvement of system reliability, safety and system failure diagnosis [16, 22].

At present, traditional component importance analysis methods include structural importance, probability importance and critical importance, etc. [1, 17, 19]. In recent years, quite a few new analysis methods have also emerged. For example, Cai [2] used GO calculation to calculate the reliability of the logistics service supply chain system accurately and found the weak links affecting the reliability of the system by analysing the minimal cut set. Yang [26] judged the influence of the change of failure probability on the average failure-free working time by increasing the failure probability of some components by five times and keeping the failure probability of other components unchanged. Ma [14], Luo [11], and Chen [3] used the failure-tree reliability analysis method to comprehensively consider the importance analysis results of the three dimensions including probability

importance, critical importance and structural importance to judge the importance of components comprehensively. Jia [10] adopted the method of failure mode effects and criticality analysis (FMECA) to obtain the importance of civil aircraft components. Scherb [20] combined the relevant structural specifications and the observed failure rate in the entire system network to determine the impact of individual components on system reliability impact is sorted to assess the importance of components. Miziuła [15] and Xue [25] adopted Birnbaum's importance, comparing the availability before and after optimization of the constructed system can effectively reduce the maintenance cost of the system and enhance its availability. Fu [5] utilized a multi-layer network parsing method to evaluate component importance.

All the above studies assume that the object is a non-repairable system and the influence of the maintenance rate parameters of the components in the repairable system is not considered, which will lead to the inaccuracy of component importance analysis results. As a typical repairable electronic system, the electric drive system needs to consider the repair and update of the components during the importance analysis.

The PMSM drive system for electric aircraft is a repairable system. In order to effectively enhance the reliability level of the PMSM drive system of electric aircraft and identify the weakness of system design, a comprehensive component importance analytical method for repairable systems based on the GO method is proposed. According to the GO diagram of the reliability simulation model of electric aircraft drive system, the system steady-state availability and failure importance of key components are calculated by using parameters such as the maintenance rate and failure rate of components, and the objective weights of the system steady-state availability and failure importance are determined by CRITIC. The comprehensive importance of key components is obtained by weighted summation. Finally, the criticality of system components is identified through the case verification and analysis results of an electric aircraft PMSM drive system. Meanwhile, an idea of system redundancy design based on the importance of components is proposed to provide data support for the early design of the system. It can be verified that the proposed method can comprehensively evaluate the vulnerabilities of the PMSM drive system, which provides an important basis for the reliability design of the electric aircraft drive system.

## 2. Materials and Methods

### 2.1. Importance calculation of key components

The PMSM drive system of a general electric aircraft is principally composed of a power source (battery pack), TMS320F28335 DSP as the core controller, an IGBT drive board composed of 2SP0115T, an IGBT three-phase bridge inverter composed of FF600R07ME4, voltage sensor, current sensor, filter, bus transceiver, amplifier, 60kW PMSM and other repairable components, all components are industrial standards. Fig. 1 shows the basic structure schematic diagram of the system.
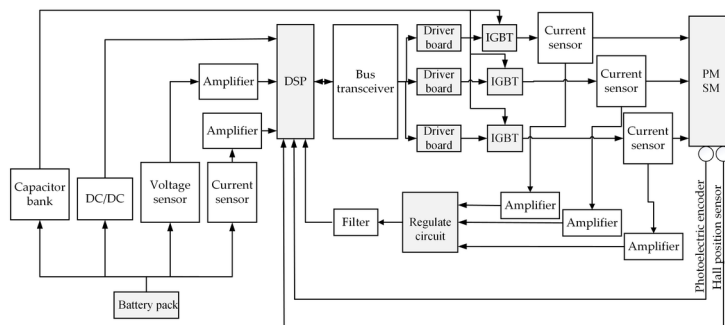


*Fig. 1. The schematic diagram of the basic structure of the electric aircraft PMSM system*

DC/DC converts battery pack energy and supplies power to DSP. As the control core, DSP has the advantage of realizing complex control algorithms and outputting high-precision pulse width modulation (PWM) through the corresponding algorithm. The PWM is connected with the external IGBT driver board through the bus transceiver to change the output power of the inverter, to achieve the purpose of PMSM control. The bus voltage is output by the voltage sensor as a differential signal and then transmitted to the DSP through the amplifier. The three-phase output current is transmitted to the DSP after amplification, conditioning and filtering. The bus current is also collected by the sensor and sent to the DSP for processing. The photoelectric encoder obtains the motor speed, rotation direction and absolute zero position, and the Hall sensor transmits the differential mode signal to the DSP for processing.

The GO method is a reliability analysis method of a success-oriented system that analyzes multi-state, time-series and process systems [4, 6, 13, 27]. The electric aircraft drive system is extremely suitable for reliability analysis using the GO method as a system with the current flow.

The basic idea of the GO method is to depict the operation, mutual relationship and logical relationship of specific units through operators (representing specific units or logical relationships) and signal flows (representing specific logistics or logical processes), and it directly translates system schematics, flowcharts, or engineering drawings into GO diagrams. The GO operation can be performed according to the operation rules of the operators and the signal flow direction after the GO map is established, and the quantitative analysis of the system's reliability can be completed.

The GO method defines 17 standard operators, which can simulate almost all combinations of component states and signal flows. Different operators correspond to different functions and simulate different components. For example, two-state unit operators can be used to model electronic components, alarms, amplifiers, batteries, safety valves, etc. Each operator has specified input and output data requirements and specified operation rules. The GO model of the PMSM drive system is established according to the basic structure schematic diagram of the system as shown in Fig. 2. Fig. 2 is equivalent to a translation of Fig. 1 using GO operators. The circles and triangles in the figure represent different types of GO operators. The number before "-" in the operator indicates the operator type, and the number after "-" indicates the operator number. Operators represent specific components or logical relationships. In Fig. 1, the power supply is used as the driving force for the entire system and is the system input, so it is represented by a single-signal generator (type 5 of GO operators). Owing to the disconnection of feedback, three-phase output current sensors, Hall sensors, and photoelectric encoders are also directly input as type 5 of GO operators. DC/DC converter, bus voltage sensor, bus current sensor, filter, conditioning circuit, DSP, bus transceiver, driver board, IGBT and PMSM only have two states of success and failure, so it is represented by a two-state unit (type 1 of GO operators). If one of the Hall sensors and the photoelectric encoder fails, the system can also work safely, accordingly, the OR gate (type 2 of GO operators) is used to indicate the relationship between the two. The signals collected to the DSP control board are indispensable, accordingly, an AND gate (type 10 of GO operators) is used to represent the logical relationship between the signals. In the same way, the relationship between the three IGBT is also AND. In this way, different components are simulated with different GO operators, and Fig. 2 is retrieved. The operation rules of specific operators are illustrated in the literature [21].

The arrow lines represent signal flows, representing specific logistics or logical processes.

When only one component fails, the equivalent failure rate $\lambda_{ri}$ of different operators can be calculated according to the operator operation rules. The calculation methods of several common operators are provided below (assuming that the failure rate follows exponential distribution):
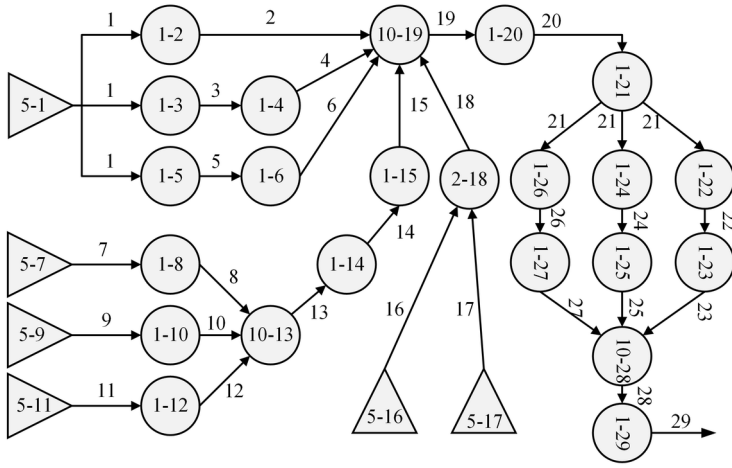
*Fig. 2. GO diagram of PMSM drive system*

(1) Two-state unit: a unit component with only one input signal and one output signal, and it has two states itself. When operating normally, the signal can pass, otherwise, the signal cannot pass:

$$\lambda_{ri} = \lambda_{si} + \lambda_{ci} \tag{1}$$

In the formula, $\lambda_{ri}$ is the equivalent failure rate of the output signal of the $\lambda_{si}$ is the equivalent failure rate of the input signal, and $\lambda_{ci}$ is the failure rate of the operator itself.

(2) Single signal generator: The operator's data is the output signal's data:

$$\lambda_{ri} = \lambda_{ci} \tag{2}$$

(3) AND gate: There are two or more signals in a parallel input signal for AND logic operation, and one signal is output. The operator itself has no data, and the output signal can succeed only when all input signals succeed. Consequently, its equivalent failure rate is the sum of all input signal failure rates:

$$\lambda_{ri} = \sum_{i=1}^{n} \lambda_{si} \tag{3}$$

(4) OR gate: If the input signals are independent of each other, as long as one of the input signals succeeds, the output will succeed. The operator itself has no data, and its equivalent failure rate is the product of the failure rates of all input signals:

$$\lambda_{ri} = \prod_{i=1}^{n} \lambda_{si} \tag{4}$$

## 2.2. Critical component importance calculation

In this paper, a component importance analysis method is proposed for the repairable PMSM drive system that integrates the steady-state availability and failure importance. The component steady-state availability is calculated by reliability parameters such as component maintenance rate and failure rate. The failure importance is calculated by the influence of the component from success state to failure state on the success probability of the system. Combined with the two to obtain the key components of importance $Y_i$, the computation formula is as follows:

$$Y_i = \omega_1 A_i + \omega_2 I_g(i) \tag{5}$$

In the formula, $A_i$ is the steady-state availability of component $i$; $I_g(i)$ is the failure importance of component $i$; $\omega_1$ is the weight of steady-state availability; $\omega_2$ is the weight of failure importance. The weight coefficient can be determined based on the CRITIC method. The calculation method of each quantity is explained below.

### 2.2.1. Component steady-state availability

Electric aircraft drive systems are repairable systems. Availability is a measure of the probability that a device is in a normal working or usable state when it begins to execute a work task at any time. Steady-state availability is one of the important reliability indicators for measuring repairable systems. It is related to the failure rate and repair rate of components, and it can provide important data for the detection of large-scale equipment and the formulation of repair strategies.

Assuming that the PMSM drive system is alternating between normal operation and downtime for maintenance, and the failure rate $\lambda_i$ and maintenance rate $\mu_i$ of each component, as well as the failure time and the completion time of maintenance, all follow the exponential distribution [23]. The electric drive system is scheduled for regular maintenance by staff, as well as emergency repairs in the event of a failure to restore its performance. The mean time between failures (MTBF) of component $i$ can be expressed as:

$$MTBF_i = \frac{1}{\lambda_i} \tag{6}$$

The mean time to repair (MTTR) of component $i$ is:

$$MTTR_i = \frac{1}{\mu_i} \tag{7}$$

The mean cycle time (MCT) of component $i$ is:

$$MCT_i = MTBF_i + MTTR_i \tag{8}$$

Then the steady-state availability of component $i$, namely the average working probability, can be expressed as:

$$A_i = \frac{MTBF_i}{MCT_i} = \frac{\mu_i}{\lambda_i + \mu_i} \tag{9}$$

In the formula, $MTBF_i, MTTR_i, MCT_i$ are in units of h.

### 2.2.2. Component failure importance

In order to identify the weak links in the electric aircraft drive system effectively, measuring the importance of each component of the system by failure importance is indispensable. The failure importance describes the influence to which the change in the failure rate of a single component of the system affects the overall reliability of the system and reflects the contribution of the component to the reliability of the system [24]. When the failure rate of component $i$ is taken as $\lambda_{ci}$ and $5\lambda_{ci}$, respectively, that is, when the failure rate of component $i$ increases by five times, the influence of the change of the component failure rate on the system reliability is calculated.

The average number of repairable system failures caused by component $i$ failures is $N_{ci}$. The calculation formula is as follows:

$$N_{ci} = P_{ci}(1) \cdot 5\lambda_{ci} \cdot (P_{ri} - P_{ri}') \tag{10}$$

In the formula, $P_{ci}(1)$ is the probability of the ith component successful operation state; $\lambda_{ci}$ is the failure rate of components; $P_{ri}$ is the probability of successful operation of the system when the component failure rate is taken $\lambda_{ci}$; $P_{ri}'$ is the probability of successful operation of the system when the component failure rate is taken $5\lambda_{ci}$.

The component failure importance $I_g(i)$ is the ratio of the average number of system failures caused by component $i$ failures to the average number of system failures $N_c$. The calculation formula is as follows:

$$I_g(i) = \frac{N_{ci}}{N_c} \tag{11}$$

### 2.2.3. Evaluation index weight

The CRITIC method is an objective weight assignment method based on evaluation indicators. As the comparative strength of samples and the conflict between indicators are fully considered, the calculation results are more objective and reasonable [9]. Suppose a system has $m$ samples and $n$ indicators, $x_{ij}$ represents the value of the $j$th evaluation index of the $i$th sample, and the evaluation matrix is shown in equation (12):

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{12}$$

The calculation steps of objective weighting are as follows:

**Step1** Normalize $x_{ij}$, $x_{ij}'$ represents the value of the $j$th evaluation index of the $i$th sample, and the standardized matrix $X'$ is obtained. The calculation formula is as follows:

$$x_{ij}' = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})} \tag{13}$$

In the formula, $\max_j(x_{ij})$ is the maximum value of $x_{ij}$ in the $j$th evaluation index; $\min_j(x_{ij})$ is the minimum value of $x_{ij}$ in the $j$th evaluation index.

**Step2** Solve the index mean $\overline{x_j}$ and index standard deviation $\sigma_j$:

$$\overline{x_j} = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \tag{14}$$

$$\sigma_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_{ij} - \overline{x_j})^2} \tag{15}$$

**Step3** Calculate the correlation coefficient $\rho_{ij}$ and information quantity $E_j$:

$$\rho_{ij} = \frac{cov(x_k', x_l')}{\sigma_k \sigma_j}, k=1,2,\ldots,n \tag{16}$$

$$E_j = \frac{\sigma_j}{\overline{x_j}}\sum_{k=1}^{n}(1-\rho_{ij}), j=1,2,\ldots,n \tag{17}$$

In the formula, $\rho_{ij}$ is the correlation coefficient between the ith index and the $j$th index after the matrix standardization, $cov(x_k', x_l')$ represents the covariance between the $k$th index and the $l$th index after matrix normalization.

**Step4** Calculate the objective weight:

$$\omega_j = \frac{E_j}{\sum_{i=1}^{n}E_j}, j=1,2,\cdots,n \tag{18}$$

In the formula, $\omega_j$ is the objective weight of the $j$th index.

## 3. Results

The component importance of a general electric aircraft drive system is analyzed, and a SIMULINK simulation model is established according to the GO diagram of the system. The electric drive system is composed of 15 repairable components. The success probability of $i$th component is $P_{ci}(1)$ and the failure rate is $\lambda_{ci}$. The probability of system success state is $P_r(1)$, and the equivalent failure rate is $\lambda_r$.

Component equivalent failure rate and maintenance rate are important parameters and calculation basis for component importance analysis. The failure rate is the change of component failure probability per unit time, and the main influencing element of maintenance rate is the component failure rate, which is crucial content for realizing specific system functions and ensuring system reliability. The equivalent failure rate can comprehensively measure the influence of all component states on the success probability of system output, it is an extension of component importance, which can provide a reference for system reliability analysis and improvement. The failure rate and maintenance rate of key components in the PMSM drive system are given according to the military standard (GJB/Z299C-2006), as shown in Table 1.

SIMULINK provides a comprehensive and efficient integrated environment, which can accomplish the modelling and simulation functions of a dynamic system for users. It is widely used in the complex simulation and design of automatic control principles and signal processing technology.

The package design of several common operators in SIMULINK is as below:

(1) Signal generator

The signal generator has no input, only output, so it can be simulated with Constant.

(2) Two-state unit

The operator has only two states, success or failure. The system model of the two-state unit is shown in Fig. 3 (a).

The embedded M-file of the Embedded MATLAB Fcn block is

function y=fcn(u)

A=u(1,:)'*u(2,:);

y=A(:,2)'

In the formula, u is the input of the GO operator, y is the output of the GO operator, and the middle operation expression is written according to the operation rules of the two-state unit.

(3) AND gate

The output signal state of AND gate is the maximum state value of the input signal flow. And gate modeling is shown in Fig. 3 (b).

The embedded M-file of the Embedded MATLAB Fcn block is

function y=fcn(u)

A=u(1,:)'*u(2,:);

y=[A(1) sum(sum(A))-A(1)]

(4) OR gate

The output signal state of OR gate is the minimum state value of the input signal flow. Or gate is shown in Fig. 3 (c):

The embedded M-file of the Embedded MATLAB Fcn block is

_Table 1. Failure rate and maintenance rate of key components_

| Component Name | Failure rate / h | Maintenance rate / h |
|---|---|---|
| Power source | 0.0000486 | 0.0001220 |
| DC/DC Converter | 0.0000700 | 0.0003860 |
| Voltage sensor | 0.0000390 | 0.0002130 |
| Amplifier | 0.0000032 | 0.0000206 |
| Bus current sensor | 0.0000420 | 0.0001880 |
| Three-phase output current sensor | 0.0000420 | 0.0001880 |
| Conditioning circuit | 0.0000059 | 0.0000202 |
| Filter | 0.0000170 | 0.0001010 |
| Hall sensor | 0.0000230 | 0.0001860 |
| Photoelectric encoder | 0.0000580 | 0.0003020 |
| DSP | 0.0000703 | 0.0003880 |
| Bus transceiver | 0.0000119 | 0.0001080 |
| Driver board | 0.0000210 | 0.0001980 |
| IGBT（Contain the capacitance plate） | 0.0000414 | 0.0003220 |
| PMSM | 0.0000150 | 0.0001090 |
| AND gate | --- | --- |
| OR gate | --- | --- |

function y=fcn(u)

u=[u(3) 1-u(1)-u(3);u(4) 1-u(2)-u(4)];

A=u(1,:)'*u(2,:);

y=[1-sum(sum(A));sum(sum(A))-A(4)]'

a)



two-state unit-Embedded
MATLAB Function

b)



AND gate-Embedded
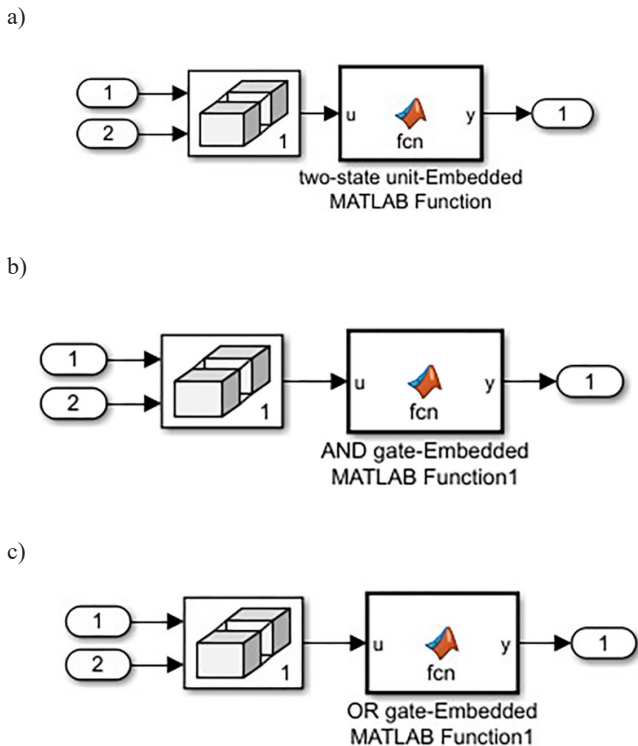MATLAB Function1

c)



OR gate-Embedded
MATLAB Function1

_Fig. 3. Encapsulation model. (a) is two-state unit-Embedded MATLAB Function. (b) is AND gate-Embedded MATLAB Function. (c) is OR gate-Embedded MATLAB Function_

Fig. 4 replaces the GO operators in Fig. 2 directly with the simulation model. According to the GO diagram of the electric aircraft drive system in Fig. 2, each operator in the GO diagram is encapsulated according to the packaging model. The state probability matrix of each operator is input through the Constant module in SIMULINK Library Browser. Finally, the SIMULINK reliability analysis simulation model of the electric aircraft drive system is shown in Fig. 4. The operation rules of these operators are described above in the article. The reliability data of components in Table 1 is input into the Constant module, and the reliability output data at each signal flow can be obtained by running the simulation.

Step1 Calculate the steady-state availability of system components:
Substitute the reliability data of system components in Table 1 into Equations (6) ~ (9), and the calculated steady-state availability is shown in Table 2.

_Table 2. Steady-state availability of key components._

| Number | Type | Component Name | Steady-state Availability |
|---|---|---|---|
| 1 | 5 | Power source | 0.7151231 |
| 2 | 1 | DC/DC Converter | 0.8464912 |
| 3 | 1 | Voltage sensor | 0.8452381 |
| 4,6,8,10,12 | 1 | Amplifier | 0.8655462 |
| 5 | 1 | Bus current sensor | 0.8173913 |
| 7,9,11 | 5 | Three-phase output current sensor | 0.8173913 |
| 14 | 1 | Conditioning circuit | 0.7739464 |
| 15 | 1 | Filter | 0.8559322 |
| 16 | 1 | Hall sensor | 0.8899522 |
| 17 | 1 | Photoelectric encoder | 0.8388889 |
| 20 | 1 | DSP | 0.8466070 |
| 21 | 1 | Bus transceiver | 0.9007506 |
| 22,24,26 | 1 | Driver board | 0.9041096 |
| 23,25,27 | 1 | IGBT（Contain the capacitance plate） | 0.8860759 |
| 29 | 1 | PMSM | 0.8790323 |
| 13,19,28 | 10 | AND gate | |
| 18 | 2 | OR gate | |

Step2 Calculate the component failure importance:
Bring reliability data into SIMULINK simulation models, and the probability of system success state $P_r(1)$ is calculated to be 0.99816610, so the equivalent failure rate $\lambda_r$ is 0.00183390. Substituting in equations (10) and (11), the failure importance of each component can be obtained in Table 3.

Step3 Calculate the weight of indicators
The electric drive system has a total of 15 components and 2 evaluation indexes. $x_{ij}$ represents the value of the $j$th evaluation index of the $i$th sample. The evaluation matrix can be obtained from equation (12).
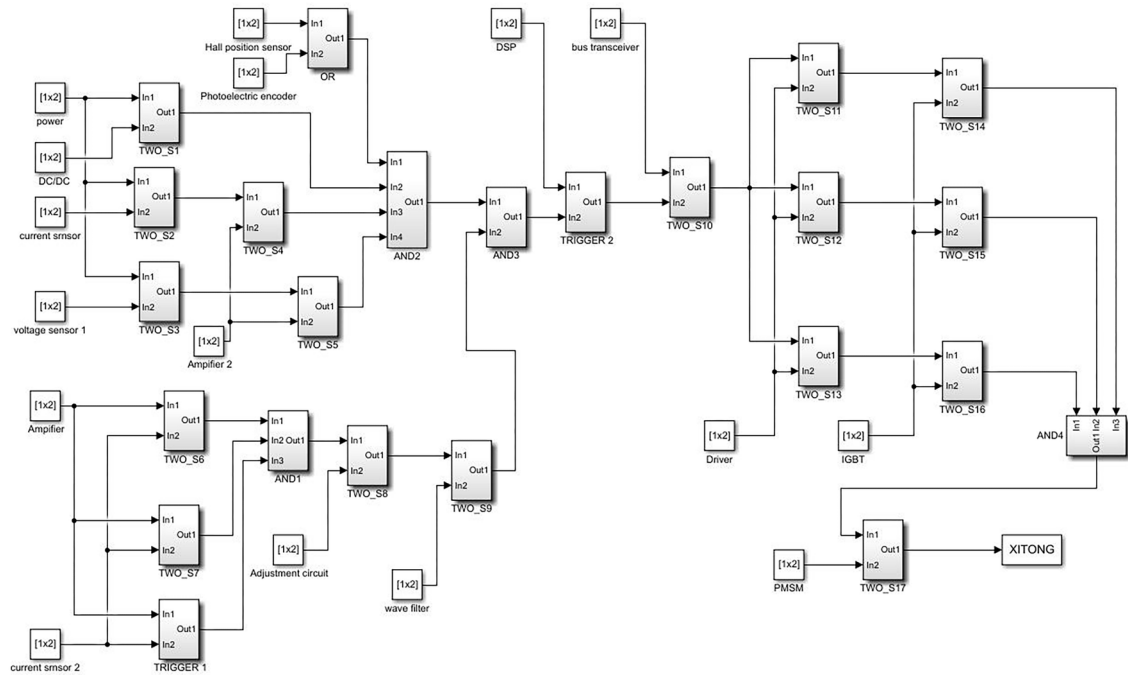
*Fig. 4. Simulation Model for Simulink Reliability Analysis of Electric Aircraft Drive System*

Table 3. *Failure importance of the key components*

| Component Name | System success probability (failure rate is $\lambda_{ci}$ ) | System success probability (failure rate is $5\lambda_{ci}$ ) | Average number of system failures caused by component faults | Failure importance |
|---|---|---|---|---|
| Power source | 0.9955 | 0.9890 | 0.0000016 | 0.0008617 |
| DC/DC Converter | 0.9948 | 0.9917 | 0.0000011 | 0.0005919 |
| Voltage sensor | 0.9945 | 0.9928 | 0.0000003 | 0.0001809 |
| Amplifier | 0.9943 | 0.9938 | 0.0000000 | 0.0000044 |
| Bus current sensor | 0.9946 | 0.9927 | 0.0000004 | 0.0002177 |
| Three-phase output current sensor | 0.9954 | 0.9892 | 0.0000013 | 0.0007103 |
| Conditioning circuit | 0.9944 | 0.9935 | 0.0000000 | 0.0000145 |
| Filter | 0.9947 | 0.9922 | 0.0000002 | 0.0001159 |
| Hall sensor | 0.9942 | 0.9942 | 0.0000000 | 0.0000000 |
| Photoelectric encoder | 0.9942 | 0.9942 | 0.0000000 | 0.0000000 |
| DSP | 0.9948 | 0.9917 | 0.0000011 | 0.0005944 |
| Bus transceiver | 0.9943 | 0.9938 | 0.0000000 | 0.0000162 |
| Driver board | 0.9944 | 0.9934 | 0.0000001 | 0.0000573 |
| IGBT (Contain the capacitance plate) | 0.9943 | 0.9937 | 0.0000001 | 0.0000678 |
| PMSM | 0.9942 | 0.9941 | 0.0000000 | 0.0000041 |
| AND gate | | | | |
| OR gate | | | | |

$$X = \begin{bmatrix} 0.7151231 & 0.0008617 \\ 0.8464912 & 0.0005919 \\ 0.8452381 & 0.0001809 \\ 0.8655462 & 0.0000044 \\ 0.8173913 & 0.0002177 \\ 0.8173913 & 0.0007103 \\ 0.7739464 & 0.0000145 \\ 0.8559322 & 0.0001159 \\ 0.8899522 & 0.0000000 \\ 0.8388889 & 0.0000000 \\ 0.8466070 & 0.0005944 \\ 0.9007506 & 0.0000162 \\ 0.9041096 & 0.0000573 \\ 0.8860759 & 0.0000678 \\ 0.8790323 & 0.0000041 \end{bmatrix}$$

Objective weighting calculation:

1. Normalized $x_{ij}$, and normalized matrix $X'$ is obtained by calculation of equation (13).

$$X' = \begin{bmatrix} 0.0000000 & 1.0000000 \\ 0.6951192 & 0.6868527 \\ 0.6884884 & 0.2098866 \\ 0.7959464 & 0.0050660 \\ 0.5411403 & 0.2526199 \\ 0.5411403 & 0.8243387 \\ 0.3112564 & 0.0168127 \\ 0.7450750 & 0.1345575 \\ 0.9250876 & 0.0000000 \\ 0.6548923 & 0.0000000 \\ 0.6957319 & 0.6897953 \\ 0.9822264 & 0.0188385 \\ 1.0000000 & 0.0664859 \\ 0.9045771 & 0.0786353 \\ 0.8673062 & 0.0047491 \end{bmatrix}$$

2. The mean value of the index $\overline{x'_j}$ is calculated from equation (14), and the standard deviation is $\sigma'_j$, which can be calculated from equation (15):

$$\overline{x'_j} = \frac{1}{15} \sum_{i=1}^{15} x'_{ij} \Rightarrow \overline{x'_1} = 0.6898658$$
$$\overline{x'_2} = 0.2659092$$
$$\sigma'_j = \sqrt{\frac{1}{15} \sum_{i=1}^{15} (x'_{ij} - \overline{x'_j})^2} \Rightarrow \sigma'_1 = 0.2573791$$

3. The correlation coefficient $\rho'_{ij}$ is calculated from equation (16), and the information amount $E'_j$ is calculated from equation (17).

$$\rho'_{ij} = -0.6127065$$

$$E'_j = \frac{\sigma'_j}{\overline{x'_j}} \sum_{k=1}^{15} (1 - \rho'_{ij}) \Rightarrow E'_1 = 2.0440027$$

$$E'_2 = 0.6016777$$

4. The objective weight $\omega_1$ of steady-state availability and the objective weight $\omega_2$ of failure importance are calculated from equation (18).

$$\omega_j = \frac{E'_j}{\sum_{j=1}^{2} E'_j} \Rightarrow \omega_1 = 0.2274189 \quad \omega_2 = 0.7725810$$
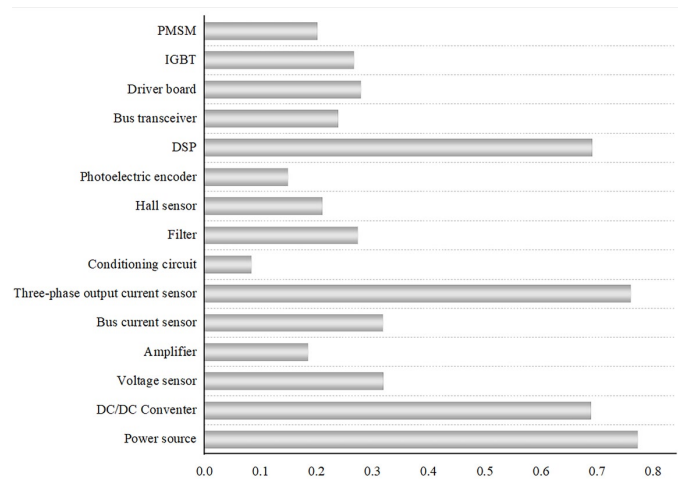


*Fig. 5. Weighted importance of key components.*

*Table 4. Weighted importance of components*

| Component Name | Steady-state availability | Failure importance | Component importance |
|---|---|---|---|
| Power source | 0.0000000 | 1.0000000 | 0.7725811 |
| DC/DC Converter | 0.6951192 | 0.6868527 | 0.6887327 |
| Voltage sensor | 0.6884884 | 0.2098866 | 0.3187297 |
| Amplifier | 0.7959464 | 0.0050660 | 0.1849272 |
| Bus current sensor | 0.5411403 | 0.2526199 | 0.3182349 |
| Three-phase output current sensor | 0.5411403 | 0.8243387 | 0.7599340 |
| Conditioning circuit | 0.3112564 | 0.0168127 | 0.0837748 |
| Filter | 0.7450750 | 0.1345575 | 0.2734007 |
| Hall sensor | 0.9250876 | 0.0000000 | 0.2103824 |
| Photoelectric encoder | 0.6548923 | 0.0000000 | 0.1489349 |
| DSP | 0.6957319 | 0.6897953 | 0.6911454 |
| Bus transceiver | 0.9822264 | 0.0188385 | 0.2379312 |
| Driver board | 1.0000000 | 0.0664859 | 0.2787847 |
| IGBT (Contain the capacitance plate) | 0.9045771 | 0.0786353 | 0.2664701 |
| PMSM | 0.8673062 | 0.0047491 | 0.2009109 |

Table 5.  Operator data of PMSM drive system after adding redundancy

| Number | Type | Component Name | Success rate ($10^{-5}$ /h) | failure rate ($10^{-5}$ /h) |
|---|---|---|---|---|
| 1~m | 5 | Power source | $1-\lambda$ | 4.86 |
| 3~n | 1 | DC/DC Converter | $1-\lambda$ | 7.00 |
| 5 | 1 | Voltage sensor | $1-\lambda$ | 3.90 |
| 6,8,10,12,14 | 1 | Amplifier | $1-\lambda$ | 0.32 |
| 7 | 1 | Bus current sensor | $1-\lambda$ | 4.20 |
| 9,11,13 | 5 | Three-phase output current sensor | $1-\lambda$ | 4.20 |
| 16 | 1 | Conditioning circuit | $1-\lambda$ | 0.59 |
| 17 | 1 | Filter | $1-\lambda$ | 1.70 |
| 18 | 1 | Hall sensor | $1-\lambda$ | 2.30 |
| 19~p | 1 | Photoelectric encoder | $1-\lambda$ | 5.80 |
| 23~o | 1 | DSP | $1-\lambda$ | 7.03 |
| 25 | 1 | Bus transceiver | $1-\lambda$ | 1.19 |
| 26,28,30 | 1 | Driver board | $1-\lambda$ | 2.10 |
| 27,29,31 | 1 | IGBT （Contain the capacitance plate） | $1-\lambda$ | 4.14 |
| 33~q | 1 | PMSM | $1-\lambda$ | 1.50 |
| 15,22,32 | 10 | AND gate | --- | --- |
| 2,4,20,21,24,34 | 2 | OR gate | --- | --- |
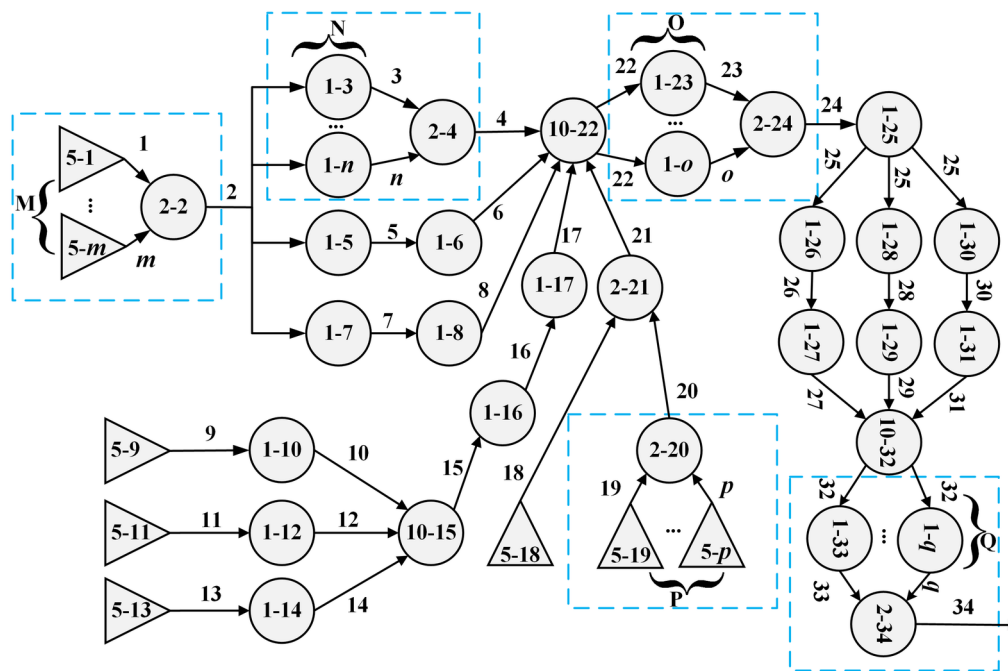


Fig. 6. GO diagram of PMSM system after adding redundancy

The objective weight of the steady-state availability of components is 0.2274189, and the objective weight of failure importance is 0.7725810. The importance of the key components obtained by comprehensive weighting is shown in Table 4. Drawing the importance of the components in Table 4 as a bar graph, the results can be analyzed more intuitively, as shown in Fig. 5.

## 4. Discussion

As we can see from Table 4 and Fig. 5, the failure of the power source, three-phase output current sensor, DSP and DC/DC converter will have a tremendous impact on the overall system. Accordingly, reliability optimization design should be carried out, such as increasing redundancy and selecting devices with higher reliability to enhance the reliability level of the electric aircraft drive system.

According to the previous analysis results, when the power source fails, it has the largest impact on the overall reliability of the system, so the redundant design of the power source is carried out. Since the system reliability is closely related to the component failure probability, the components with failure probability exceeding $5.0 \times 10^{-5}$/h, such as photoelectric encoder, DSP and DC/DC converter, are designed with redundancy to improve the system reliability. In addition, considering that PMSM plays a central role in the whole system as an executive component, redundant designed or replacement with a six-phase PMSM is required.

Table 6.  Operator data of PMSM drive system after adding redundancy

| System redundancy | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| System reliability | 0.999351295 | 0.999555078797355 | 0.999555091221017 | 0.999555091221825 |

In summary, the redundant design of the power source, photoelectric encoder, DSP, DC/DC converter and PMSM is carried out to enhance the output success probability of the electric drive system. Assuming that the redundancies of the power source, DC/DC converter, photoelectric encoder, DSP and PMSM are M, N, P, O, and Q, respectively, the corresponding signal flow are $1 \sim m$, $3 \sim n$, $19 \sim p$, $23 \sim o$, $33 \sim q$. The operator data is shown in Table 5, and the system GO diagram after adding the margin is shown in Fig. 6.

When M=N=P=O=Q=1, the reliability data of the signal flow 34 represent the reliability characteristics of the electric aircraft drive system with single redundancy. Since modern aviation technology system components generally adopt dual-redundant, triple-redundant or even quad-redundant configurations, let M=N=P=O=Q=x take integers from 1 to 4, respectively, and obtain system reliability data as shown in Table 6.

The reliability of the electric aircraft drive system increases substantially when the redundancy of power source, photoelectric encoder, DSP, DC/DC converter and PMSM are increased from 1 to 2. The electric aircraft drive system reliability varies little when the redundancy is increased from 2 to 4. Therefore, considering factors such as cost, volume, and weight, dual redundancy is a better choice to enhance system reliability. It can increase system reliability by 0.0204%.

## 5. Conclusions

This paper presents new research results and methods. According to the working characteristics of the electric aircraft drive system, a method of importance analysis of the key components of a repairable system based on the GO method is proposed. Based on the reliability simulation model of the electric aircraft drive system, the steady-state availability and fault importance of key components of the system are calculated. The objective weights of steady-state availability and fault importance of the system are determined by the CRITIC method. The importance of the key components is obtained by the weighted summation. Meanwhile, an idea of system redundancy design based on the importance of components is proposed to provide data support for the early design of the system. In the system operation stage, this method can be used to reasonably allocate inspection and maintenance resources, so as to ensure that the most crucial system units can operate normally.

The case verification and analysis results illustrate that the proposed method can comprehensively evaluate the vulnerability of the electric drive system, and provide an important basis for improving the reliability of the electric aircraft drive system. The method is more comprehensive and more reasonable than the evaluation of a single index.

In the actual evaluation of the importance of components, there are many influencing factors, such as the environment in which the system is located, the cause and degree of failure, and the level of repair personnel. These factors will affect the evaluation of the importance of components. Therefore, it is necessary to establish a more comprehensive model to evaluate the importance of components. In addition, the research object is set as a two-state system, and the components in the system also have several degraded working states during the transition from the successful state to the fault state. Therefore, follow-up research can be continued on reliability analysis and component importance analysis of polymorphic electronic systems.

## References

1. Chen DN, Wei X, Yao CY, Wang CL, Lv SJ. Continuous time T-S dynamic fault tree importance analysis method. Chinese Journal of Scientific Instrument 2020, 41(09): 232-241, http://doi.org/10.19650/j.cnki.cjsi.J2006516.
2. Cai C, Liu YQ, Zhang Y. Multi-level LSSC system reliability analysis based on go method. Statistics and Decision 2019, 35(12): 75-78, http://doi.org/10.13546/j.cnki.tjyjc.2019.12.018.
3. Chen YZ. Reliability analysis of vulnerable parts in scraper conveyor based on fault tree method. Coal Mining Machinery, 2019, 40(01): 144-145, http://doi.org/10.13436/j.mkjx.201901051.
4. Chen YQ, Xu TX, Li ZQ, Li HJ. Dynamic reliability analysis of complex polymorphic systems based on evidence go method. Systems engineering and electronics 2020, 42(01): 230-237, http://doi.org/10.3969/j.issn.1001-506X.2020.01.31.
5. Fu T, Wang D, Fan XY, Huang QH. Component Importance and Interdependence Analysis for Transmission, Distribution and Communication Systems. CSEE Journal of Power and Energy Systems, vol. 8, no. 2, 488-498, 2022, http://doi.org/10.17775/CSEEJPES.2020.05520.
6. Fan DM, Ren Y, Liu LL, et al. Algorithm based-on dynamic Bayesian networks for repairable GO methodology model. Journal of Beijing University of Aeronautics and Astronautics 2015, 41(11): 2166-2176, http://doi.org/10.13700/j.bh.1001-5965.2014.0767.
7. Gohardani AS, Doulgeris G, Singh R. Challenges of future aircraft propulsion: A review of distributed propulsion technology and its potential application for the all electric commercial aircraft. Progress in Aerospace Sciences 2011, 47(5): 369-391, https://doi.org/10.1016/j.paerosci.2010.09.001.
8. Huang J, Yang FT. Development and challenges of new energy electric aircraft. Acta Aeronautica et Astronautica Sinica 2016, 37 (01): 57-68, https://doi.org/10.7527/S1000-6893.2015.0274.
9. Hu T, Wang D, Sun Y, Huang ZY, Jiang LT. Air combat threat assessment based on improved CRITIC-LRA and grey approximation ideal solution sorting method. Acta Armamentarii 2020, 41(12): 2561-2569, https://doi.org/10.3969/j.issn.1000-1093.2020.12.022.
10. Jia BH, Guo T, Lu X, Li LY. Reliability analysis of typical components of civil aircraft. Aircraft Design 2020, 40(02): 64-68, https://doi.org/10.19555/j.cnki.1673-4599.2020.02.014.
11. Luo CK, Chen YX, He Z, Li Y, Zhang YM. Evaluation method of contribution rate of aviation equipment architecture based on fault tree analysis. Journal of National University of Defense Technology 2021, 43(01): 155-162, https://doi.org/10.11887/j.cn.202101020.
12. Li ZQ, Xu TX, Gu JY, An J, Liu YD. Reliability analysis method of a certain type of missile control system fused with uncertain information. System Engineering and Electronic Technology 2017, 39(12): 2869-2876, https://doi.org/10.3969/j.issn.1001-506X.2017.12.34.
13. Li JK, Wang Y, Wang BM. Reliability of spoiler system of an electric aircraft based on the moment go method. Acta Aeronautica Sinica 2021, 42(03): 62-70, https://doi.org/10.7527/S1000-6893.2020.23945.
14. Ma Z, Zhou HY, Li XQ, He XP. Research on the reliability of electric vehicle electric power steering system based on model-driven

architecture. Automotive Technology 2020(12): 36-42, https://doi.org/10.19620/j.cnki.1000-3703.20200057.

15. Miziuła P, Navarro J. Birnbaum Importance Measure for Reliability Systems With Dependent Components. IEEE Transactions on Reliability 2019, vol. 68, no. 2, 439-450, https://doi.org/10.1109/TR.2019.2895400.

16. Nguyen KA, Do P, Grall A. Condition-based maintenance for multi-component systems using importance measure and predictive information. International Journal of Systems Science: Operations& Logistics 2014, 1(4): 228-245, https://doi.org/10.1080/23302674.2014.983582.

17. Peng H, Coit DW, Feng Q. Component Reliability Criticality or Importance Measures for Systems With Degrading Components. IEEE Transactions on Reliability 2012, vol. 61, no. 1, 4-12, https://doi.org/10.1109/TR.2011.2182256.

18. Steiner HJ, Vratny PC, Gologan C, et al. Optimum number of engines for transport aircraft employing electrically powered distributed propulsion. CEAS Aeronautical Journa1 2015, 5(2): 157-170, https://doi.org/10.1007/s13272-013-0096-6.

19. Si SB, Yang L, Cai ZQ, Dui HY. Study on calculation method of comprehensive importance of two-state system components. Journal of Northwestern Polytechnical University 2011, 29(06): 939-947, https://doi.org/ 10.3969/j.issn.1000-2758.2011.06.021.

20. Scherb A, Garre L, Straub D. Evaluating component importance and reliability of power transmission networks subject to windstorms: methodology and application to the nordic grid. Reliability Engineering and System Safety 2019, 191: 106517-106517, https://doi.org/10.1016/j.ress.2019.106517.

21. Shen ZP, Huang XR. GO method principle and application: an analysis method of system reliability. Tsinghua University Press: Beijing, China.

22. Wang Y, Fu SS, Wu B, Huang JH, Wei XY. Towards optimal recovery scheduling for dynamic resilience of networked infrastructure. Journal of Systems Engineering and Electronics 2018, 29(05): 995-1008, https://doi.org/10.21629/JSEE.2018.05.11.

23. Wang XY, Fan QQ. Analysis of the importance of urban rail transit vehicle failure based on FMECA method. Urban Rail Transit Research 2020, 23(07): 121-124. https://doi.org/10.16037/j.1007-869x.2020.07.025.

24. Wang Z, Bao CY. The application of go method in the reliability analysis of YAG laser system. Journal of Tsinghua University (Natural Science Edition) 2007(03): 377-380, https://doi.org/10.3321/j.issn:1000-0054.2007.03.019.

25. Xue CG, Gu Y, Cao WJ, Cao HW. Optimization of opportunistic maintenance strategy for multi-component importance systems. Mechanical Design and Manufacturing 2022(02): 116-119+125, https://doi.org/10.19356/j.cnki.1001-3997.20211116.026.

26. Yang WS, Wang W, Mi GJ, Mao XJ, Yan JX, Zhong GS. Application of go method in reliability analysis of non-water-cooled all-solid-state laser system. Laser and Infrared 2010, 40(05): 479-483, https://doi.org/10.3969/j.issn.1001-5078.2010.05.007.

27. Yao AL, Huang LL, Xu TL. Reliability analysis of gas transmission station based on go method. Acta Petrolei Sinica 2016, 37(05): 688-694, https://doi.org/10.7623/syxb201605013.

28. Zhou K, Ding JY, Chen JY, Tao WW. Reliability and component importance analysis of substation automation system in multi-state mode. Electric Power Components and Systems 2019, 47(6-7): 589-604, https://doi.org/10.1080/15325008.2019.1602686.

## Appendix A

| Component Name | Failure rate / $h$ | Maintenance rate / $h$ | Steady-state availability | Failure importance | Weighted steady-state availability | Weighted failure importance | Weighted component importance |
|---|---|---|---|---|---|---|---|
| Power source | 0.0000486 | 0.0001220 | 0.7151231 | 0.0008617 | 0.0000000 | 1.0000000 | 0.7725811 |
| DC/DC Converter | 0.0000700 | 0.0003860 | 0.8464912 | 0.0005919 | 0.6951192 | 0.6868527 | 0.6887327 |
| Voltage sensor | 0.0000390 | 0.0002130 | 0.8452381 | 0.0001809 | 0.6884884 | 0.2098866 | 0.3187297 |
| Amplifier | 0.0000032 | 0.0000206 | 0.8655462 | 0.0000044 | 0.7959464 | 0.0050660 | 0.1849272 |
| Bus current sensor | 0.0000420 | 0.0001880 | 0.8173913 | 0.0002177 | 0.5411403 | 0.2526199 | 0.3182349 |
| Three-phase output current sensor | 0.0000420 | 0.0001880 | 0.8173913 | 0.0007103 | 0.5411403 | 0.8243387 | 0.7599340 |
| Conditioning circuit | 0.0000059 | 0.0000202 | 0.7739464 | 0.0000145 | 0.3112564 | 0.0168127 | 0.0837748 |
| Filter | 0.0000170 | 0.0001010 | 0.8559322 | 0.0001159 | 0.7450750 | 0.1345575 | 0.2734007 |
| Hall sensor | 0.0000230 | 0.0001860 | 0.8899522 | 0.0000000 | 0.9250876 | 0.0000000 | 0.2103824 |
| Photoelectric encoder | 0.0000580 | 0.0003020 | 0.8388889 | 0.0000000 | 0.6548923 | 0.0000000 | 0.1489349 |
| DSP | 0.0000703 | 0.0003880 | 0.8466070 | 0.0005944 | 0.6957319 | 0.6897953 | 0.6911454 |
| Bus transceiver | 0.0000119 | 0.0001080 | 0.9007506 | 0.0000162 | 0.9822264 | 0.0188385 | 0.2379312 |
| Driver board | 0.0000210 | 0.0001980 | 0.9041096 | 0.0000573 | 1.0000000 | 0.0664859 | 0.2787847 |
| IGBT (Contain the capacitance plate) | 0.0000414 | 0.0003220 | 0.8860759 | 0.0000678 | 0.9045771 | 0.0786353 | 0.2664701 |
| PMSM | 0.0000150 | 0.0001090 | 0.8790323 | 0.0000041 | 0.8673062 | 0.0047491 | 0.2009109 |
| AND gate | --- | --- | | | | | |
| OR gate | --- | --- | | | | | |
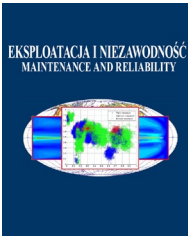
# An attempt at applying machine learning in diagnosing marine ship engine turbochargers

**Indexed by:**
Web of Science Group

Andrzej Adamkiewicz[a], Piotr Nikończuk[b]

[a] Maritime University of Szczecin, Faculty of Mechanical Engineering, ul. Wały Chrobrego 1-2, 70-500 Szczecin, Poland
[b] West Pomeranian University of Technology, Faculty of Maritime Technology and Transport, al. Piastów 17, 70-310 Szczecin, Poland

## Highlights

- Machine learning simplified the decision to renew the turbocharging system of the marine engine.

- The set of controlled parameters was minimized for the needs of diagnostic relationships.

- Ease of making maintenance decisions based on the maintenance requirement index.

- The results were verified with experimental data from engine tests of two types of turbochargers.

## Abstract

The article presents a diagnosis of turbochargers in the supercharging systems of marine engines in terms of maintenance decisions. The efficiency of turbocharger rotating machines was defined. The operating parameters of turbocharging systems used to monitor the correct operation and diagnose turbochargers were identified. A parametric diagnostic test was performed. Relationships between parameters for use in machine learning were selected. Their credibility was confirmed by the results of the parametric test of the turbocharger system and the main engine, verified by the coefficient of determination. A particularly good fit of the describing functions was confirmed. As determinants of the technical condition of a turbocharger, the relationship between the rotational speed of the engine shaft, the turbocharger rotor assembly and the charging air pressure was assumed. In the process of machine learning, relationships were created between the rotational speed of the engine shaft and the boost pressure, and the indicator of the need for maintenance. The accuracy of the maintenance decisions was confirmed by trends in changes in the efficiency of compressors.

## Keywords

machine learning, compressor diagnosis, marine ship engine, operational decision, neural network.

## 1. Introduction

Diesel engines dominate in 95% of the main propulsion of modern cargo sea vessels. As a result of normal wear and tear, as well as random interactions, they are subject to evolutionary degradation of their technical condition. Effective engine operation requires a reliable identification of the causes of wear of its systems and components, especially accelerated wear, of varying degrees of importance. Important systems, the failure or damage of which poses a threat to the performance of the transport task by the ship, include: the piston-cylinder system, the crank system and the turbocharging system, which, according to [31], causes 24.7% of all engine damage.

The analysis of the impact of the degradation of the technical condition of the turbocharging system components on the operating parameters, economy and reliability of marine engines was carried out in [4] with the use of the Konsberg turbocharged marine engine simulator. The reliability analysis of such a system was performed in [3]. The reliability model of a ship engine cooperating with a turbocharger was considered in [22], using the Weibull distribution for this purpose. The need to recognize the real effects of limiting the ef-

ficiency of rotating machinery and possible failure of the turbocharger in the turbocharging system of the marine engine was obtained based on the results of simulation tests described in [11]. The ecological consequences of the operation of a malfunctioning two-stroke engine turbocharger regarding NOx emissions are included in the complete analysis presented in the paper [13]. The publication [30] presents the results of an extended plan of experiments on a turbocharged four-stroke marine engine. By modifying the engine air intake manifold for additional compressed air supply, it was possible to increase the air pressure downstream of the turbocharger compressor, thereby generating unstable operation. The measurement results were used to determine the range of the unstable operation field on the compressor characteristics and the form of instability at any point in the engine operating field. The paper [20] presents a mathematical model constituting the basis of a simulation model of steady and transient operation of a turbocharged, slow-speed compression-ignition engine. Particular attention was paid to the study of the stability and measurement availability in difficult conditions, such as difficulties in cleaning the interscapular channels. In the work [28], a systematic analysis of friction losses in the turbocharger bearings was carried out.

---

(*) Corresponding author.
E-mail addresses: A. Adamkiewicz (ORCID: 0000-0002-8314-0660): a.adamkiewicz@pm.szczecin.pl, P. Nikończuk (ORCID: 0000-0002-7809-7653): piotr.nikonczuk@zut.edu.pl

Figure 1 shows the contaminated rotor of the ABB TPL 67-C type 6L50 DFDE turbocharger radial compressor, with salt deposits on the rotor blades emphasizing the contours of the impressions of the disassembled diffuser blades around the rotor (photo a) and a view of the rim of contaminated turbine expansion devices, after exceeding the resource working hours of the turbocharger by over 2000 (photo b) of the turbine powered by marine fuel combustion products.
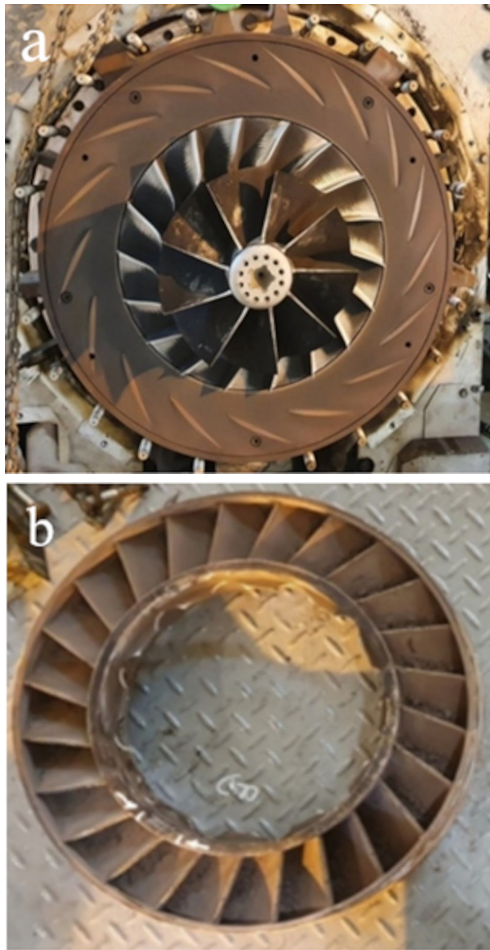


*Fig. 1. View of flow canals of a rotor in a radial compressor (a) with blades of turbine expansion instruments (b) of the ABB TPL 67-C type compressor contaminated during a normal operation of a marine engine turbocharging system*

The channels between the compressor blades are covered with contaminants contained in the aspirated sea spray and oil vapours passing through the seals between the compressor rotor and the bearings. Hard deposits of sea salt mixed with components of exhaust fumes and industrial dust in coastal areas change the surface condition of the flow channels, the geometry of the blades and, consequently, the characteristics of the machines. By increasing the imbalance of the rotor assembly, the deterioration of the technical condition of the rotor blades contributes to an increase in rotor mass and vibration of the rotor shaft and accelerates bearing wear. The likelihood of pumping increases, which precludes further operation of the turbocharging system. Figure 2 shows the view of the rim blades of the turbine expansion devices of the ABB TPL 67-C turbine with discoloration of tar deposits, carbon deposits from incomplete and incomplete fuel combustion and oil vapours from leaks in the lubrication system. The confrontation of the examples of the surface condition of the blades of the expansion

devices shown in Figures 1b and 2 confirms the possibility of large differences in the properties of the deposits on the blades at various stages of operation, their random thicknesses and distributions determining the quality and nature of energy conversion processes in the channels between the turbine blades, and consequently the efficiency compressors and turbines.



*Fig. 2. Blade rim of expansion instruments of a turbocharger of the ABB TPL 67-C type contaminated with deposits of fuel combustion products*

Experimental studies for the real 3AL25 / 30 four-stroke marine engine are presented in [31]. Based on the results of an active experiment on a technical scale, the share of contaminants and sediments of individual elements of the turbocharging system in the impact on the engine operating parameters was given, as shown in Figure 3. Contamination concerns the compressor flow channels in 56%, the turbine flow channels 22%, the air cooler 11 %, air filters 6%, compressor and turbine rotor channel cleaning systems 4%, other elements 1%.

In practice, the number of measured parameters of the turbocharging system operation does not provide sufficient information about the quality of energy conversion processes to determine the efficiency of rotating machines, and, consequently, about the need to perform services that reconstruct the technical condition of the inter-vane rotating machines channels. The technical condition of the surface of the blades has a direct impact on the efficiency of the compressor and turbine. However, to determine them, it is necessary to know the operating parameters measured during the operation of the engine and turbocharger.
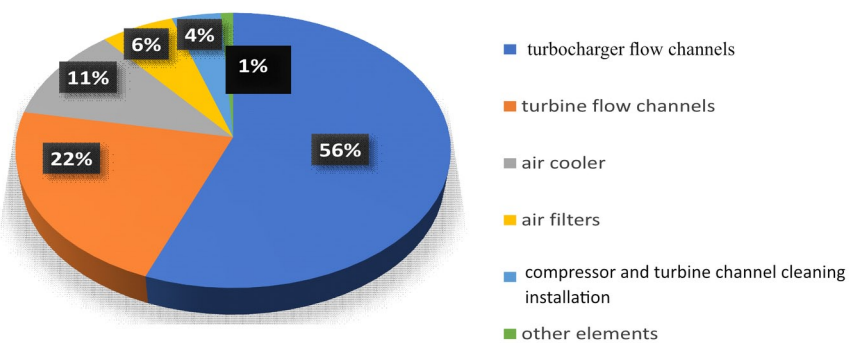


*Fig. 3. Contribution of contaminants and sludge to the effect on engine performance*

The article [6] presents the concept of diagnosing the exhaust tract for monitoring the state of the turbocharger of the ship's main engine. The flue gas flow rate indicator was proposed, characterizing the flue gas volume flow and isentropic efficiency, as two dimensionless

indicators for assessing the technical condition of the turbocharger. Moreover, the non-linear relationship between these two indicators and the measurable parameters of the exhaust tract of the turbocharger was investigated. A novel method for assessing the degradation of turbocharger efficiency has been defined. The paper [17] presents the influence of the surface roughness of the blades on the compressor efficiency and noise emission in the turbocharger of the diesel engine, while the paper [18] presents the results of numerical processing of the results of the acoustic signal parameters measurements in the turbocharger. The turbocharger condition monitoring system, especially for passenger vehicles, is presented in [14]. Possible malfunctions of turbochargers with their cause-effect relationships and frequency of occurrence were investigated. Compressor degradation (due to contamination with oil deposits and combustion product deposits, for example) was identified as the only malfunction worthy of further investigation. An approximate model of compressor efficiency reduction because of tar contamination of the blades was proposed by means of a regression function related to the actual, clean, and degraded compressor characteristics data. Based on the results of the simulation of a diesel engine based on an approximate model, it was found that when the efficiency of the compressor is reduced, three symptoms appear: increased blade angle adjustment in the turbine with variable geometry blades, which leads to an increase in the rotational speed of the rotor and an increase in air temperature at the outlet from the compressor. Limited diagnostics of malfunctions and prediction of turbocharger failures by means of the analysis of the thrust load of the turbocharger thrust bearing are presented in [33].

The methods and diagnostic models described above are based on large groups of measurement or simulation data. Neither of these methods is a solution to a diagnostic task based on a limited set of measured engine and turbo performance parameters. There is no method of assessing the efficiency of the rotating machinery of a turbocharger, necessary to assess the quality of energy conversion in a ship engine, and, consequently, to undertake maintenance activities, or replace components with new ones.

Attempts to use determined diagnostic models in the operational verification diagnostics of turbochargers encounter difficulties due to the limited availability of measurement of air and exhaust gas parameters in the turbocharging system and difficult-to-measure disturbances from accompanying processes.

In technical issues where there is no data to solve the problem based on known physical dependencies, energy conservation equations, artificial intelligence methods are often used. These methods do not operate with differential equations but are usually optimized dependency rules in the case of fuzzy logic or for neural networks a black box with weight connections between neurons with specific activation functions. Creating neural networks and learning their dependencies to obtain decision information is also called machine learning.

For the diagnosis of turbochargers, many methods and algorithms are proposed, extending the possibilities of using motor diagnostics [32]. There are known applications of fuzzy logic to control a turbocharger [10] as well as the identification of turbocharger operating parameters based on vibroacoustic signals [29] and the analysis of the vibration spectrum to identify damages [7]. There are also proposals to use neural networks to identify faults based on vibroacoustic signals from compressors [8] and piston internal combustion engines [25]. Machine learning has also been proposed to predict both damage to compressor blades [26] as well as compressor shutdown times [12]. The topic of prediction was also discussed in connection with forecasting operating parameters and potential inefficiencies of turbines [23] and predicting maintenance / servicing of turbochargers [16]. Predicting the need to repair car compressors using the service records and registered vehicle data is presented in [24]. The methods and results of the application of supervised machine learning techniques to the task of predicting the need for repair of air compressors in commercial trucks

and buses are described. The predictive models come from recorded on-board data collected during workshop visits. These data were collected over three years based on the performance of a large number of vehicles.

In the further part of the article, the authors present the concept of introducing the service necessity indicator $M$, which classifies a turbocharger for servicing. The value of the indicator is estimated based on the measured values of the engine speed and the boost air pressure from the turbocharger. An attempt was made to create a neural classifier to which known neural networks with feed-forward backpropagation were applied. For the computational work, measurement data from a marine engine operated in standard conditions were used, and a graphical interface for teaching *nntool* neural networks in Matlab was used.

## 2. Problem formulation

Deterioration of compressor and turbine blade surfaces, changes the aerodynamics of the air and exhaust flow, reducing the flow capacity of the inter-blade canals of the rotating machines, which is reflected in [17]:

– change in the rotor unit speed;
– decrease of efficiency of turbocharger and engine machines;
– increase of engine fuel consumption,
– generation of vibrations resulting from the unbalance of the turbocharger rotor unit;

with simultaneous change of thermal and flow parameters of air and exhaust gases in the turbocharging system. The turbocharging system of the engine consists of a turbocharger with an air filter, noise silencer and charging air cooler. Fig. 4 shows a diagram of the system with marked control planes.
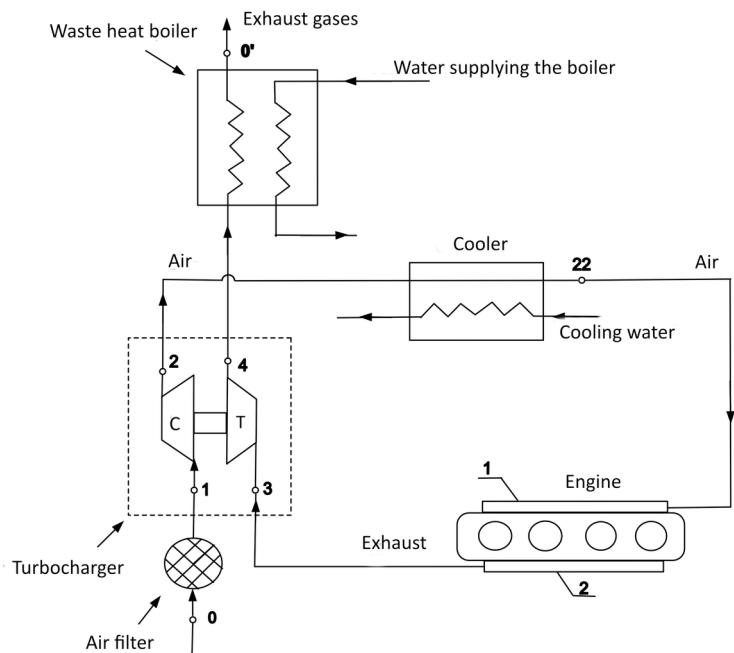


*Fig. 4. Diagram of the turbocharging system for a marine diesel engine: 1 – charging air reservoir; 2 – exhaust gas collector*

This paper studies a turbocharger consisting of a rotor unit mounted on bearings embedded in a hull with an exhaust gas collector. The rotor unit includes a radial compressor with straight blades and a rim of turbine blades, shafts connected by couplings, supported in bearing seats with thrust axial and sealing rings. A diagram of the turbocharger rotor unit with labelled control planes for the needs of a mathematical model is shown in Figure 5.
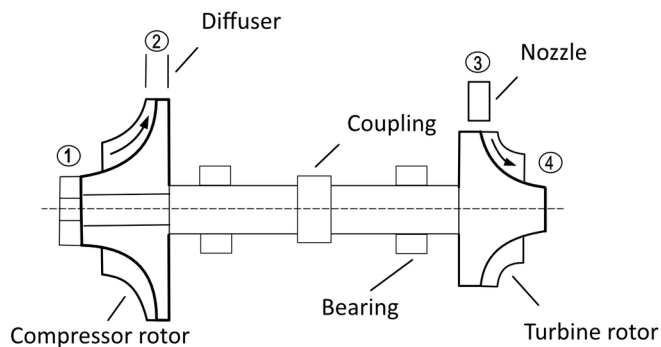
Fig. 5. Diagram of a turbocharger rotor unit with control plane numbers

Standard condition monitoring systems of turbocharging systems used on ships do not ensure the determination of compressor and turbine efficiency based on the results of current measurements of operational parameters. Not all necessary parameters of thermodynamic processes can be reliably accessed by measurements. If the key operational issue is to maintain the turbocharging system of the engine in the condition of the highest possible efficiency which does not decrease during its operation, it is necessary to control systematically the relationships between operational parameters in confrontation with admissible and limiting parameters. Operational parameters used for the evaluation of correctness of energy transformations in the turbocharger should serve as an argument for formulating service conclusions. However, the condition of task realization is the availability of operational parameters of the turbocharger subunits determining the methods of decision making or application of machine learning.

## 3. Efficiency of turbocharger rotating machines

In the mathematical model built, it was assumed that the measure of the quality of the air compression and exhaust gas expansion process is the isentropic efficiency of the compressor and turbine. The compression process was considered to be an adiabatic irreversible one with increasing entropy due to internal friction in the flow canals of the compressor. Figure 6 shows the air compression process in enthalpy-entropy coordinates, from the state of $1*(p*1, t*1)$ at the compressor inlet, to the state of $2*(p*2, t*2)$ at the compressor air outlet.
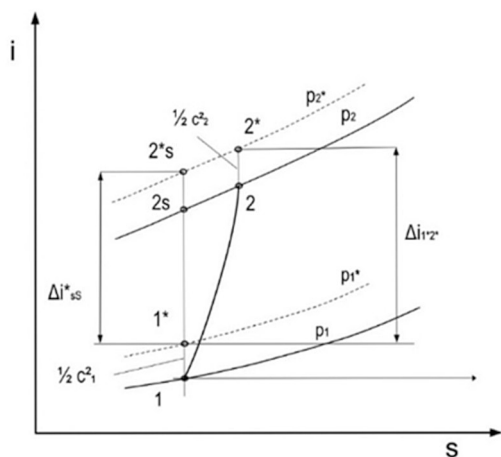


Fig. 6. The air compression process in enthalpy-entropy coordinates

According to the notational convention used in the model, the isentropic efficiency of the compressor $\eta^*_c$ is defined as the ratio of the enthalpy gain of isentropic compression $\Delta i^*_{sC}$ to the work of the actual compression $l^*_C = \Delta i^*_{12}$

$$\eta^*_C = \frac{\Delta i^*_{sC}}{\Delta i^*_{12}} = \frac{\Delta i^*_{sC}}{l^*_{C12}} = \frac{i^*_{2s} - i^*_1}{i^*_2 - i^*_1} \qquad (1)$$

Treating air as semi-perfect gas and assuming for $i^* = c_{pa}T^*$ the average specific heat of transformation $c_{p12} = idem$ as constant, relation (1) will take the form of:

$$\eta^*_C = \frac{T^*_{2s} - T^*_1}{T^*_2 - T^*_1} \qquad (2)$$

The isentropic efficiency of compression in the total parameters between state $1*(p*1,T*1)$ and state $2*(p*2,T*2)$ considers the conversion of kinetic energy of the air in state $2*$ into static pressure energy, thus increasing in the outlet diffuser, the pressure of the charging air. By substituting the equation of the isentropic conversion into formula (2), the relation for the isentropic efficiency of compression expressed in the air state parameters determined by measuring the total pressures and temperatures at the air inlet and outlet from the compressor was obtained

$$\eta^*_C = \frac{\left(\frac{p^*_2}{p^*_1}\right)^{\left(\frac{\kappa-1}{\kappa}\right)_a} - 1}{\left(\frac{T^*_2}{T^*_1}\right) - 1} \qquad (3)$$

where the properties of the air in transformation are described by the isentropic exponent for the air $\kappa a = 1.4$. The exhaust gas heat conversion process between the states $3*(p^*_3, T^*_3)$ and $4*(p^*_4, T^*_4)$ as the adiabatic transformation of the exhaust gas into the enthalpy-entropy coordinates is shown in Figure 7. Using the symbols shown in the graph (Fig. 7), the efficiency of the turbine $\eta_T$ is defined as the ratio of the actual available enthalpy drop in the turbine $\Delta i^*_{34}$ between states $3*(p*3,T*3)$ - $4*(p*4,T*4)$, internal operation of the turbine $1^* T_{34}$, to the isentropic available exhaust enthalpy drop $\Delta i^*_{sT}$ between states $3* - 4_s*$.

$$\eta^*_T = \frac{\Delta i^*_{34}}{\Delta i^*_{sT}} = \frac{l^*_{T34}}{\Delta i^*_{sT}} = \frac{i^*_3 - i^*_4}{i^*_3 - i^*_{4s}} \qquad (4)$$

From the physics of conversion, the internal work of the turbine is smaller than the disposed isentropic enthalpy drop in the turbine due to frictional losses in real processes. Treating the exhaust gas as semi-perfect and assuming for $i^* = c_{pg}T^*$ the average specific heat of transformation to be $c_{p12} = idem$ the defining relationship (4) is formulated in the following form:

$$\eta^*_T = \frac{\Delta i^*_{34}}{\Delta i^*_{sT}} = \frac{T^*_3 - T^*_4}{T^*_3 - T^*_{4s}} \qquad (5)$$

By substituting the isentropic equation into relation (5), we obtained the isentropic efficiency of the turbine expressed by measurable parameters of the state of exhaust gases before and after the turbine

$$\eta_T = \frac{1 - \left(\frac{T^*_4}{T^*_3}\right)}{1 - \left(\frac{p^*_4}{p^*_3}\right)^{\left(\frac{K-1}{K}\right)_g}} \qquad (6)$$

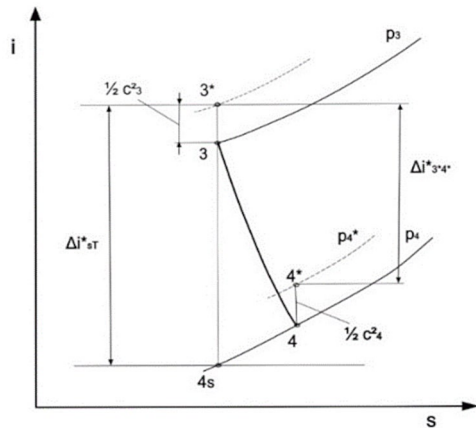where $K_g = 1,33$ is the average isentropic exponent for marine fuel exhaust.

*Fig. 7. Expansion process of exhaust gases in a single-stage turbine in en-thalpy-entropy coordinates*

Determination of compressor efficiency according to equation (3) and turbine efficiency according to relation (6) requires reliable measurements of air parameters in states 1 and 2 and exhaust parameters in states 3 and 4, in Figure 6.

Based on the experience of studies [31], [9], [15], [13] it was assumed that the informative character of operational diagnostics should ensure the assessment of the technical condition of a turbocharger based on values of observed operational parameters and enable to make important maintenance decisions based on a generalized parameter.

## 4. Parametric diagnostic test

Cleaning of inter-blade channels of rotating machines belongs to the important and characteristic turbocharger services. The prerequisite for the identification of the sought-after malfunction is the knowledge of the relationships between the state characteristics and the parameters of the thermal-fluid signal and operational diagnostics]. Identification of relations, with simultaneous recognition of operational parameter measurement availability has been performed based on the population of marine engine turbochargers of several types.

To recognize  measurability, variability, and relationships between the operational parameters of the turbocharging system, a passive operational experiment was carried out using a MAN type TCA55 turbocharger, cooperating with a Sulzer 6RTA48TB type main propulsion engine of 7 368 kW. The measurements were performed using standard control and measuring equipment [5]. Selected measurement results, in the form of courses of values of operational parameters of a turbocharger of the  MAN TCA55 type: charging pressure $p_2$ (Fig. 8), main engine rotational speed $n_{ME}$, turbocharger rotor speed $n_{TC}$ (Fig. 9), air temperature in the engine room $t_0$, oil behind the turbocharger bearing $t_{ol2}$ and charging air behind the cooler $t_{22}$ as a function of operation time (Fig. 10), represented by the numbers of consecutive observations, are shown in Figs. 8-10.
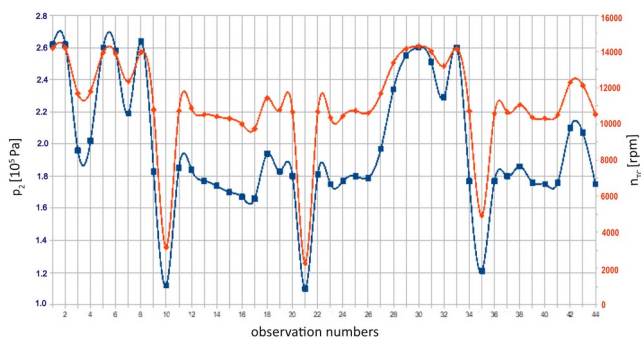


*Fig. 8. Variability of charging  pressure and turbocharger rotor speed as a function of observation numbers*

The expected parameter of the charge exchange system is the charging air pressure $p_2$, whose strong relationships with the turbocharger rotor unit speed and engine shaft speed are shown by the runs in Figures 9 and 10.
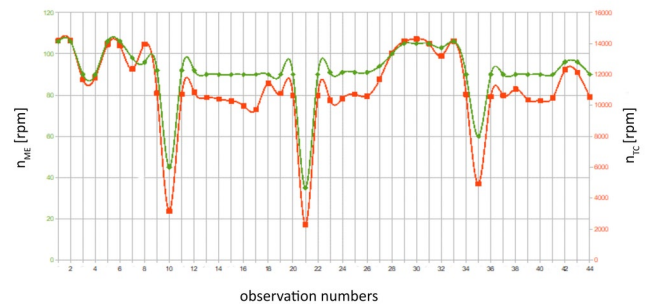


*Fig. 9. The main engine speed and turbocharger rotor speed runs as a function of observation numbers*

While the runs of the parameters in fig. 8 and 9 creating mutual relationships showed their diagnostic usefulness, the confrontation of the courses of air temperature in the engine room (temperature of sucked air at the beginning of compression), oil temperature behind the turbocharger bearing and charging air behind the cooler shown in fig. 10, are the parameters of different subunits and they do not show mutual relations.
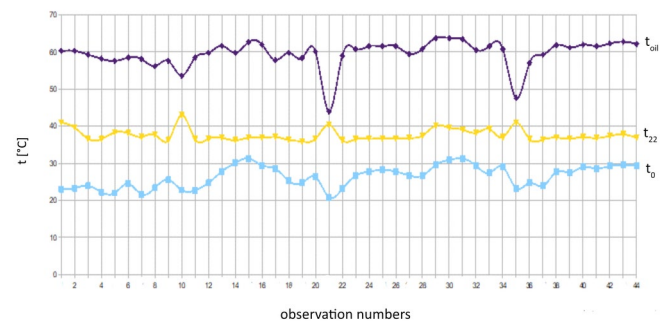


*Fig. 10. Engine room air temperature, oil temperature behind the turbocharger bearing and charging air temperature after the cooler as a function of observation numbers*

The value of  air temperature behind the cooler is kept constant at $t_{22} = 45.5°C = idem$. This temperature is regulated by a three-way valve in the LT (Low Temperature) cooling system, controlled by an air temperature sensor behind the cooler. Thus, the charging air temperature measured after the cooler is not a diagnostic parameter, it merely confirms  correct functioning of the system.

During the period of the passive operational experiment, none of the monitored parameters of the turbocharger operation reached the limiting or admissible value. To evaluate the  efficiency of the compressor based on formula (3) and  turbine efficiency based on formula (6) it is necessary to know the measured air temperature directly behind the compressor $T^*_2$ and the exhaust gas pressure before the turbine . In many types of turbochargers, as well as in the tested one, the manufacturers did not foresee the technical possibility to make this type of  measurements.

Therefore, the operational experiment was extended to include measurements performed on  turbocharging systems of four main engines of the  MAK M43 type with a turbocharger of the  ABB TPL 77 type, equipped with the possibility of measuring the air temperature behind the compressor $t_2$ [2]. During one observation, the following were measured: air temperature at the turbocharger inlet $t_1$, air temperature $t_2$ and air pressure $p_2$ after the compressor, air temperature at the air cooler outlet $t_{22}$, exhaust gas temperature before the turbine $t_3$

and after the turbine $t_4$, rotational speed of the turbocharger rotor unit $n_{TC}$, pressure drop at the air cooler $\Delta p_{2-22}$, and fuel delivery indicator at the fuel pump $x$. Measurement observations were made once a day and marked with consecutive numbers. The correctness of the measurements was verified by determining the quality measure of the model fit using the coefficient of determination $R^2$. Figure 11 presents graphically an example of air pressure after the compressors as a function of the turbocharger rotor speed of the four investigated engines [2]. The values of determination coefficients shown in the figure are in the range of 0.9-1.0, confirming their particularly good model fit and thus the diagnostic usefulness of the measured charging pressure.
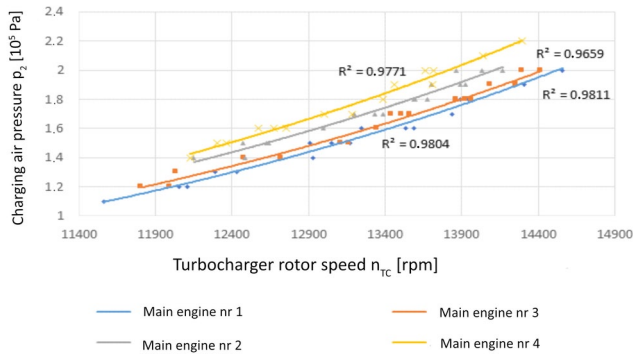


*Fig. 11. Air pressure after compressor $p_2$ versus rotor speed $n_{TC}$ of ABB TPL 77 turbocharger*

Using the results of measurements obtained during these observations, compressor efficiencies were calculated in accordance with equation (3) in various technical states of the surfaces of compressor inter-blade canals, before and after washing, before and after replac-

*Table 1.  Calculated compressor efficiencies*

| Engine number / Observation number | nr 1 $\eta_c$[%] | nr 2 $\eta_c$[%] | nr 3 $\eta_c$[%] | nr 4 $\eta_c$[%] |
|---|---|---|---|---|
| 1 | 38.53 | 47.70 | 51.53 | - |
| 2 | 27.45 | 36.64 | 43.93 | 76.59 |
| 3 | 29.15 | 33.70 | 41.73 | 43.62 |
| 4 | 23.88 | 33.79 | 41.62 | 42.69 |
| 5 | 31.03 | 41.84 | 44.52 | 44.74 |
| 6 | 14.32 | 17.61 | 34.94 | 40.68 |
| 7 | 44.15 | 52.27 | - | 51.17 |
| 8 | 20.69 | 30.42 | 29.79 | 36.50 |
| 9 | 14.48 | 24.84 | 30.98 | 30.24 |
| 10 | - | - | - | - |
| 11 | 43.89 | 53.41 | - | 53.37 |
| 12 | - | - | - | - |
| 13 | - | 47.85 | 49.21 | 53.27 |
| 14 | 34.89 | 44.20 | 46.85 | 47.52 |
| 15 | 33.33 | 41.94 | 48.43 | 47.97 |
| 16 | 20.31 | 28.94 | 33.79 | 36.26 |
| 17 | 39.61 | 50.12 | 40.22 | - |
| 18 | 31.24 | 40.85 | 45.07 | - |
| 19 | 8.33 | - | - | - |
| 20 | 26.51 | 16.56 | 47.40 | 39.24 |
| 21 | 42.56 | 37.19 | 53.15 | - |
| 22 | - | - | 40.69 | - |
| 23 | - | 49.73 | 46.33 | 54.75 |

ing the dirty air filter. The results of these calculations are presented in Table 1.

The obtained results of calculations for each of the  engines, from the whole measurement series, have been  confronted with the reference parameters of engine no. 1, presented in table  2 [1]. The compressor efficiency adopted the values fitting correctly into the ranges of reference parameters, consistent with the physics of wear and aging processes.

*Table 2.  Chosen reference parameters of  the ABB TPL 77 type  turbocharger engine*

| Charging air press. $p_2$ [$10_5$ Pa] | 1.074 | 2.394 | 2.943 | 3.302 |
|---|---|---|---|---|
| Load [%] | 50 | 85 | 100 | 110 |
| Engine Power $P_e$ [kW] | 3900 | 6630 | 7800 | 8580 |
| Compressor efficiency [%] | 6.02 | 50.80 | 57.10 | 59.24 |

Figure 12 shows  trend patterns of decreasing compressor efficiency of the four engines at selected operating time intervals. They confirmed the nature of compressor efficiency changes and revealed differences in the rate of deterioration, when sailing in different hydrometeorological conditions. Only during short operation intervals, from 4 to 6 days (observation numbers), the tendency for cyclic decrease in efficiency and then its increase, after the compressor canal cleaning, turbine cleaning or air filter replacement, has been noted.
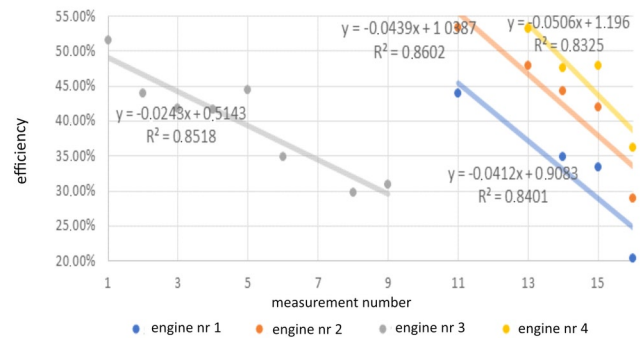


*Fig. 12. Trends in compressor efficiency reduction for four engines in selected operational periods*

Among the set of parameters of  mathematical models of turbocharging systems, the symptoms of process diagnostics of the turbocharging system included such parameters of the main operational process as:
  – rotational speed of the main power engine $n_{ME}$;
  – rotational speed of turbocharger rotor unit  $n_{TS}$;
  – air temperature and pressure before $t_1$, $p_1$ and after the compressor $t_2$, $p_2$;
  – temperature and pressure of exhaust gas before $t_3$, $p_3$ and after the turbine $t_4$, $p_4$; in relation to the fuel rail setting, and selected parameters of  associated processes.

The lack of possibility to measure  exhaust gas pressure before the turbine in turbochargers, which is necessary to determine the expansion of ($p^*_4/p^*_3$) the exhaust gas  in the turbine, makes it impossible to determine the efficiency of the turbine in operation, according to relation (6). With the limited usefulness of the parametric diagnostic test for undertaking maintenance services on turbocharger systems, an attempt has been made to apply machine learning to turbocharger diagnosis.

## 5. Turbocharger diagnostics with the application of machine learning

The authors proposed a classification of the technical condition of a marine engine turbocharger based on the measurement signals of the main engine shaft speed and the charging air pressure at the compressor outlet. These parameters were selected basing on the analysis of available measurement signals of various turbocharged engines and the relationships between engine and turbocharger parameters. Figures 8 and 9 show clear correlations between the main engine shaft speed and the turbocharger rotor speed (Figure 9) and the charging air pressure and the turbocharger rotor speed (Figure 8). Based on these relationships, a strong dependence of charging air pressure on the main engine shaft speed was assumed.

It was found that a neural network with two input signals of the main engine speed $n_{ME}$ and the charging air pressure $p_2$ during machine learning would enable determination of the transformation function $f_{nn}$ of sets $p_2$ and $n_{ME}$ into a set of service necessity index $M$

$$(p_2, n_{ME}) \xrightarrow{f_m} M \qquad (7)$$

The relationship described by equation (7) was realized by the neural network shown in Figure 13. Multilayer neural networks taught by back propagation error are a classic method that is perfectly implemented in many utilities, including the Matlab *nntool* tool.
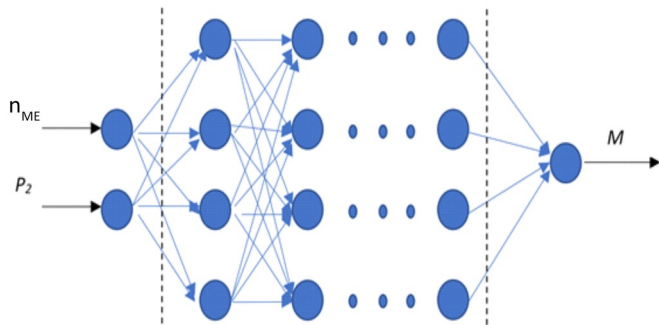


*Fig. 13. A neural network implementing transformation (7)*

## 6. Numeric experimental studies

For the proposed method of identifying the need for turbocharger overhaul, simulation has been carried out using the results of measurements based on the TCA55 turbocharger shown in Chapter 4, Figures 8-10. Figure 14 shows the dependence of the charging air pressure $p_2$ on the main engine shaft speed $n_{ME}$ at different engine loads.
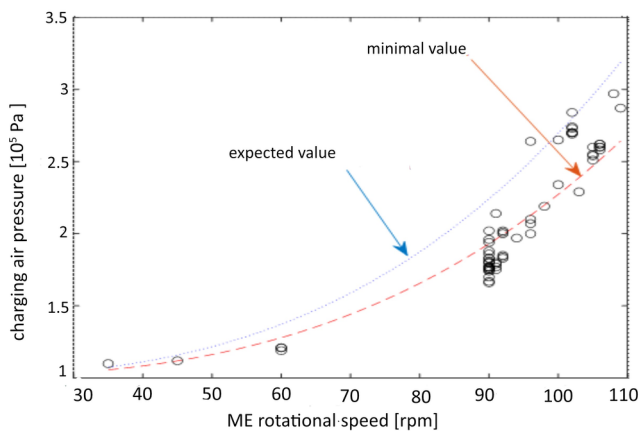


*Fig. 14. A function describing the relationship between the expected values of air charge pressure $p_{2ex}$ and a function describing the minimum acceptable values of air charge pressure $p_{2min}$ in relation to the main engine shaft speed (ME)*

In Figure 14, the blue curve indicates the expected value of air charging pressure $p_{2ex}$ for the maximum turbocharger efficiency, determined by selecting maximum air charging pressure values for chosen main engine shaft speeds. The equation of the expected value curve $p_{2ex} = f(n_{ME})$ was established as a trend line equation for these measurement points. The selected points and the equation of the trend line are shown in Figure 15.
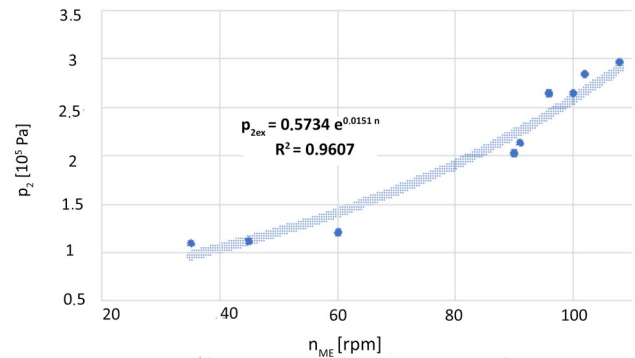


*Fig. 15. Equation of the function of the expected values of charging air pressure*

Figure 14 also shows in red the curve of the $p_{2min}$ limit air charging values. The measurement results below the limiting values indicate the turbocharger is in a condition requiring service. For each measurement point the value of the index $M$ has been assigned; for all points lying on the blue curve and above it, the need for servicing does not occur and the value of the coefficient is $M = 0$. For the points on the red line and below the curve the need for servicing occurs and the value of the index is $M = 1$. For the points situated between the lines the coefficient value was assumed to be linearly dependent on the measured value of the $p_{2m}$ pressure

$$M = \frac{p_{2ex} - p_{2m}}{p_{2ex} - p_{2min}} \qquad (8)$$

A plot of the variation of the $M$-value as a function of boost pressure and engine shaft speed is shown in Figure 16.
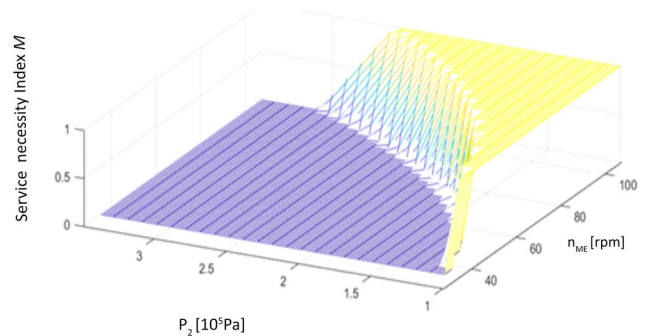


*Fig. 16. Variability of the service necessity index M*

The machine learning experiment was carried out in MATLAB using the *nntool* interface. Neural networks with one input signal can have a quite simple structure, e.g., for mapping the trigonometric function (sine or cosine), a simple 1-7-1 network with one hidden layer containing 7 neurons is enough. In the case of the described modelling of the $M$ index depending on two input parameters, the neural network with one hidden layer did not bring positive results. During subsequent experiments, the network size was determined as 2-80-80-80-80-1 that is 4 layers hidden with 80 neurons in each layer. The
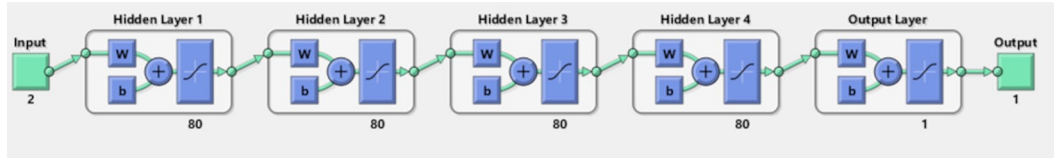
*Fig. 17. Neural network structure*

structure of the network is presented in Figure 17. In the hidden layers for all neurons, the *tansig* transfer function was used, which returns numerical values in the range <-1: 1>, while in the output layer the *logsig* function was used, which has the output value range <0: 1>.

To network learning , a set of cases was generated containing 500 elements in the range of pressure and rotational speed values as shown in Figure 16. The values of the $M$ index were determined according to equation (8). The data is shown in Figure 18. Case values are summarized in the form of a matrix

$$\begin{bmatrix} n_{ME} \\ p_2 \\ M \end{bmatrix} \qquad (9)$$

with a size of 3x500. The first two lines containing the numerical values of the $n_{ME}$ engine speed and the boost pressure $p_2$ were the input values of the neural network (input), while the last line containing the values of the service necessity index $M$ constituted the expected network responses (target). The set of cases was created in the following way: a set of 500 random values of the $n_{ME}$ engine rotational speed in the value range <35: 110> [rpm] was generated. Then the values of $p_{2min}$ and $p_{2ex}$ were determined for these values. Where $p_{2ex}$ was calculated according to the equation presented in Figure 15. For the purposes of the experiment, $p_{2min} = 0.75p_{2ex}$ was assumed. In the next step, 500 values of pressure $p_{2m}$ were drawn from the interval <$p_{2min}$: $p_{2ex}$>. With sets of $n_{ME}$ and $p_2$ values, the service necessity coefficient $M$ was calculated according to equation (8). The numerical values of the $M$ index used as the target set are in the range <0: 1>, therefore the authors concluded that there is no need to normalize the data.

The gradient descent with momentum weight and *bias* learning function was used to adapt the network, i.e., to change the weight value and the bias value. The function calculates the weight change for each neuron based on the input data, neuron error, weight or bias, learning rate coefficients and momentum.

The default parameters of the *trainlm* and *learngdm* functions were used in the network learning:

$lp\_lr$ = 0.01 learning rate,

$lp\_m$ = 0.9 momentum,

$epochs$ =1000 - maximum number of epochs,

$goal$ = 0 expected mapping error,

$max\_fail$ = 6 validation error for a specified number of consecutive epochs,

$max\_fail\_epochs$ = 50 number of epochs with invariant validation error

$min\_grad$ =1e-7 minimal performance gradient,

$mu$ = 0,001 momentum change factor (initial value),

$mu\_dec$ = 0.1 reduction factor of $mu$,

mu_inc = 10 growth factor of $mu$,

$mu\_max$ = 1e10 maximum value of $mu$,

$time$ = inf maximum time of network training [s].

Figure 19 shows the process of learning the network with a division into the quality of mapping the training, validation, and test sets. Network training was completed after 211 iterations after the network mapping error condition was satisfied by subsequent generations (*max_fail*). The best representation of the validation set, obtained in 161 iteration, was marked on the run. Over the next 50 epochs, the mapping error did not improve, which was the basis for completing network learning.



*Fig. 18. Data for neural network learning*



*Fig. 19. The process of training a neural network*
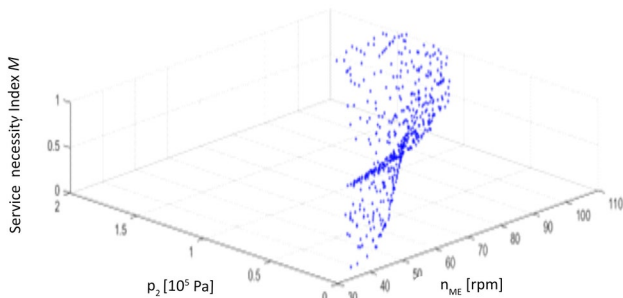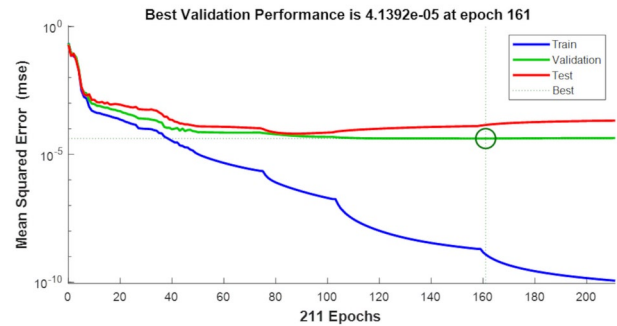
The *trainlm* function was selected to train the network, which implements non-linear Levenberg-Marquardt optimization. The *trainlm* function is often recommended as a supervised first choice algorithm. This function randomly divides the provided set of cases into the following subsets*: training set*, which is 1/2 of the case set, *validation set*, which is 1/4 of the case set, and *test set*, which is 1/4 a collection of cases. The validation set is used to evaluate the learning quality of the network and to decide whether to stop training early if the quality of the network's response to the validation vector (*max_fail*) does not improve or remains the same for a specified number of consecutive epochs. The test set is used to monitor whether the network is generalizing correctly, but its results have no effect on the training progress.

As a result of the network learning process, the correct mapping of the sets of boost pressure $p_2$ and the rotational speed of the $n_{ME}$ main engine into the service necessity index set was obtained. Figure 20 shows the results of linear regression R mapping individual sets by the neural network and their generalization.

After learning the neural network based on the generated set of cases, the correctness of the neural network response was verified using the real measurement data presented in Figure 14. The results of the measurements were returned by the neural network to the correct values of the maintenance necessity factor $M$. The obtained values of the index $M$ were plotted on the course of the index variability (Fig. 16). The results are presented graphically in Figure 21.
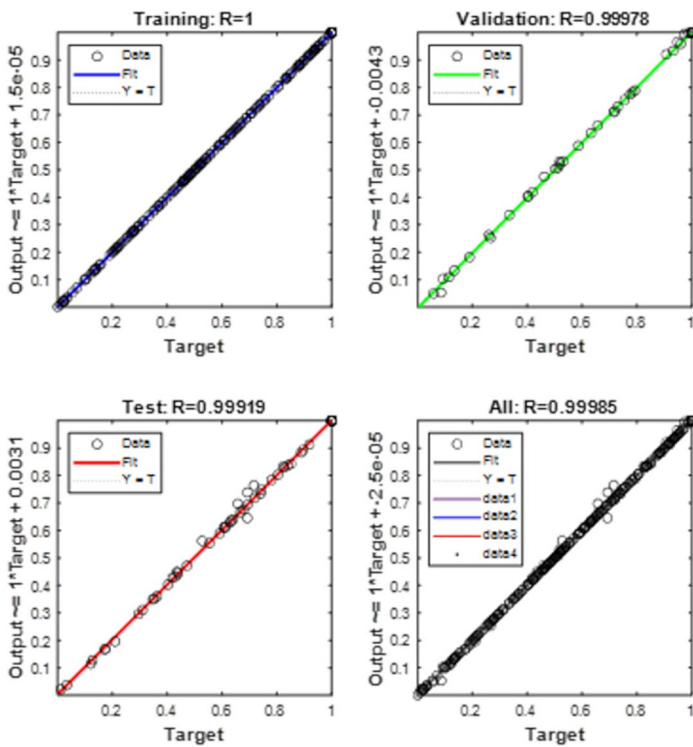
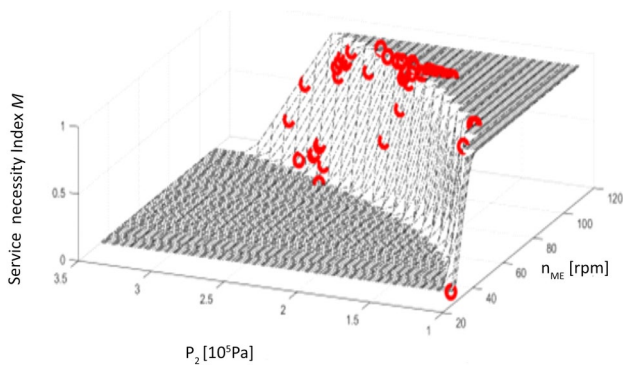*Fig. 20. The result from the neural network learning process in the form of linear regression*



*Fig. 21. M index values obtained from neural network for measurement points*

## 7. Summary

Methods of control of operational parameters and supervision of turbocharging system functioning in marine engines, which are currently used, do not provide full information for making maintenance/service decisions, in accordance with the adopted operational strategy, classification society regulations and engine manufacturer's regulations. Current control and archiving systems of operational parameters of turbocharging systems may not always be expert systems with diagnostic reasoning. There are strong diagnostic relations between some of the operational parameters, resulting both from models of energy processes and from the obtained parametric diagnostic test results. They can be the basis for searching for methods of inference other than parametric ones The experimentally found strong diagnostic relationships can be used to search for decisions based on other sources of information, such as machine learning. A turbocharger service necessity index $M$ was proposed. Its values depended on the marine engine shaft speed $n_{ME}$ and compressor outlet air pressure $p_2$. The process of teaching the artificial neural network the dependence of the indicator $M = f(n_{ME}, p_2)$ was carried out. The neural network responses were verified using the data from measurements on a real marine engine and promising results were obtained. The simulation experiment conducted in this research demonstrated the usefulness of the proposed method in utilitarian diagnosis of turbochargers.

The authors conducted the first attempt to use artificial neural networks to classify a turbocharger for service work. Multilayer neural networks taught by back propagation error are a classic method. In the experiment, the authors focused on obtaining satisfactory results in the mapping of the proposed service necessity coefficient $M$ based on the measured values of the engine speed $n_{ME}$ and the boost pressure $p_2$. This is the first step of the authors in the use of neural networks for the service classification of a turbocharger. The next planned step is to extend the classification of the need for service to include the mechanical element of the turbocharger, i.e., turbine blades or compressor flow channels. These elements are shown in figures 1 and 2. After determining the measured parameters, based on which it is possible to indicate the place of maintenance of the turbocharger. The study of the problem during the research showed that there are other strong diagnostic relationships to be applied in decision support using artificial neural networks. One of them is the relationship of charge air pressure with turbocharger rotor unit speed and charge air pressure with fuel rail indicator [2]. This provides a basis for further experiments supporting turbocharger service decisions based on the values of these parameters. The authors plan experiments using methods other than error backpropagation.

## References

1. ABB Turbo Systems Ltd: Operational Manual TPL 77-A, 2007
2. Adamkiewicz A., Zeńczak W.: Diagnostyka systemu turbodoładowania okrętowego silnika o zapłonie samoczynnym podczas eksploatacji. DIAGO 2016 Technicka Diagnostika Asociace Technickych Diagnostiku Ćeske Republiku, o.s., ISSN 1210-311X, Z1, ROĆNIK XXV, Abstrakt book, CD 5-14
3. Anantharaman M., Khan F., Garaniya V., Lewarn B., Reliability Assessment of Main Egine Subsystems Considering Turbocharger Failure as a Case Study. TransNav the International Journal on Marine Navigation and Safety of Sea Transportation 2018;12(2):271-276. Doi: 10.12716/1001.12.02.06
4. Anantharaman M, Islam R, Sardar A, Garaniya V, Khan F. Impact of defective turbocharging system on the safety and reliability of large marine diesel engine. TransNav : International Journal on Marine Navigation and Safety of Sea Transportation. 2021;15(1):189-198.
5. Burzyński K.: Analiza parametrów dopuszczalnych i granicznych stosowanych w diagnostyce cieplnych maszyn wirnikowych. Praca dyplomowa inżynierska, Akademia Morska w Szczecinie, Wydział Mechaniczny, Szczecin 2015.
6. Cui X, Yang C, Serrano JR, Shi M. A performance degradation evaluation method for a turbocharger in a diesel engine. R Soc Open Sci. 2018 Nov 14;5(11):181093. doi: 10.1098/rsos.181093. PMID: 30564402; PMCID: PMC6281929.
7. Deng, X. W., Gu, Y. J., Fang, L. P., Ren, Z. X., & Han, Y. P., Study of Intelligent Fault Diagnosis Method for Turbo-Generator Unit Based on Support Vector Machine and Knowledge. Applied Mechanics and Materials 2014; 543–547; 1057–1063. Trans Tech Publications, Ltd, https://doi.org/10.4028/www.scientific.net/amm.543-547.1057
8. Divya M.N, Modeling a Fault Detection Predictor in Compressor using Machine Learning Approach based on Acoustic Sensor Data,

(IJACSA) International Journal of Advanced Computer Science and Applications 2021;12, (9); 650-667

9.  Dong H., Zhao Z., Fu J., Liu J., Li J., Liang K., Zhou Q., Experiment and simulation investigation on energy management of a gasoline vehicle and hybrid turbocharger optimization based on equivalent consumption minimization strategy, Energy Conversion and Management, 2020; 226. Doi.org/10.1016/j.enconman.2020.113518.

10. Főző l., Andoga R., Madarász L., Kolesár J., Judičák J., Description of an intelligent small turbocompressor engine with variable exhaust nozzle, SAMI 2015, IEEE 13th International Symposium on Applied Machine Intelligence and Informatics, January 22-24, 2015, Herlany, Slovakia

11. Guan C, Theotokatos G, Chen H. Analysis of Two Stroke Marine Diesel Engine Operation Including Turbocharger Cut-Out by Using a Zero-Dimensional Model. Energies. 2015; 8(6):5738-5764. https://doi.org/10.3390/en8065738

12. Hipple S. M., Bonilla-Alvarado H., Pezzini P., Shadle L., Bryden K. M., Using Machine Learning Tools to Predict Compressor Stall, Journal of Energy Resources Technology 2020; 142; 072305-1- 072305-9

13. Hountalas D.T.; Sakellaridis N.F.; Pariotis E.; Antonopoulos A.K.; Zissimatos L.; Papadakis N., Effect of turbocharger cut out on two-stroke marine diesel engine performance and NOx emissions at part load operation. In Proceedings of the ASME 12th biennial conference on engineering systems design and analysis, Copenhagen, Denmark, 25–27 July 2014. ESDA2014-20514

14. Hriadel D. Health Status Monitoring of Turbocharger for Passenger Vehicle Applications, Master's Thesis, Czech Technical University in Prague, Department of Control Engineering, Prague, January 2019

15. Knežević V., Orović J., Stazić L., Čulin J., Fault Tree Analysis and Failure Diagnosis of Marine Diesel Engine Turbocharger System, Journal of Marine Science and Engineering 2020; 8, (12), doi:10.3390/jmse8121004

16. Lau C., Maier M., Knowledge-based predictive maintenance for olefins turbocompressors, 2004 AIChE Spring National Meeting, Conference Proceedings; 611 – 625

17. Liu C, Cao Y, Ding S, Zhang W, Cai Y, Lin A. Effects of blade surface roughness on compressor performance and tonal noise emission in a marine diesel engine turbocharger. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering. 2020;234(14):3476-3490. Doi:10.1177/0954407020927637

18. Liu C., Cao Y., Liu Y.2, Zhang W., Ming P., Numerical investigation of marine diesel engine turbocharger compressor tonal noise. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering. 2020;234 (1):71-84. Doi: 10.1177/0954407019841808

19. MAN B&W Turbocharger. Cleaning the Turbine – Benefits of Dry Cleaning at operating load. Diesel Customer Information. MAN B&W Diesel AG 86224 Augsburg Cus 228, 06/06

20. Medica, V., Račić, N., & Radica, G. (2009). Performance simulation of marine slow-speed diesel propulsion engine with turbocharger under aggravated conditions. Strojarstvo, 51(3), 199-212

21. Nguyen-Schafer, H. Rotodynamics of Automotive Turbochargers. Springer. 2015,XV, 362 p. 222. Chapter 2 Thermodynamics of Turbochargers. ISBN: 978-3-3-319-17643-7.

22. Nnaji, O.E., Nkoi, B., Lilly, M.T., Le-ol, A.K.: Evaluating the Reliability of a Marine Diesel Engine Using the Weibull Distribution. Journal of Newviews in Engineering and Technology (JNET). 2, 2, 1–9 (2020).

23. Pawełczyk M, Fulara S, Sepe M, De Luca A, Badora M. Industrial gas turbine operating parameters monitoring and data-driven prediction. Eksploatacja i Niezawodność – Maintenance and Reliability 2020; 22 (3): 391–399, http://dx.doi.org/10.17531/ein.2020.3.2

24. Prytz, Rune et al. (2015). "Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data." In: Engineering applications of artificial intelligence 2015;41: 139–150. Doi: 10.1016/j.engappai.2015.02.009

25. Tabaszewski M, Szymański G. M. Engine valve clearance diagnostics based on vibration signals and machine learning methods. Eksploatacja i Niezawodność – Maintenance and Reliability 2020; 22 (2): 331–339, http://dx.doi.org/10.17531/ein.2020.2.16.

26. Taylor J.V, Conduit B., Dickens A., Hall C., Hillel M., Miller R. J., Predicting the Operability of Damaged Compressors Using Machine Learning, Journal of Turbomachinery 2020; 142; 051010-1: 051010-8, DOI: 10.1115/1.4046458,

27. Turbocharging Efficiencies - Definitions and Guidelines for Measurement and Calculation. International Council on Combustion Engines. CIMAC Working Group "Turbocharger Efficiency" and approved by CIMAC in May 2007, Frankfurt, Germany, Number 27/2007

28. Vanhaelst, R., Kheir, A., Czajka, J., A systematic analysis of the friction losses on bearings of modern turbocharger. Combustion Engines,2015; 55(1), 22-31

29. Varbanets R., Fomin, O., Píštěk V., Klymenko V., Minchev D., Khrulev A., Kučera P., Acoustic method for estimation of marine low-speed engine turbocharger parameters. Journal of Marine Science and Engineering, 2021;9(3), 321.

30. Vrettakos NA. Analysis and characterization of a marine turbocharger's unstable performance. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment. 2018;232(3):293-306. doi:10.1177/1475090217693118

31. Witkowski K.: Research on Influence of Condition Elements the Supercharger System On The Parameters Of The Marine Diesel Engine Journal of KONES Powertrain and Transport, Vol. 20, No. 1 2013

32. Yi W., Hailong L., Gengxuan C., Jiawei Y., Fault Diagnosis of Marine Turbocharger System Based on an Unsupervised Algorithm, Journal of Electrical Engineering & Technology DA - 2020; 05(01), https://doi.org/10.1007/s42835-020-00375-z

33. Zhang, J., Sun, H., Hu, L., and He, H., Fault diagnosis and failure prediction by thrust load analysis for a turbocharger thrust bearing. In Turbo Expo: Power for Land, Sea, and Air. 2010;44014: 491-498
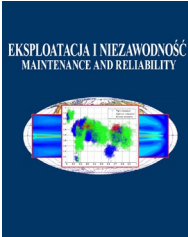
# Identification of crashworthiness indicators of column energy absorbers with triggers in the form of cylindrical embossing on the lateral edges using artificial neural networks

Indexed by:
Web of Science Group

## Mirosław Ferdynus [a,*], Jakub Gajewski [a]

**Lublin University of Technology, Department of Machine Construction & Mechatronics, Lublin, Poland**

## Highlights

- The possibility of using artificial neural networks to identify the most advantageous variants of column energy is shown

- The considered design variants differ in geometric parameters and the position of the trigger

- The research was carried out with the use of FEM, and the models were validated by the experiment

- It has been shown that the use of neural networks to predict the properties of the energy absorber is possible with a slight error in relation to the time-consuming multi-variant FEM analyzes.

## Abstract

The paper presents the possibility of neural network application in order to identify the most advantageous design variants of column energy absorbers in terms of the achieved energy absorption indicators. Design variants of the column energy absorber made of standard thin-walled square aluminium profile with triggers in the form of four identical cylindrical embossments on the lateral edges were considered. These variants differ in the diameter of the trigger, its depth and position. The geometrical parameters of the trigger are crucial for the energy absorption performance of the energy absorber. The following indicators are studied: PCF (Peak Crushing Force), MCF (Mean Crushing Force), CLE (Crash Load Efficiency), SE (Stroke Efficiency) and TE (Total Efficiency). On the basis of numerical studies validated by experimentation, a neural network has been created with the aim of predicting the above-mentioned indices with an acceptable error for an energy absorber with the trigger of specified geometrical parameters and position. The paper demonstrates that the use of an effective multilayer perceptron can successfully speed up the design process, saving time on multivariate time-consuming analyses.

## Keywords

crashworthiness indicators, energy absorber, thin- walled tube, artificial neural network.

## List of abbreviations used in the article

ANN- Artificial Neural Network,
CFE- Crash Force Efficiency,
CFRP- Carbon Fiber Reinforced Polymer,
CLE- Crash Load Efficiency,
EA- Energy Absorbed,
FE- Finite Element,
FE- Finite Element Method,
FMTS- Foam-filled Multi-cell Thin-walled Structures,
GA- Genetic Algorithm,
MCF- Mean Crushing Force,
MOD- Multi-objective Optimization Design,
PCF- Peak Crushing Force,
RP- Reference Point,
SEA- Specific Energy Absorbed,
SE- Stroke Efficiency,
TE- Total Efficiency,

TOPSIS- Technique for Ordering Preferences by Similarity to ideal Solution,
TW- Thin-Walled,

## 1. Introduction, motivation and objectives of the research

In the times of increasingly intensive traffic on the roads and higher speeds reached by vehicles, the passive safety of motor vehicles is becoming a very important issue often determining the survival of the traveling people. One of the key aspects of passive safety is a properly designed crumple zone, of which crash boxes are a very important element. Crash boxes are mounted on the side members inside the engine compartment, connected by a rigid beam, which is hidden under the bumper and is designed to absorb energy in low-speed crashes (15-20 km/h) protecting the above-mentioned side members from damage, which is costly and time-consuming to repair. This makes a big difference to vehicle operation in terms of repair costs and time out of service. The presence of crash boxes at higher speeds is also

(*) Corresponding author.
E-mail addresses: M. Ferdynus (ORCID: 0000-0003-0348-631X): m.ferdynus@pollub.pl, J. Gajewski (ORCID: 0000-0001-8166-7162): j.gajewski@pollub.pl

important, as they are able to absorb quite a large portion of energy for their size, which consequently leads to much less destruction of this part of the vehicle and in the case of extremely high energies increases the chances of survival of the driver and passengers.

In the modern design process, the artificial neural network technique is often used among other things because of its high predictive capability, which makes it possible to arrive at optimal solutions at a lower cost. Simplifying, a typical ANN acts like a black box, transforming input data into output data by applying various learning procedures. An ANN mimics biological neurons in that it has many non-linear computational agents working in parallel. All these computational agents are tightly coupled by weights, which are carefully modified by learning algorithms to improve their prediction performance or to search for optimal, maximum or minimum values. In this paper, ANN is used to predict the values of energy-absorption indicators in relation to the geometrical parameters of the trigger and its position.

The main objective of this study is to present the possibility of using a neural network to identify the most favourable design variants of column energy absorbers in terms of the energy absorption rates achieved. Design variants of the energy absorber in the form of a thin-walled square column made of aluminium alloy with triggers in the form of four identical cylindrical embossments on the lateral edges were considered. The variants differ in the diameter of the trigger, its depth and position. The geometrical parameters of the trigger are crucial to the energy absorption performance of the absorber. The main task of the work was therefore, based on numerical studies verified experimentally, to create a neural network that would predict energy-absorption rates with acceptable error for an energy absorber with a trigger with specific geometrical parameters and position.

## 2. State of art: an overview

Designers of crumple zones intended to absorb impact energy must satisfy two major, often conflicting requirements: minimizing the initial load while maximizing the amount of energy absorbed. The initial load at the moment of impact must not be too high to avoid excessive decelerations during vehicle impact (biomechanical reasons). On the other hand, the main requirement is the maximum possible energy absorption and dissipation capacity to make these zones effective. Therefore, it is important, the search for an optimal design of these elements, maximize the energy absorption and minimize the ratio of the initiating peak crushing force (PCF) to the mean crushing force (MCF).

Thin-walled metal tubes, are broadly used as energy absorbers because they are relatively cheap and efficient in absorbing energy. Square, rectangular or circular tubes are commonly used but other cross-section shapes have also been analyzed for their suitability [7, 31], as well as multi-cornered [1, 35] multi-cell [12, 28, 42] and bi-tubal designs [38, 45]. The behavior of these types of structures during axial, oblique and lateral crushing has been extensively studied over the past decades. This research has been reflected in numerous publications, among which review articles have been published more and more over the years [6, 9, 40, 43]. Abramowicz [2] presented deep considerations based on analytical description and experimental studies of thin-walled structures subjected to axial crushing force, although the first paper appeared much earlier in the 1960s [5]. However, mainly due to the underdevelopment of numerical methods, it took another decades before the field could develop. Many relevant papers were presented in the 1st International Symposium on Structural Crashworthiness held in Liverpool in 1983, which were collected in a special issue of the International Journal of Mechanical Science (9/10, 1983). At that time, valuable books also appeared [49–51].

In his work [6] Alghamdi was one of the first to systematize the works that appeared until the end of the 20th century, dividing energy absorbers according to the shape of the cross-section but also the way of operation. Nurick et al [43], on the other hand, limited their review to axially impacted tubular structures to which imperfections in the form of recesses, notches or combinations thereof were introduced, but also presented interesting results for "prebuckle" structures. Hollow or foam-filled structures were considered. A separate section describes the behavior of structures that are ripped into strips during axial impact. On the other hand, publication [9] based on an extensive literature review, presents a comprehensive review of recent developments in the crashworthiness of TW tubes used in vehicles, with a focus on topics that have emerged in the last fifteen years, such as crash optimization design and energy absorption responses of unconventional thin-walled components, including multi-cell tubes, functionally graded thickness tubes. Both hollow and foam-filled structures operating under axial, lateral, oblique, and bending loads were considered, including foam-filled structures with graded functionality. The paper [40] mainly discusses the current state of knowledge on energy absorption of gradient structures and materials and the effect of gradient properties on their crashworthiness. These advanced energy-absorbing structures and materials primarily include thin-walled structures with variable diameter/width/wall thickness/ strength, variable density cellular materials and their filler structures, and other hybrid structures with multiple graded properties. It is undoubtedly a very developing direction, pursued both by academia and industry. An example of a publication where crashworthiness issues are presented in various aspects of aero-structure design and testing is work of Xianfeng Yang et al. [41] presenting a systematic review of the literature and specific helicopter design solutions. Also noteworthy are the very valuable, even fundamental books [19, 27] on which crashworthiness researchers are based, as evidenced by the fact how often they are quoted.

In the work of Langseth at al [24], extrusions with square cross-section, made of aluminum alloy, were subjected to static and dynamic analysis. Both experimental studies and numerical simulations (LS-DYNA code) were carried out. The specimens were prepared with a small trigger in the form of embossments on opposite walls. Tests were conducted for two variants of its position: in the middle plane and on top. The behavior of the structure was tested at different tup weights and different impact velocities. Both the initiating peak crushing force PCF and the mean crushing force MCF were calculated.

In publication [4], Alavi Nia et al conducts a study of aluminum thin-walled structures with triangular, square, hexagonal and octagonal cross-section made in two multi-cell and one hollow variants. It has been shown that regardless of the configuration of the location of the inner walls of multi-cells, the SEA (Specific Energy Absorbed) is always greater than for the profile in a simple form, while the variant in which the inner walls of multi-cells connect with the outer profile in the middle of the walls and not at the corners is much more advantageous. Another very important conclusion that disqualifies multi-cell profiles from certain applications is that there is a 30 to 120% increase in PCF force compared to a plain profile, and this result is obtained for quasi-static tests and can be even higher for dynamic tests. Such an increase will result in a drastic increase in overload during the initial crush phase. Unfortunately, CLE values are not under investigation and it is uncertain whether the CLE value (Crash Load Efficiency) ultimately decreased or increased despite the increase in PCF and MCF.

The paper [25] analyzed the effect of the placement of triggers in the form of grooves on square section profiles on the PCF (called maximum compressive repulsive force by the authors) and the amount of energy absorbed. The optimal groove layout was obtained by introducing indentations with a pitch corresponding to the fold wavelength accurately estimated from the computer simulation. An important conclusion is that excessive concentration of triggers although leads to a reduction in PCF the member undergoes destruction which is a combination of global buckling and progressive folding and as a result does not absorb energy as well. Incidentally, this is one of the first papers, and unfortunately few, in which the triggering mechanism was properly appreciated. In many papers the role of triggers is down-

played as if their role was limited only to the initiation of the crushing process, while their influence on the achieved parameters, especially CLE, is undeniable.

In a paper published by Rai et al [34], two decades later, the effect of the triggering mechanism on the crushing force efficiency of aluminum tubular absorbers was investigated. Different triggering mechanisms such as cut-out, circumferential notch and end- fillet were investigated using a validated numerical model. Based on the numerical and experimental results, it was found that tubes made of aluminum showed better crashworthiness compared to steel tubes. According to these results, the developed trigger mechanisms significantly change the crash performance of tubular absorbers. The crushing force efficiency doubled with the most effective trigger mechanism, while the stroke and specific energy absorption efficiencies decreased by 4% and 15%, respectively.

Very interesting results concerning the triggering mechanism and circumferential stresses in thin-walled tubes of square and circular cross-sections, whose geometry was perturbed by the use of corrugations are presented in [11]. Undoubtedly, this work should be continued as far as the determination of energy absorption coefficients is concerned. On the other hand, in the work of Zhang et al. [46] numerical investigations were presented concerning the course of the crushing process and the improvement of energy absorption achieved by applying pyramidal embossments arranged in a certain pattern on the whole lateral surface of the specimen in two configurations but also with different number of pyramids placed on the specimen depending on the size of a single base element. Also this valuable work demands, after overcoming technological problems, experimental verification.

A very valuable publication is the work of Karagiozova and Jones [20] where the dynamic elastic-plastic buckling of thin-walled square tubes was studied from the point of view of the propagation of the elastic-plastic stress wave that originated from axial impact load. The influence of the impact velocity and the striking mass on the development of the buckling shape was discussed when considering the transient deformation process. The wave behavior in square tubes was also compared with geometrically equivalent circular tubes.

Axial crushing of square crash boxes was also the subject of a very interesting publication by Jafarzadeh-Aghdam and Schröder [18], where the authors addressed the problems of experimental validation and the large number of irreproducible test results. There is an interesting description of the influence of stress wave propagation, existing imperfections as a source of this irreproducibility.

At the turn of the centuries, a new class of tubular energy absorbers emerged, which are classic axially loaded thin-walled metal columns filled with foam or honeycomb structures. Energy absorbers of this type absorb energy during the crushing process both through phenomena in the metal profile itself and in the filling. Due to the existence of many parameters affecting the energy absorption capacity of such columns and the interaction between the outer shell and the filling, optimization procedures had to be developed. One of the first attempts in this field was published by Zarei and Kroger [44], who performed an optimization of a tube filled with aluminum foam and also analyzed hollow tubes.

In the first decade of the 21st century, multi-cell columns, both empty and filled with foam, were implemented as energy absorbers [12]. Yin et al [42] presented the results obtained using the non-linear LS-DYNA finite element code, for six types of foam-filled multi-cell thin-walled structures (FMTS) with different cell numbers. During the process of multi-objective optimization design (MOD), four kinds of commonly used metamodels was established to reduce the computational cost of crash simulations by the finite element method. A very interesting concept of energy absorbing structure was presented in Luo's work [28], where the energy absorption properties of square tubes have been significantly increased by replacing ultra-thin solid walls into sandwich walls. On the other hand, W. Liu et al. in paper [26], performed multi-objective crash optimization of star sandwich tubes. A rather promising approach to improve the energy efficiency

and impact performance of thin-walled structures was proposed by Zhang et al in their paper [48], by introducing a thickness gradient in the cross-section. Experimental studies were first carried out for square tubes with two types of thickness distribution, and only then FE numerical analyses were carried out to confirm the experimental results. On the other hand, in the paper [33], the authors addressed the issue of evaluating the crashworthiness of newly designed multicellular structures with different structural forms (i.e. square, hexagonal, octagonal, decagonal and circular). The optimisation problem consisted of minimising the PCF and maximising the SEA and the variables were geometrical parameters related to the cross-sectional area and, more specifically, to the ratio of the inner and outer tube dimensions. Geometrical dimensions of these structures were optimized using ANNs and GA (genetic algorithm) by considering three different scenarios. The optimal structures were compared together from the crashworthiness point of view by considering two conflicting crashworthiness indicators namely SEA and PCF using a decision making method called TOPSIS (technique for ordering preferences by similarity to ideal solution).

The paper [16] presents the results of numerical tests of impact and energy absorption capacity of thin-walled aluminum columns, having a square cross-section and spherical indentations on their lateral surfaces. The numerical models were validated using an experiment conducted on the drop hammer. The crushing behavior of the columns and some crashworthiness indicators were investigated. The most beneficial design/construction alternatives in terms of achieved crash performance have been indicated. Numerical analysis of the thin-walled structure with different trigger locations was also conducted in article [36].

The article by Zhang & Huh [47] presents a numerical study of the behaviour of longitudinally grooved square tubes made of steel during dynamic axial compression. The simulations took into account phenomena such as the distribution of effective plastic strain and thickness changes resulting from the stamping process. Apart from force-displacement characteristics, only PCF and SEA were determined.

The influence of triggers in the form of cylindrical symmetrical indentations located at the edges of columns on the performance of the structure and energy-absorption indicators was analysed in a paper [15], where advanced numerical studies with experimental validation were demonstrated. The present work is a continuation of those considerations.

Optimisation of structural crashworthiness and energy absorption has become an important research topic due to its proven benefits for public safety. Fang et al provide a comprehensive review of the important studies on design optimization for structural crashworthiness and energy absorption [13]. The design criteria used in crashworthiness and energy absorption studies are reviewed and surrogate modelling for evaluating these criteria is discussed. Multi-objective optimization, optimization under uncertainties and topological optimization from concepts, algorithms to applications in crashworthiness studies are also reviewed.

The analysis of bibliography indicates a recent publication on the application of artificial intelligence methods with energy-absorption analysis. Baykasoğlu et al [10] described designs of novel lattice structure filled square thin-walled tubes by using a compromise programming based multi-objective crashworthiness optimization procedure. An artificial neural network is employed for predicting values of the objective functions. Laban et al [23] provides insights into the sensitivity of braided carbon-kevlar round tubes to external damages and neural network-based models that can predict the consequences of damages on the crush behaviour. Paygozar and Dizaji analyzed a damper of high-rise structures using artificial neural network [32]. Literature studies have also shown interesting publications on the use of neural networks and FEM for crashworthiness studies. In a paper [29], Marzbanrad and Elbrahimi conducted a numerical crushing study of thin-walled circular aluminium tubes to investigate their behaviour under axial impact loading for five crushing parameters using multi-

objective optimisation with a weighted sum method. To improve the accuracy of the optimisation process, artificial neural networks were used. A genetic algorithm was also implemented. Finite element code, capable of evaluating the crushing parameters, from which the output is used to train and test the developed neural networks. In Mirzaei et al article [30] the multi-objective optimization of cylindrical aluminium tubes under axial impact load was presented. A back-propagation neural network was constructed as the model to formulate the mapping between variables and objectives. The finite element software ABAQUS/Explicit was used to generate the training database for the network, as in our research. Validating the results of finite element model, several impact tests are carried out using drop hammer. In [37] the finite element model of the CFRP tube was developed using the Tsai-Wu failure criterion to model the crush characteristics. A series of FEM simulations were conducted considering different fiber directions and the number of layers to generate enough data for constructing the artificial neural network.

The literature analysis has shown that more and more research interest is directed towards coupling numerical simulations validated by experiment and analysis using neural networks. There are few works that would give answers to questions concerning the optimal configuration of geometrical parameters of triggers and their location from the point of view of crashworthiness. It was the authors' ambition to at least partially fill this gap. Due to the fact that the triggering mechanism is most effective in simple structures, the consideration was limited to square section structures, which despite the emergence of research on multicell, multicorner and bitubal structures are still widely used due to their simplicity and cheapness. Not without significance is the fact that geometrically more complex structures, despite showing a much higher SEA value compared to simple ones, generate a much higher PCF value which affects the overloads to which the protected object is subjected. This eliminates this type of structures from certain applications.

In this paper [21], the crashworthiness of composite rectangular tubes was analysed using experimental and ANN techniques. Based on the experimentally obtained values of different impact strength parameters under different loading conditions, ANN models were constructed to determine the optimum cross-sectional ratio of the laminated composite to achieve the target mechanical properties, such as load carrying capacity and energy absorption. There are many publications in which artificial intelligence has been used to predict specific parameters. In paper [8] the authors illustrated this using fatigue crack length as an example, while paper [3] applied the network to aircraft engine failure prediction. An interesting look at energy efficiency issues on another technical facility is shown in publication [22].

The literature review shows that the issue of predicting the values of energy-absorbing indices using ANNs is novel and can significantly speed up the process of developing an energy-absorber with the desired parameters. The considerations carried out in this work are of universal character, however, they were conducted on the basis of one type of trigger - a cylindrical symmetrical embossment on the side edge. The influence of geometrical parameters of triggers (diameter of cylindrical embossment and its depth) and their location on the achieved energy absorption indicators was studied. The neural networks created on the basis of validated numerical tests allowed predicting what results in terms of energy absorption rates will be achieved by an energy absorber with a trigger of any set of geometrical parameters and its position. This makes it possible to obtain the result much faster and at a lower cost. This opens the way to the issue of synthesis of energy-absorber parameters with the desired energy-absorbing indicators. The conducted studies using neural networks also gave an answer to the question of the level of significance of a given parameter on the obtained indicators, as the same effect can be obtained by appropriately changing different geometrical parameters and position.

## 2. Crashworthiness indicators

Among the various indices for assessing structural crashworthiness, specific energy absorption (SEA) is commonly used to quantify the energy absorption capacity of different types of energy absorbers or different structural materials. It is the quotient of the energy absorbed and the mass of the energy absorber. The value of the absorbed energy EA is commonly obtained by determining the area under the crushing force-displacement (shortening) curve over a segment equal to the crushing distance. Since in the considerations carried out in this work a constant drop energy is assumed and the mass of all energy absorber models is identical - the SEA index loses its comparative aspect and is practically invariable for all columns considered and amounts to SEA= 14,617 kJ/kg. Besides, the EA and SEA energy values do not show by themselves whether the energy absorption process was effective, so the following indicators will be used in the considerations: PCF, MCF, CLE, SE and TE (abbreviations will be explained below). The initial peak crushing force PCF, occurring in the first milliseconds after impact, is an extremely important factor for biomechanical reasons and, therefore, a maximum reduction is sought at the design stage. Reducing this force is the simplest way to maximise the crash load efficiency index CLE, which is given by the formula:

$$CLE = \frac{MCF}{PCF} \cdot 100\% \qquad (1)$$

Alternatively, it is defined as crash force efficiency CFE:

$$CFE = \frac{MCF}{PCF} \qquad (1a)$$

Another crashworthiness indicator is stroke efficiency SE, which represents the deformation capacity of an absorber and it is defined as follows:

$$SE = \frac{U}{L_o} \qquad (2)$$

where, U represents its maximum shortening (crushing distance) and $L_o$ its initial length.

The most desirable value of the SE factor is the highest, corresponding to the highest value of the crushing distance. The optimum structure during the crushing process should be able to use its entire available length to absorb the impact energy, obviously if the impact energy is large enough to cause such a large deformation. This is why SE is one of the basic indicators of crushing performance. A combination of the CLE (CFE) and the stroke efficiency SE was proposed by Hanssen et al. [17] as total efficiency TE to assess the whole performance of an energy absorber. This can be a percentage or a dimensionless value depending on whether the factor is CFE or CLE, as shown below:

$$TE = CLE \times SE \qquad [\%] \qquad (3)$$

$$TE = CFE \times SE \qquad [-] \qquad (3a)$$

## 4. Subject of the investigation

The subjects of the study were thin-walled aluminum tubes of square section with four dents formed by cylindrical embossing at the corners on each tube. All tube models are characterized by the cross-sectional dimensions □40x1.2 and the constant height l=200 mm. The dent's geometry was described by the diameter **D** and the depth **g** which

can easily be related to the overall dimension of the column section, while the position of the trigger is defined by the dimension **h**, which is the distance of its centre plane from the bottom edge as shown in Figure 1.



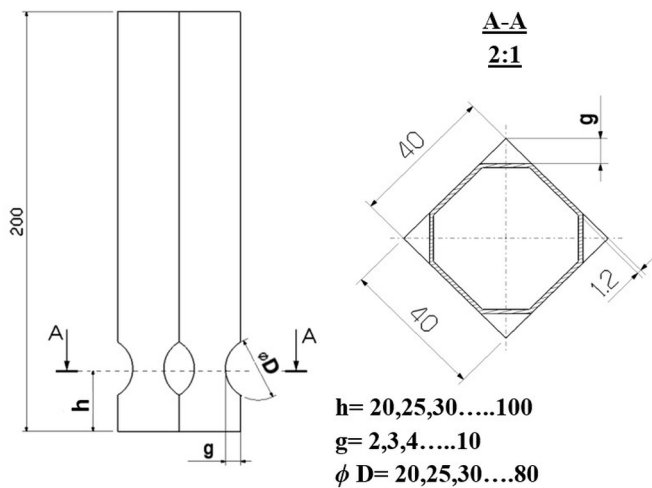h= 20,25,30…..100
g= 2,3,4…..10
$\phi$ D= 20,25,30….80

*Fig. 1. Construction drawing of the column model with cross-section in the trigger location plane*

The research program included the realization of a series of models differing in geometrical parameters of dent and its position. In the first round of research, a total of 103 models were numerically analyzed in order to obtain not only results, but above all data to create a neural network and predict the behavior and energy absorption capacity for any data configuration (**D, g, h**). The range of diameters of dent **D** was from 20 to 80 mm (50-200% of column width) and its depth **g** from 2 to 10 mm (5-25% of column width). The trigger position **h** ranged from 20 mm to 100 mm (mid surface) measured from the bottom edge. The following model designation has been adopted: the first letter denotes the type of trigger D (dent), then the subsequent dash gives the diameter of the trigger **D**, its depth **g** and its position **h** preceded by the letter **h**. For example, a model with a trigger diameter of 40 mm and a depth of 8 mm placed at a height of 30 mm was designated D-40-8-h30.



*Fig. 2. FE model a) Column model with trigger on example of D-40-5-h30 model b) Modifiable sketch of trigger position and its geometrical parameters*

## 5. FE model

An analysis of the crushing behavior of the tested columns was carried out using FEM simulations. The Abaqus 2019 code of the Explicit method was used. The model of the thin-walled column with an embossment was made directly in the Part module (Figure 2a), instead of importing the geometry from Catia v5 system, as it was the case in previous publications [14, 15]. This approach allows, by editing the sketch, to quickly modify the geometric parameters, which

significantly reduces the time needed to perform subsequent structural variations, as shown in Figure 2b.

During the development of the FE model, this column was placed between two rigid plates, to which its edges were connected by means of *Tie* bonds. At the geometric centers of the plates, reference points RP were created, where the impact force (at the bottom point) and acceleration, velocity, displacement (at the top point) were recorded during the numerical simulation. The assembly model with reference points is shown in Figure 3a, and with the mesh created using 4-node shell elements S4R in Figure 3b. It can also be seen that a partitioning technique was used to create a uniformly divided mesh.

The contact domain was set up as the *General Contact* option (all with self). In the tangential plane, the contact properties of *Penalty type* were assumed, with a coefficient of friction equal to 0.2. The behaviour in the normal plane was declared as the *Hard Contact* option. On the lower rigid plate all degrees of freedom were disabled, while on the upper plate, which takes the impact, only displacement in the vertical direction was allowed.

The column was subjected to an impact load with a kinetic energy of E= 1.47 kJ corresponding to a mass of m= 60 kg falling with an initial velocity of $V_0$= 7 m/s. Since the research program included experimental validation of the model, the properties of the EN AW6063-T6 material from which the column is made were determined by static tensile testing of samples cut fr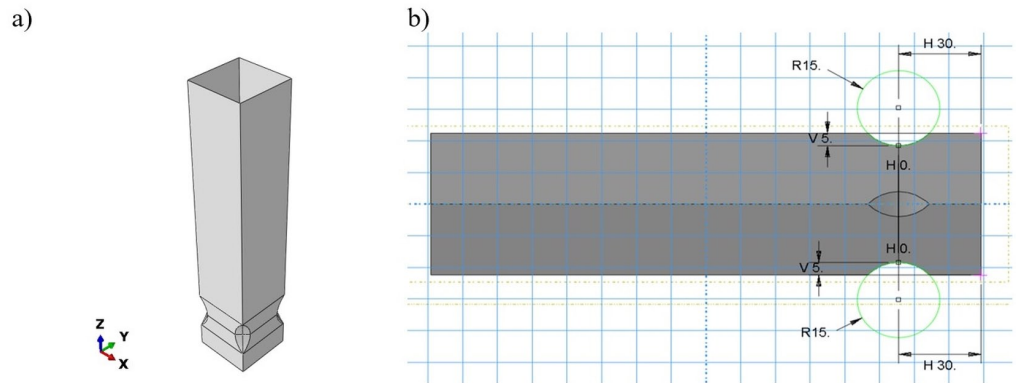om the profile by waterjet. The material properties obtained in this way is shown in Table 1. Since aluminum alloys do not exhibit a significant sensitivity to the strain rate [39], a tri-linear material model was applied, neglecting the effect of the strain rate yet taking account of strain hardening. Constitutive relation does not cover the failure criterion. The numerical analysis of column crushing was carried out in two stages. In the first stage, a buckling analysis was carried out which resulted in buckling modes, of which the first will be needed for further computations. In the second stage, a non-linear geometrical problem of impact with a rigid plate of given energy is analysed, where the first buckling mode, according to which the real structure behaves, is implemented as a geometrical imperfection by means of special procedure. The omission of this procedure would lead to results of the deformation form, and therefore of the other results, significantly deviating from the reality represented by the experiment. The first mode of column buckling is shown in Figure 3c.

*Table 1. Mechanical properties of aluminum alloy AW6063-T6*

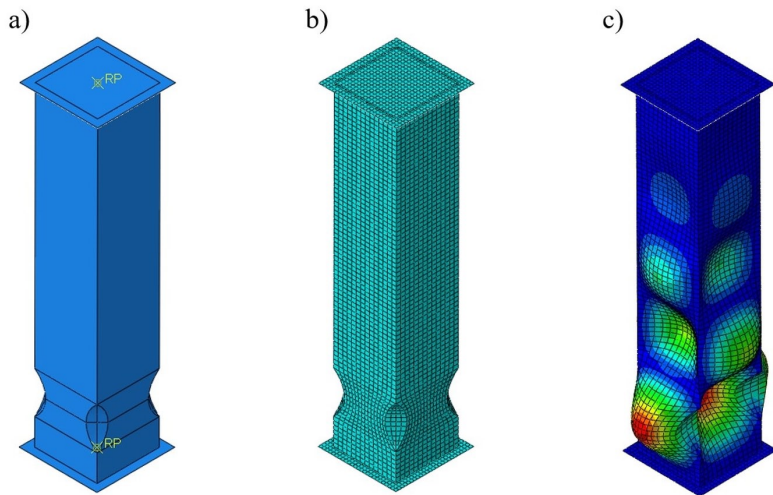| Tri- linear characteristic | | | | Elongation [-] |
|---|---|---|---|---|
| Young's Modulus E [MPa] | 70 000 | Yield Strength $\sigma_Y$ [MPa] | 200 | 0 |
| Density $\rho$ [kg/m³] | 2 700 | | 249.5 | 0.00248 |
| Poisson's ratio $\upsilon$ [-] | 0.33 | Ultimate tensile strength $\sigma_{ult}$ [MPa] | 271.8 | 0.0598 |

Fig. 3. FE model a) Assembly view with rigid plates and reference points b) with mesh c) first buckling mode

## 6. Experimental verification of numerical model

Impact tests were carried out using an Instron CEAST 9350HES drop hammer rig, a general view of which is shown in Figure 4a. Specimen D-40-6-h30 was selected for verification testing, and their designation was extended to include the S1,S2,S3 end. The columns were plugged at the ends by attaching cubes with holes for air evacuation during the crushing process. The column and cubes were mounted on measuring table equipped with a piezoelectric force sensor as can be seen in Figure 4b. The specimens were made from standard aluminum extrusions with square cross-section of 40 mm × 40 mm and wall thickness of 1.2 mm by cold stamping of their edges using a hydraulic press and an instrumentation - a special die consisting of a divided inner part, an outer part and punch, which is shown in Figure 4c.
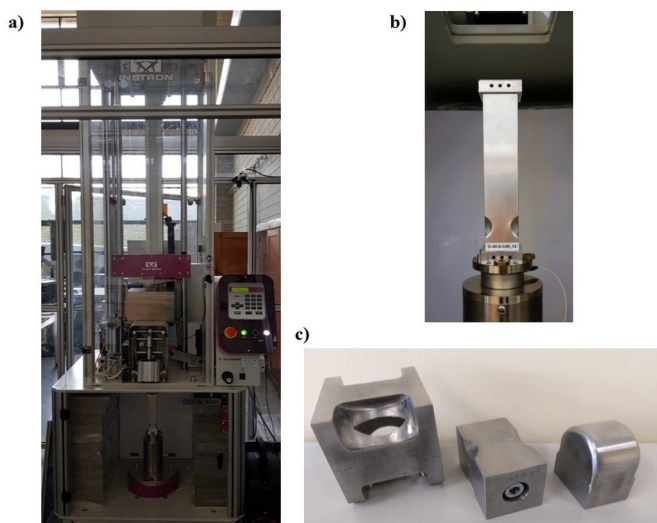


Fig. 4. Investigation and manufacture of triggered columns (a) General view of the drop hammer rig (Instron CEAST 9350HES) (b) Specimen D-40 -6-h30_S1 mounted on a test bench (c) Instrumentation for making dents

As a result of the tests carried out, the load-shortening characteristics were obtained, which are presented in Figure 5 together with the curve for model D-40-6-h30_FEM, obtained by FEM calculations. As can be seen, the waveforms obtained from the drop tower are very close to those obtained from the FEM simulation. This is also a result of validation procedures consisting mainly in a precise adjustment of the mesh density, the main purpose of which was to obtain a similar crush distance. Subsequent force peaks occur at similar deformation stages (similar displacement) and their number is identical as is the number of folds that form as a result of the impact. It also shows that the material model adopted is sufficient to represent well the phenomena occurring during the crush. The FEM model is reliable and gives a good description of reality in terms of characteristics. The deformation forms produced as a result of the impact, despite the apparent slight skewness of specimens S1 and S2, correspond very well to those obtained from the FEM calculations, as can be seen in Figure 6.

As a consequence, the quality of the model is also very good in terms of calculating the energy absorption indicators. Table 2 shows the results of the calculations, which were based on the characteristics mentioned above. The calculations for the experiment are based on three tests and have been averaged for comparison with the FEM results. Similar validation studies on a different machine and at a different drop energy have already been carried out and presented in the paper [15], however, the purpose of the present publication required them to be carried out again due to the change in material, column height and the higher drop energy that can be achieved on the machine the authors have in their possession. The relatively small discrepancies between the calculated values, shows that the developed FEM model can be the basis for extensive parametric analyses and gives credence to the conclusions that will be drawn from it. More comprehensive experimental studies are planned for other projects in the future.
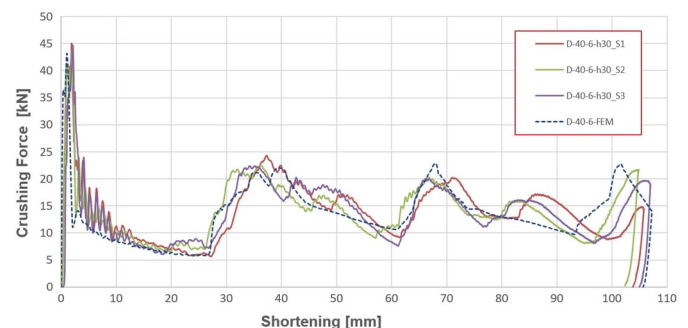


Fig. 5. Crushing force- shortening diagram obtained from experiment and FEM simulation for D-40-6-h30 models

## 7. Parametric study

Using the advanced and experimentally validated FE model, described in Chapter 5, by varying the geometric parameters of the crush initiator and its location, various model configurations were created and
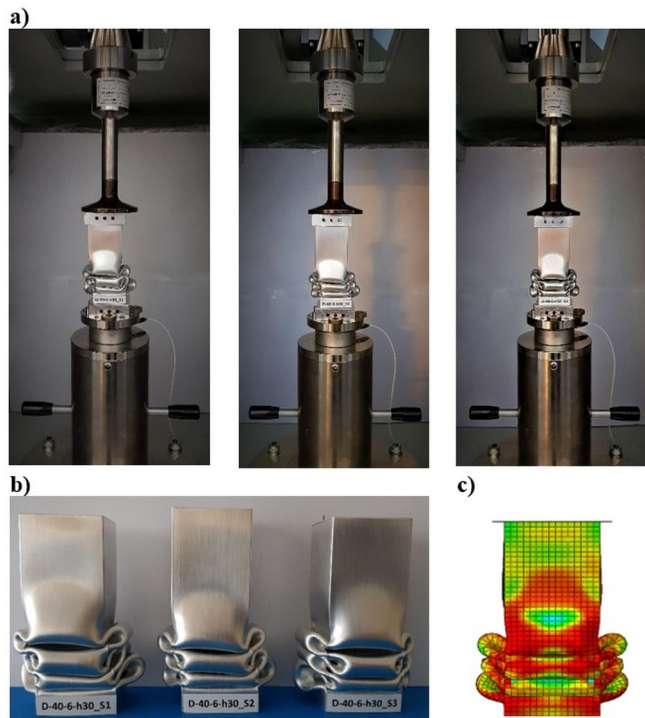
Table 2. Crashworthiness indicators obtained from experiment and FEM simulation, comparison

| Model | EA [J] | U3$_{MAX}$ [mm] | MCF [kN] | PCF [kN] | CLE [%] | SE [-] | TE [%] |
|---|---|---|---|---|---|---|---|
| D-40-6-h30_**S1** | 1484,656 | 105,80 | 14,0324 | 45,0554 | 31,14 | 0,5290 | 16,48 |
| D-40-6-h30_**S2** | 1475,578 | 104,85 | 14,0739 | 43,5271 | 32,33 | 0,5242 | 16,95 |
| D-40-6-h30_**S3** | 1480,098 | 106,99 | 13,8345 | 44,6539 | 30,98 | 0,5349 | 16,57 |
| Experiment Mean Value | 1480,111 | 105,88 | 13,9803 | 44,4121 | 31,49 | 0,5294 | 16,67 |
| D-40-6-h30_**FEM** | 1468,047 | 107,25 | 13,6877 | 43,1998 | 31,68 | 0,5363 | 16,99 |
| Difference [%]* | 0,82 | 1,30 | 2,09 | 2,73 | 0,63 | 1,30 | 1,95 |
| *– in relation to experimental values | | | | | | | |

subjected to numerical simulation of an impact with an energy corresponding to the experiment, in order to determine the crush force-shortening characteristic, which in turn is the basis for determining the energy absorption indicators. Examples of this type of characteristic are shown in Figure 7, with the results presented for a fixed trigger position relative to the base, which is h=30 mm. The individual curves present the results according to a certain key, showing how changing the diameter of the cylindrical embossment influences the curves at a certain embossment depth.

The calculated indicators for this group of models are presented in Table 3. It can easily be seen that the key parameter is the embossing depth, while its diameter is of significance, but much smaller. The value of the crush initiation force PCF is practically determined by the embossing depth. It can be seen not only in the table, but also by analysing successive diagrams (Figure 7), where the value of PCF decreases with increasing embossing depth, while within individual diagrams PCF remains practically constant with little dependence on embossing diameter. It can also be seen from the graphs that in the initial phase of the crush, when the thin-walled structure disturbed by the presence of the trigger is deformed under impact, the value of shortening accompanying the transition of the structure to the phase of forming the first fold increases with the increase of the diameter of the embossing at constant depth. This is very well seen in particular in Figure 7c. The amount of energy absorbed in this crushing phase of the structure is therefore quite varied and affects the subsequent behaviour of the structure.

The minimisation of the PCF force is critical to maximise the value of the CLE index, as the MCF varies little with the change in trigger



Fig. 6. Deformation forms of model D-40-6-h30 (a) deformed specimens just after the test (b) comparison of deformed specimens (c) deformation shape of the FEM model
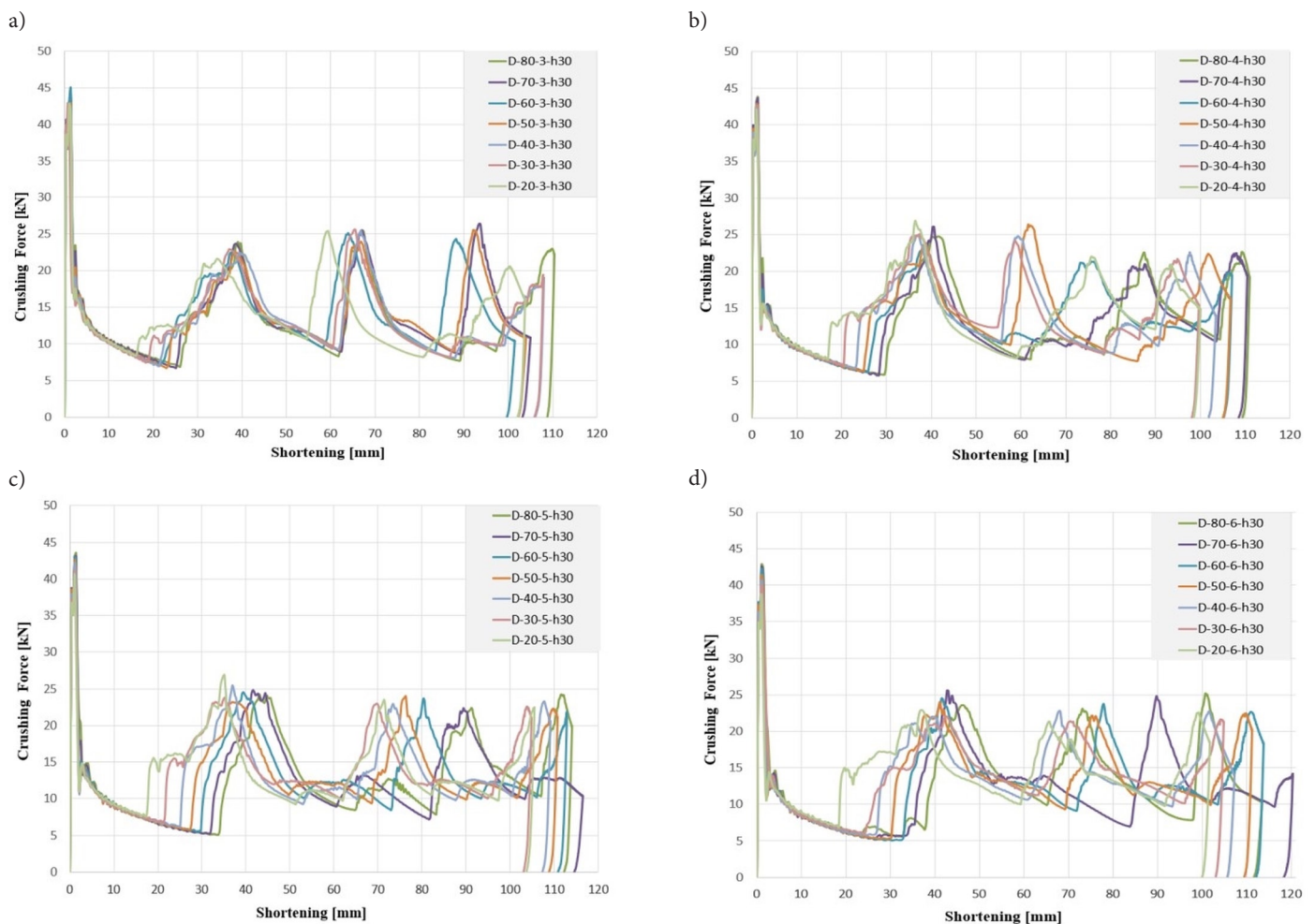


Fig. 7. Crushing force vs shortening diagrams obtained for columns with different trigger geometric parameters and constant trigger position. The individual graphs refer to the changing depth of the trigger for different dents' diameters: a) D=20-80 mm, **g=3 mm**, h=30mm, b) D=20-80 mm, **g=4 mm**, h=30mm, c) D=20-80 mm, **g=5 mm**, h=30mm, d) D=20-80 mm, **g=6 mm**, h=30mm
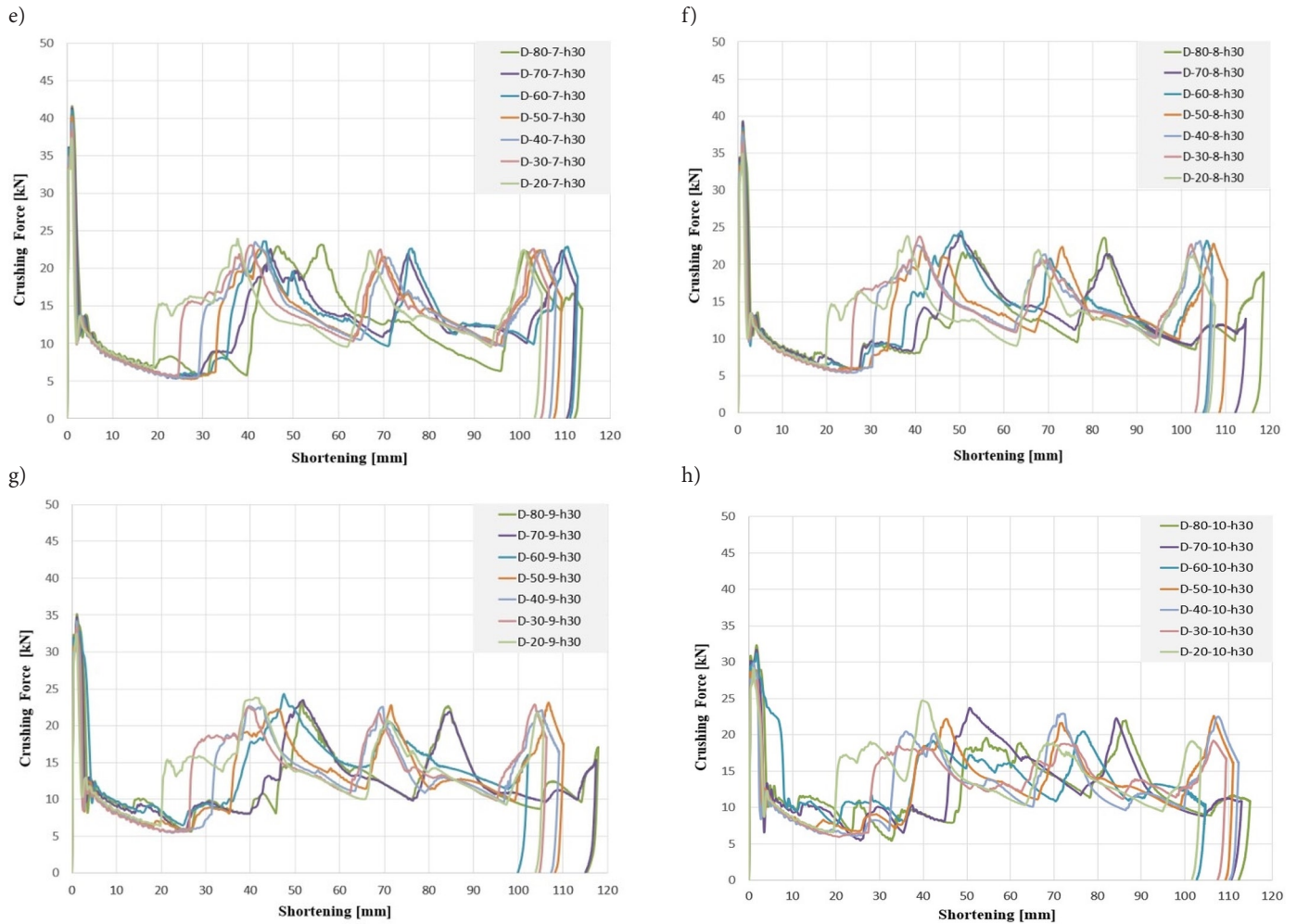
e)



f)



g)



h)



Fig. 7. (cont.) Crushing force vs shortening diagrams obtained for columns with different trigger geometric parameters and constant trigger position. The individual graphs refer to the changing depth of the trigger for different dents' diameters: e) D=20-80 mm, **g=7 mm**, h=30mm, f) D=20-80 mm, **g=8 mm**, h=30mm, g) D=20-80 mm, **g=9 mm**, h=30mm, h) D=20-80 mm, **g=10 mm**, h=30mm

Table 3. Crashworthiness indicators obtained from FEM numerical calculations for specific trigger geometric parameters - Part 1 (fixed trigger position **h**=30 mm, its diameter **D** and depth **g** are varied)

| No | Model Designation | D [mm] | g [mm] | h [mm] | PCF [kN] | MCF [kN] | CLE [%] | SE [-] | TE [%] |
|----|---|---|---|---|---|---|---|---|---|
| 1 | D-20-2-h30 | 20 | 2 | 30 | 47,084 | 12,952 | 27,507 | 0,5669 | 15,595 |
| 2 | D-20-3-h30 | 20 | 3 | 30 | 45,636 | 14,124 | 30,949 | 0,5196 | 16,082 |
| 3 | D-20-4-h30 | 20 | 4 | 30 | 43,877 | 14,678 | 33,454 | 0,4999 | 16,724 |
| 4 | D-20-5-h30 | 20 | 5 | 30 | 41,457 | 13,903 | 33,537 | 0,5277 | 17,697 |
| 5 | D-20-6-h30 | 20 | 6 | 30 | 39,513 | 14,412 | 36,475 | 0,5089 | 18,562 |
| 6 | D-20-7-h30 | 20 | 7 | 30 | 37,211 | 13,967 | 37,533 | 0,5250 | 19,706 |
| 7 | D-20-8-h30 | 20 | 8 | 30 | 34,664 | 13,637 | 39,342 | 0,5380 | 21,165 |
| 8 | D-20-9-h30 | 20 | 9 | 30 | 32,317 | 13,873 | 42,927 | 0,5287 | 22,696 |
| 9 | D-20-10-h30 | 20 | 10 | 30 | 29,381 | 14,152 | 48,168 | 0,5185 | 24,975 |
| 10 | D-30-2-h30 | 30 | 2 | 30 | 47,537 | 13,339 | 28,060 | 0,5505 | 15,456 |
| 11 | D-30-3-h30 | 30 | 3 | 30 | 45,940 | 13,595 | 29,593 | 0,5396 | 15,968 |
| 12 | D-30-4-h30 | 30 | 4 | 30 | 44,467 | 14,711 | 33,084 | 0,4988 | 16,501 |
| 13 | D-30-5-h30 | 30 | 5 | 30 | 42,271 | 13,988 | 33,091 | 0,5243 | 17,350 |
| 14 | D-30-6-h30 | 30 | 6 | 30 | 40,897 | 13,978 | 34,179 | 0,5247 | 17,934 |
| 15 | D-30-7-h30 | 30 | 7 | 30 | 40,897 | 13,776 | 33,685 | 0,5323 | 17,929 |
| 16 | D-30-8-h30 | 30 | 8 | 30 | 36,308 | 13,966 | 38,466 | 0,5249 | 20,191 |
| 17 | D-30-9-h30 | 30 | 9 | 30 | 33,389 | 13,789 | 41,298 | 0,5317 | 21,958 |

| 18 | D-30-10-h30 | 30 | 10 | 30 | 30,140 | 13,423 | 44,535 | 0,5476 | 24,388 |
|---|---|---|---|---|---|---|---|---|---|
| 19 | D-40-2-h30 | 40 | 2 | 30 | 47,567 | 13,123 | 27,588 | 0,5601 | 15,451 |
| 20 | D-40-3-h30 | 40 | 3 | 30 | 46,233 | 13,569 | 29,348 | 0,5406 | 15,867 |
| 21 | D-40-4-h30 | 40 | 4 | 30 | 45,123 | 14,185 | 31,437 | 0,5171 | 16,256 |
| 22 | D-40-5-h30 | 40 | 5 | 30 | 43,200 | 13,443 | 31,119 | 0,5459 | 16,988 |
| 23 | D-40-6-h30 | 40 | 6 | 30 | 41,152 | 13,688 | 33,262 | 0,5363 | 17,837 |
| 24 | D-40-7-h30 | 40 | 7 | 30 | 39,516 | 13,572 | 34,345 | 0,5407 | 18,570 |
| 25 | D-40-8-h30 | 40 | 8 | 30 | 37,758 | 13,682 | 36,236 | 0,5361 | 19,425 |
| 26 | D-40-9-h30 | 40 | 9 | 30 | 34,100 | 13,441 | 39,416 | 0,5456 | 21,505 |
| 27 | D-40-10-h30 | 40 | 10 | 30 | 30,958 | 13,046 | 42,141 | 0,5619 | 23,678 |
| 28 | D-50-2-h30 | 50 | 2 | 30 | 47,599 | 13,481 | 28,322 | 0,5468 | 15,486 |
| 29 | D-50-3-h30 | 50 | 3 | 30 | 46,506 | 14,118 | 30,357 | 0,5197 | 15,778 |
| 30 | D-50-4-h30 | 50 | 4 | 30 | 45,222 | 13,751 | 30,408 | 0,5334 | 16,218 |
| 31 | D-50-5-h30 | 50 | 5 | 30 | 43,493 | 13,254 | 30,474 | 0,5536 | 16,871 |
| 32 | D-50-6-h30 | 50 | 6 | 30 | 42,108 | 13,194 | 31,334 | 0,5561 | 17,426 |
| 33 | D-50-7-h30 | 50 | 7 | 30 | 40,316 | 13,428 | 33,306 | 0,5463 | 18,195 |
| 34 | D-50-8-h30 | 50 | 8 | 30 | 37,891 | 13,292 | 35,080 | 0,5519 | 19,360 |
| 35 | D-50-9-h30 | 50 | 9 | 30 | 34,221 | 13,320 | 38,924 | 0,5505 | 21,427 |
| 36 | D-50-10-h30 | 50 | 10 | 30 | 31,633 | 13,195 | 41,713 | 0,5558 | 23,183 |
| 37 | D-60-2-h30 | 60 | 2 | 30 | 47,597 | 13,370 | 28,089 | 0,5496 | 15,438 |
| 38 | D-60-3-h30 | 60 | 3 | 30 | 46,378 | 14,467 | 31,194 | 0,5073 | 15,824 |
| 39 | D-60-4-h30 | 60 | 4 | 30 | 45,130 | 13,705 | 30,367 | 0,5352 | 16,251 |
| 40 | D-60-5-h30 | 60 | 5 | 30 | 43,954 | 12,999 | 29,574 | 0,5645 | 16,695 |
| 41 | D-60-6-h30 | 60 | 6 | 30 | 42,560 | 12,894 | 30,295 | 0,5693 | 17,248 |
| 42 | D-60-7-h30 | 60 | 7 | 30 | 41,185 | 13,008 | 31,583 | 0,5639 | 17,810 |
| 43 | D-60-8-h30 | 60 | 8 | 30 | 38,553 | 13,700 | 35,535 | 0,5352 | 19,018 |
| 44 | D-60-9-h30 | 60 | 9 | 30 | 36,382 | 14,354 | 39,453 | 0,5110 | 20,159 |
| 45 | D-60-10-h30 | 60 | 10 | 30 | 32,490 | 14,002 | 43,096 | 0,5239 | 22,577 |
| 46 | D-70-2-h30 | 70 | 2 | 30 | 47,574 | 13,279 | 27,912 | 0,5533 | 15,445 |
| 47 | D-70-3-h30 | 70 | 3 | 30 | 46,589 | 13,982 | 30,011 | 0,5251 | 15,757 |
| 48 | D-70-4-h30 | 70 | 4 | 30 | 45,332 | 13,288 | 29,313 | 0,5522 | 16,187 |
| 49 | D-70-5-h30 | 70 | 5 | 30 | 44,327 | 12,608 | 28,443 | 0,5827 | 16,575 |
| 50 | D-70-6-h30 | 70 | 6 | 30 | 43,039 | 12,210 | 28,369 | 0,6016 | 17,067 |
| 51 | D-70-7-h30 | 70 | 7 | 30 | 41,237 | 13,057 | 31,662 | 0,5620 | 17,795 |
| 52 | D-70-8-h30 | 70 | 8 | 30 | 39,686 | 12,814 | 32,288 | 0,5727 | 18,492 |
| 53 | D-70-9-h30 | 70 | 9 | 30 | 36,262 | 12,497 | 34,462 | 0,5875 | 20,246 |
| 54 | D-70-10-h30 | 70 | 10 | 30 | 33,020 | 12,990 | 39,340 | 0,5647 | 22,216 |
| 55 | D-80-2-h30 | 80 | 2 | 30 | 47,579 | 13,474 | 28,320 | 0,5452 | 15,440 |
| 56 | D-80-3-h30 | 80 | 3 | 30 | 46,639 | 13,292 | 28,500 | 0,5520 | 15,733 |
| 57 | D-80-4-h30 | 80 | 4 | 30 | 45,530 | 13,221 | 29,038 | 0,5551 | 16,118 |
| 58 | D-80-5-h30 | 80 | 5 | 30 | 44,644 | 12,872 | 28,832 | 0,5701 | 16,437 |
| 59 | D-80-6-h30 | 80 | 6 | 30 | 43,581 | 12,939 | 29,690 | 0,5673 | 16,843 |
| 60 | D-80-7-h30 | 80 | 7 | 30 | 41,973 | 12,883 | 30,694 | 0,5695 | 17,479 |
| 61 | D-80-8-h30 | 80 | 8 | 30 | 39,730 | 12,378 | 31,156 | 0,5927 | 18,467 |
| 62 | D-80-9-h30 | 80 | 9 | 30 | 36,995 | 12,443 | 33,635 | 0,5895 | 19,827 |
| 63 | D-80-10-h30 | 80 | 10 | 30 | 34,144 | 12,772 | 37,407 | 0,5745 | 21,490 |

*Table 4. Crashworthiness indicators obtained from FEM numerical calculations for specific trigger geometric parameters- Part 2 (fixed trigger diameter **D**=40 mm, its depth g and position h are varied)*

| No | Model Designation | D [mm] | g [mm] | h [mm] | PCF [kN] | MCF [kN] | CLE [%] | SE [-] | TE [%] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | D-40-2-h20 | 40 | 2 | 20 | 47,768 | 13,305 | 27,853 | 0,5509 | 15,345 |
| 2 | D-40-2-h30 | 40 | 2 | 30 | 47,567 | 13,123 | 27,588 | 0,5601 | 15,451 |
| 3 | D-40-2-h40 | 40 | 2 | 40 | 47,832 | 13,364 | 27,939 | 0,5484 | 15,322 |
| 4 | D-40-2-h50 | 40 | 2 | 50 | 48,522 | 13,099 | 26,996 | 0,5593 | 15,099 |
| 5 | D-40-2-h60 | 40 | 2 | 60 | 48,838 | 13,748 | 28,150 | 0,5328 | 14,998 |
| 6 | D-40-2-h70 | 40 | 2 | 70 | 48,847 | 14,092 | 28,849 | 0,5203 | 15,009 |
| 7 | D-40-2-h80 | 40 | 2 | 80 | 48,635 | 14,395 | 29,598 | 0,5093 | 15,075 |
| 8 | D-40-2-h90 | 40 | 2 | 90 | 48,984 | 14,574 | 29,753 | 0,5032 | 14,971 |
| 9 | D-40-2-h100 | 40 | 2 | 100 | 49,264 | 14,845 | 30,134 | 0,4937 | 14,878 |
| 10 | D-40-4-h20 | 40 | 4 | 20 | 45,416 | 12,827 | 28,243 | 0,5714 | 16,138 |
| 11 | D-40-4-h30 | 40 | 4 | 30 | 45,123 | 14,185 | 31,437 | 0,5171 | 16,256 |
| 12 | D-40-4-h40 | 40 | 4 | 40 | 44,656 | 13,539 | 30,318 | 0,5412 | 16,410 |
| 13 | D-40-4-h50 | 40 | 4 | 50 | 45,242 | 14,494 | 32,037 | 0,5056 | 16,197 |
| 14 | D-40-4-h60 | 40 | 4 | 60 | 46,608 | 13,429 | 28,813 | 0,5453 | 15,711 |
| 15 | D-40-4-h70 | 40 | 4 | 70 | 47,464 | 13,739 | 28,946 | 0,5333 | 15,436 |
| 16 | D-40-4-h80 | 40 | 4 | 80 | 47,694 | 13,787 | 28,907 | 0,5313 | 15,358 |
| 17 | D-40-4-h90 | 40 | 4 | 90 | 47,982 | 13,476 | 28,086 | 0,5436 | 15,266 |
| 18 | D-40-4-h100 | 40 | 4 | 100 | 48,573 | 13,522 | 27,839 | 0,5418 | 15,082 |
| 19 | D-40-6-h20 | 40 | 6 | 20 | 41,533 | 13,522 | 32,557 | 0,5347 | 17,409 |
| 20 | D-40-6-h30 | 40 | 6 | 30 | 41,152 | 13,688 | 33,262 | 0,5363 | 17,837 |
| 21 | D-40-6-h40 | 40 | 6 | 40 | 41,340 | 13,522 | 32,709 | 0,5516 | 18,043 |
| 22 | D-40-6-h50 | 40 | 6 | 50 | 41,909 | 13,522 | 32,265 | 0,5577 | 17,996 |
| 23 | D-40-6-h60 | 40 | 6 | 60 | 42,358 | 13,522 | 31,923 | 0,5480 | 17,493 |
| 24 | D-40-6-h70 | 40 | 6 | 70 | 42,502 | 13,522 | 31,815 | 0,5534 | 17,608 |
| 25 | D-40-6-h80 | 40 | 6 | 80 | 42,563 | 13,522 | 31,769 | 0,5666 | 17,999 |
| 26 | D-40-6-h90 | 40 | 6 | 90 | 42,927 | 13,522 | 31,500 | 0,5612 | 17,678 |
| 27 | D-40-6-h100 | 40 | 6 | 100 | 43,653 | 13,522 | 30,976 | 0,5600 | 17,346 |
| 28 | D-40-8-h20 | 40 | 8 | 20 | 36,941 | 12,649 | 34,241 | 0,5797 | 19,851 |
| 29 | D-40-8-h30 | 40 | 8 | 30 | 37,758 | 13,682 | 36,236 | 0,5361 | 19,425 |
| 30 | D-40-8-h40 | 40 | 8 | 40 | 37,026 | 12,599 | 34,027 | 0,5818 | 19,798 |
| 31 | D-40-8-h50 | 40 | 8 | 50 | 39,158 | 12,986 | 33,163 | 0,5630 | 18,671 |
| 32 | D-40-8-h60 | 40 | 8 | 60 | 39,993 | 13,209 | 33,028 | 0,5545 | 18,315 |
| 33 | D-40-8-h70 | 40 | 8 | 70 | 40,164 | 12,843 | 31,976 | 0,5692 | 18,202 |
| 34 | D-40-8-h80 | 40 | 8 | 80 | 40,321 | 12,722 | 31,552 | 0,5763 | 18,184 |
| 35 | D-40-8-h90 | 40 | 8 | 90 | 40,395 | 12,838 | 31,781 | 0,5717 | 18,170 |
| 36 | D-40-8-h100 | 40 | 8 | 100 | 40,526 | 12,995 | 32,066 | 0,5643 | 18,095 |
| 37 | D-40-10-h20 | 40 | 10 | 20 | 31,032 | 13,188 | 42,498 | 0,5564 | 23,646 |
| 38 | D-40-10-h30 | 40 | 10 | 30 | 30,958 | 13,046 | 42,141 | 0,5619 | 23,678 |
| 39 | D-40-10-h40 | 40 | 10 | 40 | 31,137 | 12,589 | 40,431 | 0,5826 | 23,557 |
| 40 | D-40-10-h50 | 40 | 10 | 50 | 35,219 | 12,799 | 36,341 | 0,5731 | 20,828 |
| 41 | D-40-10-h60 | 40 | 10 | 60 | 36,485 | 12,876 | 35,291 | 0,5696 | 20,102 |
| 42 | D-40-10-h70 | 40 | 10 | 70 | 36,633 | 12,777 | 34,878 | 0,5733 | 19,997 |
| 43 | D-40-10-h80 | 40 | 10 | 80 | 36,804 | 12,490 | 33,937 | 0,5872 | 19,927 |
| 44 | D-40-10-h90 | 40 | 10 | 90 | 39,109 | 12,928 | 33,056 | 0,5674 | 18,757 |
| 45 | D-40-10-h100 | 40 | 10 | 100 | 39,010 | 12,866 | 32,981 | 0,5704 | 18,812 |

*Table 5. Crashworthiness indicators obtained from FEM numerical calculations for specific trigger geometric parameters- Part 3 (other configurations)*

| No | Model Designation | D [mm] | g [mm] | h [mm] | PCF [kN] | MCF [kN] | CLE [%] | SE [-] | TE [%] |
|----|-------------------|--------|--------|--------|----------|----------|---------|--------|--------|
| 1 | D-20-2-h20 | 20 | 2 | 20 | 47,884 | 13,463 | 28,115 | 0,5428 | 15,261 |
| 2 | D-20-3-h100 | 20 | 2 | 100 | 49,331 | 14,493 | 29,379 | 0,5086 | 14,943 |
| 3 | D-20-6-h60 | 20 | 6 | 60 | 42,124 | 13,841 | 32,858 | 0,5326 | 17,501 |
| 4 | D-20-10-h20 | 20 | 10 | 20 | 28,956 | 13,459 | 46,480 | 0,5461 | 25,382 |
| 5 | D-20-10-h100 | 20 | 10 | 100 | 39,195 | 13,730 | 35,031 | 0,5260 | 18,427 |
| 6 | D-40-10-h20 | 40 | 10 | 20 | 30,982 | 13,244 | 42,747 | 0,5558 | 23,757 |
| 7 | D-40-10-h80 | 40 | 10 | 80 | 38,145 | 13,014 | 34,118 | 0,5662 | 19,317 |
| 8 | D-60-5-h80 | 60 | 5 | 80 | 47,061 | 12,989 | 27,601 | 0,5676 | 15,666 |
| 9 | D-60-7-h50 | 60 | 7 | 50 | 40,929 | 12,994 | 31,747 | 0,5671 | 18,003 |
| 10 | D-80-2-h20 | 80 | 2 | 20 | 47,901 | 13,350 | 27,870 | 0,5521 | 15,386 |
| 11 | D-80-2-h100 | 80 | 2 | 100 | 49,318 | 14,586 | 29,575 | 0,5054 | 14,946 |
| 12 | D-80-6-h60 | 80 | 6 | 60 | 44,499 | 12,337 | 27,724 | 0,5978 | 16,574 |
| 13 | D-80-10-h20 | 80 | 10 | 100 | 39,555 | 13,396 | 33,867 | 0,5509 | 18,656 |

depth. The highest energy-absorbing CLE indicators are obtained for maximum embossing depth and minimum embossing diameter.

In order to broaden the base of results, which is necessary due to the quality of the neural network being built and its predictive capabilities, further design variations of the energy absorbers were modelled and numerical calculations were carried out according to an identical procedure. The results of these calculations are shown in Table 4, where the results are given for a fixed trigger diameter and the depth of the embossing and its position (in the range up to 100 mm) was changed. The next Table 5 contains the results of models characterized by mostly boundary geometric parameter values. Obtaining these data significantly improved the prediction quality of the network.

## 8. Multilayer perceptron networks study

Neural networks are signal processing mathematical models. The most popular structure of Artificial Neural Network is multilayer perceptron. This network model consists of numerous neurons set in layers. Regardless of the type, each neural network uses numerical input values and sets numerical output values.

In regression issues, the aim is to estimate the value of the continuous output variable. The values of PCF, CLE, SE, TE indicators from the numerical experiment are the values of output variables. Trigger parameters were taken as input variables. These are position (h), depth (g) and diameter (D), which are shown in Tables 3-5. These parameters are discussed in detail in Chapter 4.

Regression problems are represented by data sets, in which the output variable is numerical. Figure 8 shows one of the most frequently used activation functions used in the research - the logistic activation function.
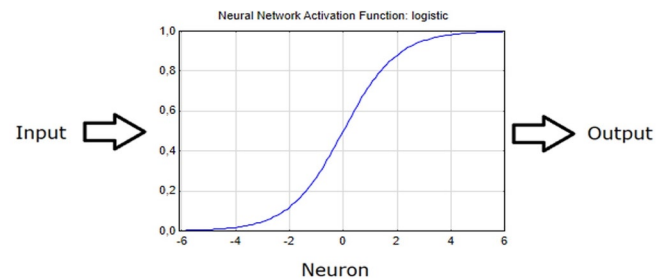


*Fig. 8. Logistic activation function*

This is an S-function. Output values bound between 0 and 1, normalizing the output of each neuron. This type of activation function enables clear predictions. For x above 2 or below -2, the prediction value is close to 1 or 0. The logistical function is smooth and easily differentiated, these features enable the network training algorithm to work.

The function can be represented by the equation:

$$u(x) = \frac{1}{1 + e^{-x}} \qquad (4)$$

Two MLP neural networks were used in the numerical study. Based on preliminary experiments, it was decided to test networks with four and five neurons in the hidden layer. Table 6 lists the values of the training rates for the MLP networks with respect to predicted indicators.

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) training method is a quasi-Newton method. In numerical optimization, the algorithm

*Table 6. Artificial neural network operation parameters*

| Network | Quality | | | Training algorithm | Error function | Activation (Hidden) | Activation (Output) |
|---------|---------|---------|-----------|--------------------|----------------|---------------------|---------------------|
| | Training | Testing | Validation | | | | |
| MLP 3-4-4 | 0,895826 | 0,904728 | 0,977966 | BFGS 34 | SOS | Logistic | Logistic |
| MLP 3-5-4 | 0,893355 | 0,892867 | 0,977768 | BFGS 27 | SOS | Logistic | Logistic |

*Table 7. Networks sensitivity (training set)*

| Network | Diameter D | Depth g | Position h |
|---------|-----------|---------|-----------|
| MLP 3-4-4 | 2,364994 | 17,73557 | 4,105513 |
| MLP 3-5-4 | 2,262062 | 17,07618 | 3,951909 |
| Mean value | 2,313528 | 17,40587 | 4,028711 |

*Table 8. Correlation rates*

| | PCF | | | CLE | | | SE | | | TE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Training | Testing | Valida-tion | Training | Testing | Valida-tion | Training | Testing | Valida-tion | Training | Testing | Valida-tion |
| MLP 3-4-4 | 0,98828 | 0,99415 | 0,99515 | 0,94571 | 0,97153 | 0,99600 | 0,66706 | 0,66181 | 0,92606 | 0,98225 | 0,99143 | 0,99466 |
| MLP 3-5-4 | 0,98645 | 0,98932 | 0,99428 | 0,94480 | 0,97031 | 0,99385 | 0,66068 | 0,62273 | 0,92870 | 0,98150 | 0,98913 | 0,99424 |
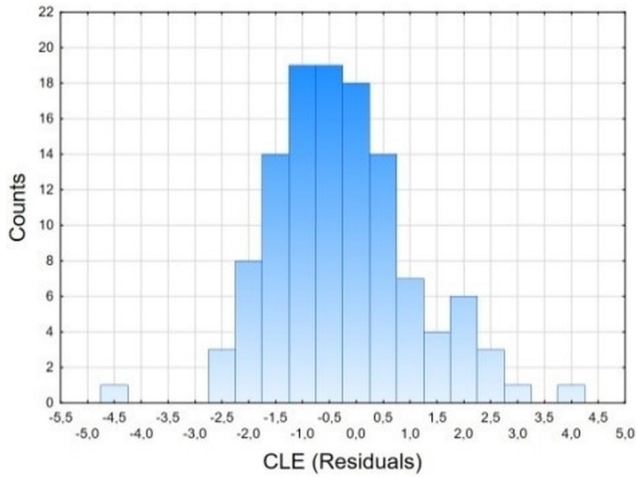


Fig. 9. Residual analysis for the CLE parameter
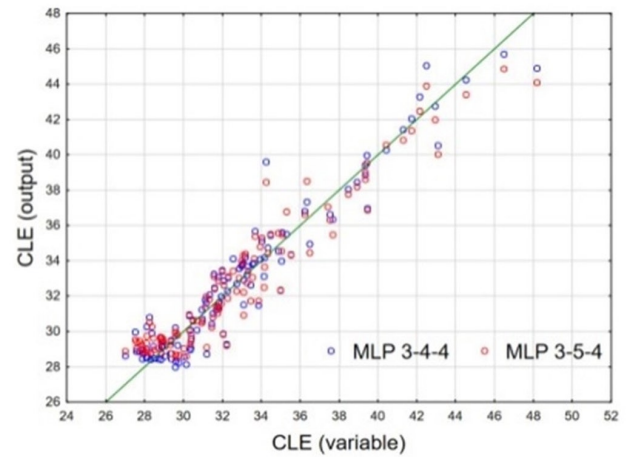


Fig. 12. Analysis of the forecast to the observed value for the indicator
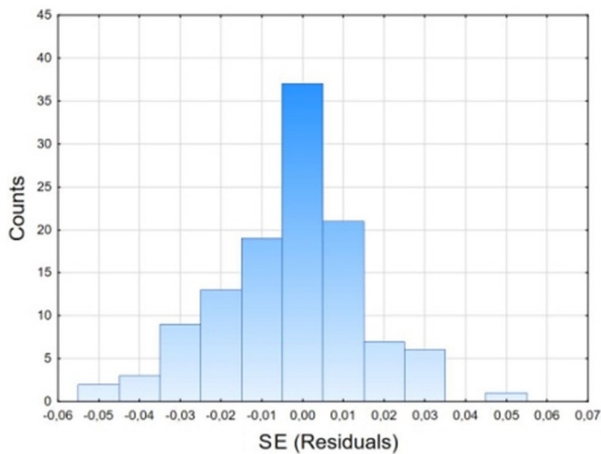


Fig. 10. Residual analysis for the SE parameter
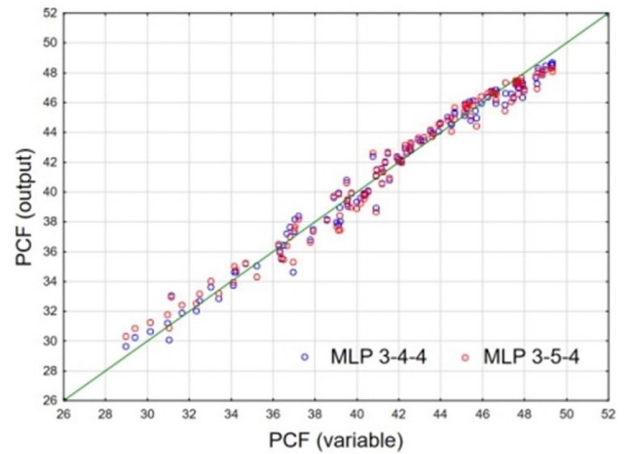


Fig. 13. Analysis of the forecast to the observed value for the indicator
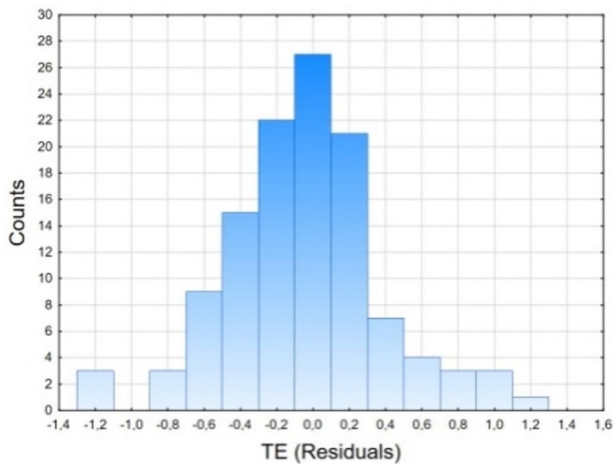


Fig. 11. Residual analysis for the TE parameter
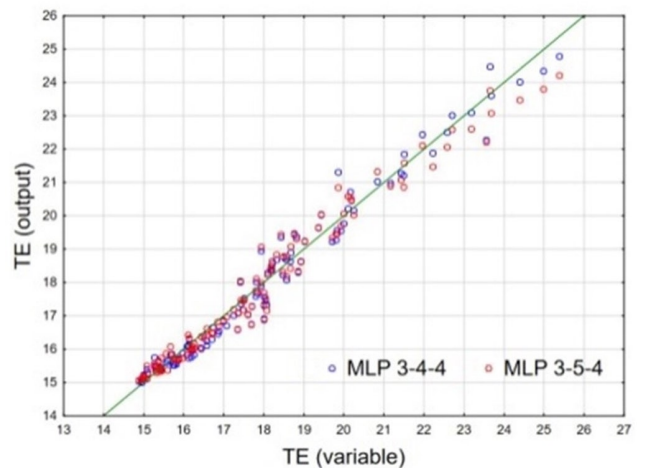


Fig. 14. Analysis of the forecast to the observed value for the indicator

BFGS is an iterative method for solving non-linear optimization problems. The error function is the sum of the squares (SOS) of the deviations between the reference value and the network output.

The sensitivity analysis shows that the depth of the trigger is the most important parameter (Table 7). Another important parameter is the position of the initiator. The least important is the diameter of the dent.

The parameters of neural models, except SE, show high correlation for all datasets: training, testing and validation (Table 8).

It is worthwhile to examine the distribution of the residuals, i.e. the differences between the output variable and its prediction (Figures 9-11).

For both active networks, the histograms indicate that the residuals are more or less normally distributed around zero, which is in line with the general assumption of normal noise in the data. Wider histograms indicate more noise. A histogram that is too narrow compared to the amount of noise in the data would indicate that the network is overfitting. Histogram examination shows that the residues are close to normal distribution with a zero average.

It is a good indication that the network has discovered the assumed noise model. We assume that noise on the target variables is normally distributed with zero mean and an unknown variance.

Figures 12-14 show the relationship of the results obtained by the neural network depending on the actual value. These graphs show the effectiveness of the two MLP networks analysed.

The figures 15-22 show the values of the analysed indicators depending on the geometric parameters and position of the trigger.

It can be seen from the above graphs that in some cases, it is possible to fit a surface plot representing the function of the dependence of the CLE, PCF, SE and TE parameters on the input geometric parameters.

Figures 15 and 17 show that especially changing the dent depth definitely affects the CLE, PCF, TE and SE parameters. The diameter of the trigger particularly influences the determination of the SE parameter.

Figures 16 and 18 indicate that a significant change in CLE and PCF indicators occurs with a high depth and low trigger position configuration.

## 9. Final conclusions

The conducted research has shown that artificial neural networks can be applied to predict the values of crashworthiness indicators with an acceptable error.
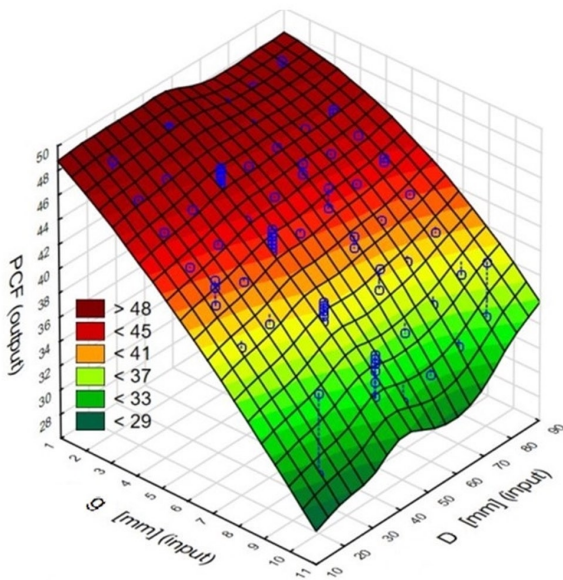


*Fig. 15. PCF value as a function of depth g and diameter D*
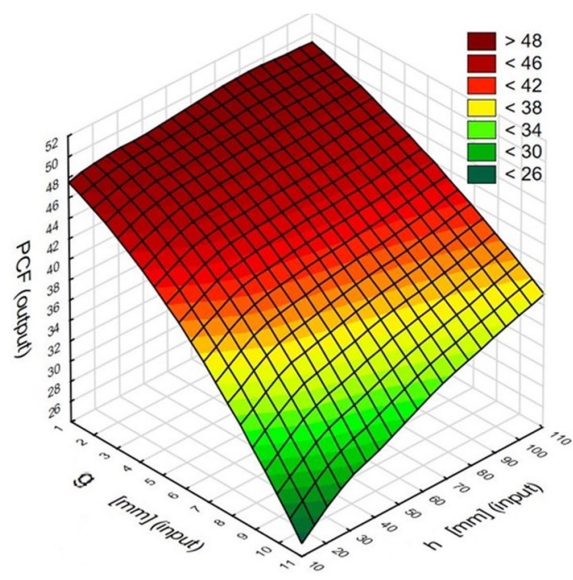


*Fig. 16. PCF value as a function of depth g and position h*
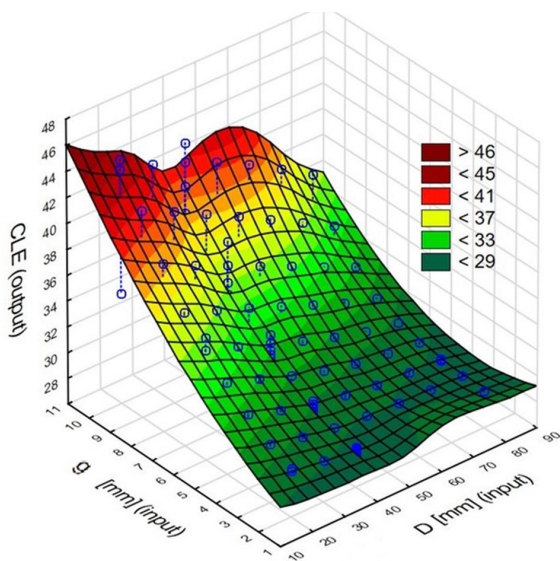


*Fig. 17. CLE value as a function of depth g and diameter D*
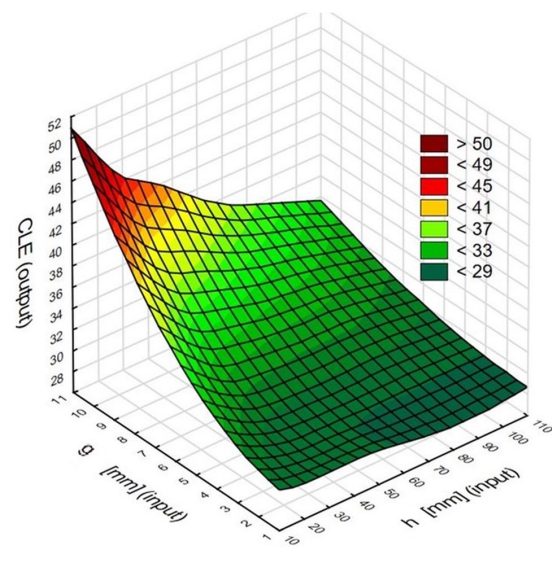


*Fig. 18. CLE value as a function of depth g and position h*
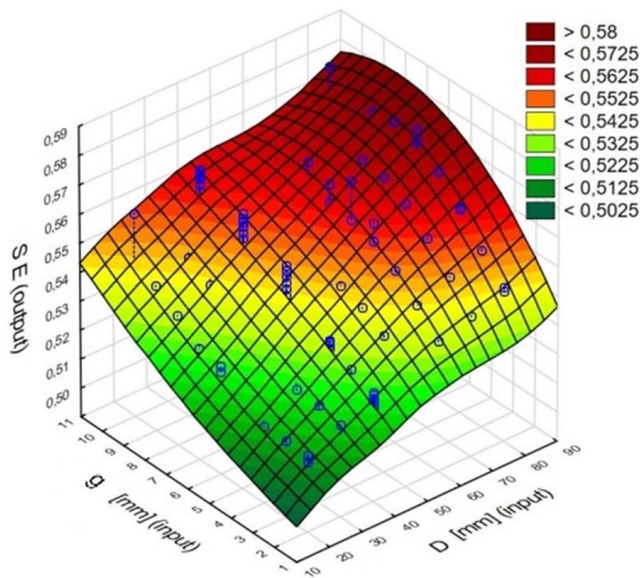
*Fig. 19. SE value as a function of depth g and diameter D*
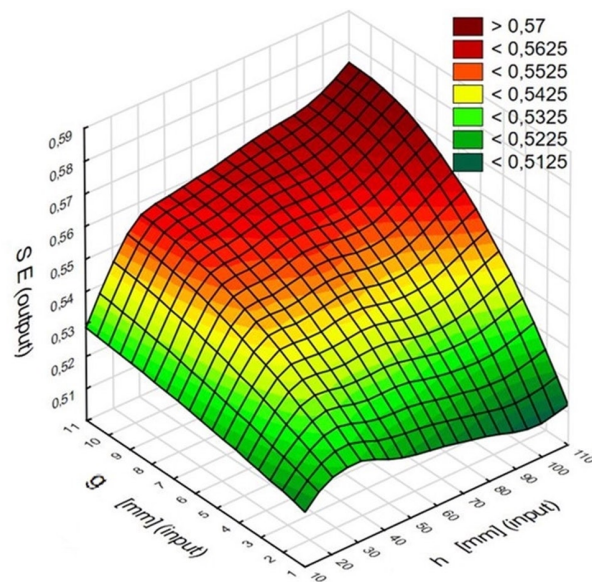


*Fig. 20. SE value as a function of depth g and position h*
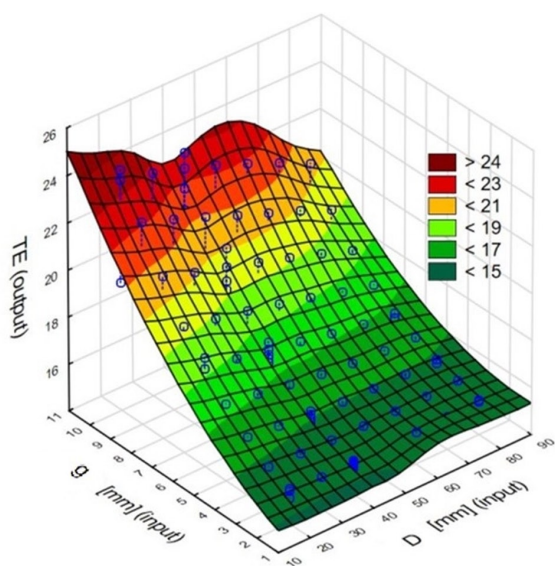


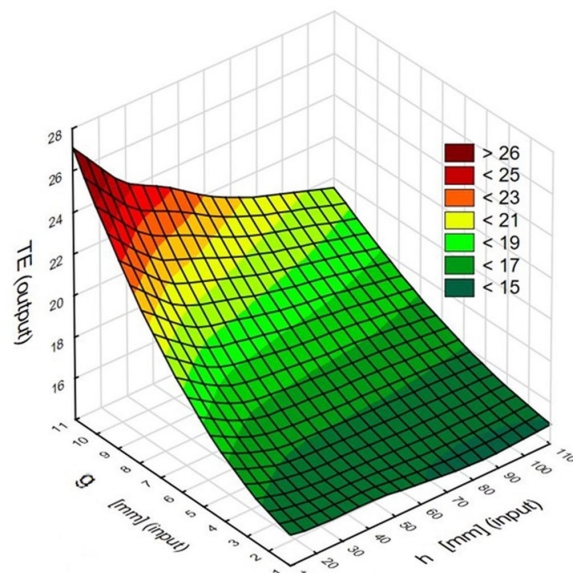*Fig. 21. TE value as a function of depth g and diameter D*



*Fig. 22. TE value as a function of depth g and position h*

There are minor differences between the network models studied. It is important to note that the simpler the model the faster it operates. Quality differences between networks with four and five neurons in the hidden layer are below 2% (Table 6).

From the network sensitivity analysis it can be concluded that by far the most important parameter for the performance of the model is the trigger depth. The position of the trigger and the diameter affect the performance of the MLP to a lesser extent. This conclusion should be taken into account when designing energy-absorbing models.

The overall quality of the network performance was about 90% for the training set and over 97% for the test data. The correlation coefficients indicate that the SE coefficient was the most difficult to predict. For the PCF, CLE and TE coefficients the prediction quality was 95-99% (Table 8).

In order to confirm the universality of the neural network performance, an additional numerical experiment was introduced. The random models, including two with parameters outside the range that was used to training the ANN were tested (D-80-4-h140, D-100-8-h50). Figure 23 shows the crushing force-shortening characteristics and Table 9 shows the indicator results for the random test specimens.
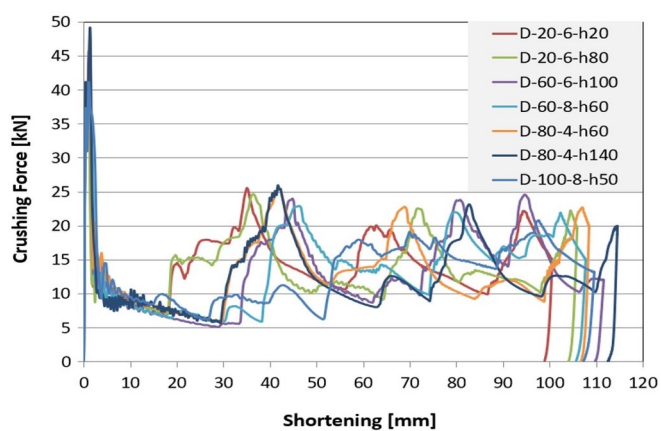


*Fig. 23. Crushing force vs shortening diagrams obtained for testing models*

Predictions of indicators were then made for the test specimens and their values were compared with the results obtained by the FE

---

| Designation | PCF [kN] | MCF [kN] | CLE [%] | SE [-] | TE [%] |
|---|---|---|---|---|---|
| D-20-6-h20 | 38,898 | 14,657 | 37,680 | 0,5022 | 18,923 |
| D-20-6-h80 | 40,751 | 13,915 | 34,148 | 0,5293 | 18,074 |
| D-60-6-h100 | 45,712 | 13,191 | 28,857 | 0,5580 | 16,103 |
| D-60-8-h60 | 39,416 | 13,190 | 33,461 | 0,5388 | 18,859 |
| D-80-4-h60 | 47,370 | 13,597 | 28,703 | 0,5422 | 15,563 |
| D-80-4-h140 | 48,918 | 12,861 | 26,291 | 0,5721 | 15,040 |
| D-100-8-h50 | 41,058 | 13,442 | 32,739 | 0,5472 | 17,915 |

slightly exceed 4,8%. Good prediction properties are shown by both networks when it comes to the calculation of TE, which is after all the product of SE and CLE, despite the relatively high differences obtained in the prediction of individual factors. This is because these factors are predicted most often simultaneously with over- and under-estimates.

As a practical aspect of the research carried out, it should be emphasised that the use of multilayer perceptron can successfully speed up the design process.

The study carried out, although it concerns an energy absorber with a specific type of trigger, can be successfully extended to energy absorbers with other types of trigger, provided that it is characterised by several geometrical parameters whose interaction is difficult to grasp. As part of such work, a neural network is planned for an energoabsorber with a trigger in the form of spheroidal embossments on the

*Table 10. Crashworthiness indicators obtained from MLP 3-4-4 NN and relative error values in relation to FE simulation results*

| | | | | | | Relative error (difference) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Designation | PCF [kN] | MCF [kN] | CLE [%] | SE [-] | TE [%] | $\Delta_{PCF}$ [%] | $\Delta_{MCF}$ [%] | $\Delta_{CLE}$ [%] | $\Delta_{SE}$ [%] | $\Delta_{TE}$ [%] |
| D-20-6-h20 | 39,434 | 14,355 | 36,243 | 0,5137 | 18,218 | 1,38 | 2,06 | 3,81 | 2,29 | 3,72 |
| D-20-6-h80 | 42,000 | 13,780 | 32,753 | 0,5381 | 17,463 | 3,07 | 0,97 | 4,08 | 1,67 | 3,38 |
| D-60-6-h100 | 45,965 | 13,275 | 29,894 | 0,5537 | 16,612 | 0,55 | 0,64 | 3,59 | 0,77 | 3,16 |
| D-60-8-h60 | 39,838 | 12,918 | 32,502 | 0,5698 | 18,627 | 1,07 | 2,06 | 2,86 | 5,55 | 3,13 |
| D-80-4-h60 | 47,196 | 13,141 | 28,558 | 0,5565 | 15,784 | 0,37 | 3,35 | 0,50 | 2,64 | 1,42 |
| D-80-4-h140 | 48,718 | 13,508 | 28,027 | 0,5407 | 15,439 | 0,41 | 5,03 | 6,60 | 5,48 | 2,65 |
| D-100-8-h50 | 42,603 | 12,568 | 29,946 | 0,5827 | 17,614 | 3,76 | 6,50 | 8,53 | 6,49 | 1,68 |

*Table 11. Crashworthiness indicators obtained from MLP 3-5-4 NN and relative error values in relation to FE simulation results*

| Designation | PCF [kN] | MCF [kN] | CLE [%] | SE [-] | TE [%] | $\Delta_{PCF}$ [%] | $\Delta_{MCF}$ [%] | $\Delta_{CLE}$ [%] | $\Delta_{SE}$ [%] | $\Delta_{TE}$ [%] |
|---|---|---|---|---|---|---|---|---|---|---|
| D-20-6-h20 | 39,683 | 14,525 | 37,339 | 0,5053 | 18,644 | 2,02 | 0,90 | 0,91 | 0,62 | 1,48 |
| D-20-6-h80 | 40,692 | 13,757 | 33,970 | 0,5351 | 18,099 | 0,14 | 1,14 | 0,52 | 1,09 | 0,14 |
| D-60-6-h100 | 46,598 | 13,201 | 29,660 | 0,5588 | 15,956 | 1,94 | 0,08 | 2,78 | 0,15 | 0,91 |
| D-60-8-h60 | 39,827 | 13,073 | 32,688 | 0,5632 | 18,421 | 1,04 | 0,89 | 2,31 | 4,34 | 1,99 |
| D-80-4-h60 | 47,560 | 13,391 | 28,821 | 0,5504 | 15,556 | 0,40 | 1,51 | 0,41 | 1,51 | 0,04 |
| D-80-4-h140 | 49,069 | 13,596 | 27,958 | 0,5445 | 15,022 | 0,31 | 5,71 | 6,34 | 4,83 | 0,12 |
| D-100-8-h50 | 41,025 | 12,927 | 31,535 | 0,5695 | 17,911 | 0,08 | 3,83 | 3,68 | 4,07 | 0,03 |

method. The predicted indicator values along with the percentage difference from the FE calculations are shown in Tables 10 and 11.

In terms of predicting the PCF values, the maximum differences are found in the MLP-3-4-4 network and are about 3.75%, while for the MLP-3-5-4 network they are maximum 2%. Regarding the MCF and CLE indicator, the maximum differences in both networks were obtained for out-of-range specimens and are between 5-8.5%, where again the MLP 3-5-4 network had better prediction properties. The prediction of the SE indicator, also obtains maximum differences for out-of-range specimens. These differences are between 5.5 and 6.5% and apply to MLP3-4-4 networks, while for MLP3-5-4 networks they

lateral surfaces, where some results on experimentally validated numerical analysis have already been shown [16].

## References

1.  Abbasi M, Reddy S, Ghafari-Nazari A, Fard M. Multiobjective crashworthiness optimization of multi-cornered thin-walled sheet metal members. Thin-Walled Structures 2015. doi:10.1016/j.tws.2014.12.009, https://doi.org/10.1016/j.tws.2014.12.009.
2.  Abramowicz W. Thin-walled structures as impact energy absorbers. Thin-Walled Structures, 2003. doi:10.1016/S0263-8231(02)00082-4, https://doi.org/10.1016/S0263-8231(02)00082-4.
3.  Al-Garni A, Abdelrahman W, Abdallah A. ANN-based failure modeling of classes of aircraft engine components using radial basis functions. Eksploatacja i Niezawodnosc 2019; 21(2): 311–317, https://doi.org/10.17531/ein.2019.2.16.
4.  Alavi Nia A, Parsapour M. Comparative analysis of energy absorption capacity of simple and multi-cell thin-walled tubes with triangular,

square, hexagonal and octagonal sections. Thin-Walled Structures 2014; 74: 155–165, https://doi.org/10.1016/j.tws.2013.10.005.

5. Alexander J M. An approximate analysis of the collapse of thin cylindrical shells under axial loading. Quarterly Journal of Mechanics and Applied Mathematics 1960; 13: 10–15.

6. Alghamdi A A A. Collapsible impact energy absorbers: An overview. Thin-Walled Structures 2001. doi:10.1016/S0263-8231(00)00048-3, https://doi.org/10.1016/S0263-8231(00)00048-3.

7. Ali M, Ohioma E, Kraft F, Alam K. Theoretical, numerical, and experimental study of dynamic axial crushing of thin walled pentagon and cross-shape tubes. Thin-Walled Structures 2015. doi:10.1016/j.tws.2015.04.007, https://doi.org/10.1016/j.tws.2015.04.007.

8. Badora M, Sepe M, Bielecki M et al. Predicting length of fatigue cracks by means of machine learning algorithms in the small-data regime. Eksploatacja i Niezawodnosc 2021; 23(3): 575–585, https://doi.org/10.17531/EIN.2021.3.19.

9. Baroutaji A, Sajjia M, Olabi A G. On the crashworthiness performance of thin-walled energy absorbers: Recent advances and future developments. Thin-Walled Structures 2017. doi:10.1016/j.tws.2017.05.018, https://doi.org/10.1016/j.tws.2017.05.018.

10. Baykasoğlu A, Baykasoğlu C, Cetin E. Multi-objective crashworthiness optimization of lattice structure filled thin-walled tubes. Thin-Walled Structures 2020. doi:10.1016/j.tws.2020.106630, https://doi.org/10.1016/j.tws.2020.106630.

11. Chen D H, Ozaki S. Circumferential strain concentration in axial crushing of cylindrical and square tubes with corrugated surfaces. Thin-Walled Structures 2009; 47(5): 547–554, https://doi.org/10.1016/j.tws.2008.10.003.

12. Chen W, Wierzbicki T. Relative merits of single-cell, multi-cell and foam-filled thin-walled structures in energy absorption. Thin-Walled Structures 2001. doi:10.1016/S0263-8231(01)00006-4, https://doi.org/10.1016/S0263-8231(01)00006-4.

13. Fang J, Sun G, Qiu N et al. On design optimization for structural crashworthiness and its state of the art. Structural and Multidisciplinary Optimization 2017; 55(3): 1091–1119, https://doi.org/10.1007/s00158-016-1579-y.

14. Ferdynus M, Kotełko M, Kral J. Energy absorption capability numerical analysis of thin-walled prismatic tubes with corner dents under axial impact. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20(2): 252–289, https://doi.org/10.17531/ein.2018.2.10.

15. Ferdynus M, Kotełko M, Urbaniak M. Crashworthiness performance of thin-walled prismatic tubes with corner dents under axial impact - Numerical and experimental study. Thin-Walled Structures 2019. doi:10.1016/j.tws.2019.106239, https://doi.org/10.1016/j.tws.2019.106239.

16. Ferdynus M, Rozylo P, Rogala M. Energy absorption capability of thin-walled prismatic aluminum tubes with spherical indentations. Materials 2020; 13(19): 1–19, https://doi.org/10.3390/ma13194304.

17. Hanssen A G, Langseth M, Hopperstad O S. Static and dynamic crushing of square aluminum extrusions with aluminum foam filler. International Journal of Impact Engineering 2000. doi:10.1016/S0734-743X(99)00169-4, https://doi.org/10.1016/S0734-743X(99)00169-4.

18. Jafarzadeh-aghdam N, Schröder K. Thin-Walled Structures Mechanism of reproducible axial impact of square crash boxes. Thin-Walled Structures 2022; 176(February): 109062, https://doi.org/10.1016/j.tws.2022.109062.

19. Jones N. Structural Impact. 1990. doi:10.1017/cbo9780511624285, https://doi.org/10.1017/cbo9780511624285.

20. Karagiozova D, Jones N. Dynamic buckling of elastic-plastic square tubes under axial impact - II: Structural response. International Journal of Impact Engineering 2004; 30(2): 167–192, https://doi.org/10.1016/S0734-743X(03)00062-9.

21. Kazi M K, Eljack F, Mahdi E. Design of composite rectangular tubes for optimum crashworthiness performance via experimental and ANN techniques. Composite Structures 2022; 279(July 2021): 114858, https://doi.org/10.1016/j.compstruct.2021.114858.

22. Kosucki A, Stawiński Ł, Malenta P et al. Energy consumption and energy efficiency improvement of overhead crane's mechanisms. Eksploatacja i Niezawodnosc 2020; 22(2): 322–330, https://doi.org/10.17531/ein.2020.2.15.

23. Laban O, Gowid S, Mahdi E, Musharavati F. Experimental investigation and artificial intelligence-based modeling of the residual impact damage effect on the crashworthiness of braided Carbon/Kevlar tubes. Composite Structures 2020. doi:10.1016/j.compstruct.2020.112247, https://doi.org/10.1016/j.compstruct.2020.112247.

24. Langseth M, Hopperstad O S, Berstad T. Crashworthiness of aluminum extrusions: Validation of numerical simulation, effect of mass ratio and impact velocity. International Journal of Impact Engineering 1999; 22(9): 829–854, https://doi.org/10.1016/S0734-743X(98)00070-0.

25. Lee S, Hahn C, Rhee M, Oh J E. Effect of triggering on the energy absorption capacity of axially compressed aluminum tubes. Materials and Design 1999; 20(1): 31–40, https://doi.org/10.1016/s0261-3069(98)00043-0.

26. Liu W, Lin Z, He J et al. Crushing behavior and multi-objective optimization on the crashworthiness of sandwich structure with star-shaped tube in the center. Thin-Walled Structures 2016; 108: 205–214, https://doi.org/10.1016/j.tws.2016.08.021.

27. Lu G, Yu T. Energy Absorption of Structures and Materials. 2003. doi:10.1533/9781855738584, https://doi.org/10.1533/9781855738584.

28. Luo Y, Fan H. Energy absorbing ability of rectangular self-similar multi-cell sandwich-walled tubular structures. Thin-Walled Structures 2018; 124(March 2017): 88–97, https://doi.org/10.1016/j.tws.2017.11.042.

29. Marzbanrad J, Ebrahimi M R. Multi-Objective Optimization of aluminum hollow tubes for vehicle crash energy absorption using a genetic algorithm and neural networks. Thin-Walled Structures 2011; 49(12): 1605–1615, https://doi.org/10.1016/j.tws.2011.08.009.

30. Mirzaei M, Shakeri M, Sadighi M, Akbarshahi H. Crashworthiness design for cylindrical tube using neural network and genetic algorithm. Procedia Engineering 2011; 14: 3346–3353, https://doi.org/10.1016/j.proeng.2011.07.423.

31. Nia A A, Hamedani J H. Comparative analysis of energy absorption and deformations of thin walled tubes with various section geometries. Thin-Walled Structures 2010. doi:10.1016/j.tws.2010.07.003, https://doi.org/10.1016/j.tws.2010.07.003.

32. Paygozar B, Dizaji S A. Investigating Energy Absorption Accessible by Plastic Deformation of a Seismic Damper Using Artificial Neural Network. Procedia Structural Integrity 2019; 21(January): 138–145, https://doi.org/10.1016/j.prostr.2019.12.095.

33. Pirmohammad S, Marzdashti S E. Crushing behavior of new designed multi-cell members subjected to axial and oblique quasi-static loads. Thin-Walled Structures 2016; 108: 291–304, https://doi.org/10.1016/j.tws.2016.08.023.

34. Rai V, Ghasemnejad H, Watson J W et al. Developed trigger mechanisms to improve crush force efficiency of aluminium tubes. Engineering Structures 2019. doi:10.1016/j.engstruct.2019.109620, https://doi.org/10.1016/j.engstruct.2019.109620.

35. Reddy S, Abbasi M, Fard M. Multi-cornered thin-walled sheet metal members for enhanced crashworthiness and occupant protection. Thin-Walled Structures 2015. doi:10.1016/j.tws.2015.03.029, https://doi.org/10.1016/j.tws.2015.03.029.

36. Rogala M, Gajewski J, Ferdynus M. Numerical analysis of the thin-walled structure with different trigger locations under axial load. IOP Conference Series: Materials Science and Engineering 2019. doi:10.1088/1757-899X/710/1/012028, https://doi.org/10.1088/1757-899X/710/1/012028.

37. Seyedi M R. A Study of Multi-Objective Crashworthiness Optimization of the Thin-Walled Composite Tube under Axial Load. 2020:

438–452.

38. Sharifi S, Shakeri M, Fakhari H E, Bodaghi M. Experimental investigation of bitubal circular energy absorbers under quasi-static axial load. Thin-Walled Structures 2015. doi:10.1016/j.tws.2014.12.008, https://doi.org/10.1016/j.tws.2014.12.008.

39. Winzer R, Glinicka A. The static and dynamic compressive behaviour of selected aluminium alloys. Engineering Transactions 2011; 59(2): 85–100.

40. Xu F, Zhang X, Zhang H. A review on functionally graded structures and materials for energy absorption. Engineering Structures 2018; 171(February): 309–325, https://doi.org/10.1016/j.engstruct.2018.05.094.

41. Yang X, Ma J, Wen D, Yang J. Crashworthy design and energy absorption mechanisms for helicopter structures: A systematic literature review. Progress in Aerospace Sciences 2020; 114(June): 100618, https://doi.org/10.1016/j.paerosci.2020.100618.

42. Yin H, Wen G, Liu Z, Qing Q. Crashworthiness optimization design for foam-filled multi-cell thin-walled structures. Thin-Walled Structures 2014; 75: 8–17, https://doi.org/10.1016/j.tws.2013.10.022.

43. Yuen S C K, Nurick G N. The energy-absorbing characteristics of tubular structures with geometric and material modifications: An overview. Applied Mechanics Reviews 2008; 61(1–6): 0208021–02080215, https://doi.org/10.1115/1.2885138.

44. Zarei H R, Kröger M. Optimization of the foam-filled aluminum tubes for crush box application. Thin-Walled Structures 2008; 46(2): 214–221, https://doi.org/10.1016/j.tws.2007.07.016.

45. Zhang X, Cheng G, Wang B, Zhang H. Optimum design for energy absorption of bitubal hexagonal columns with honeycomb core. International Journal of Crashworthiness 2008; 13(1): 99–107, https://doi.org/10.1080/13588260701731732.

46. Zhang X, Cheng G, You Z, Zhang H. Energy absorption of axially compressed thin-walled square tubes with patterns. Thin-Walled Structures 2007; 45(9): 737–746, https://doi.org/10.1016/j.tws.2007.06.004.

47. Zhang X, Huh H. Energy absorption of longitudinally grooved square tubes under axial compression. Thin-Walled Structures 2009. doi:10.1016/j.tws.2009.07.003, https://doi.org/10.1016/j.tws.2009.07.003.

48. Zhang X, Wen Z, Zhang H. Axial crushing and optimal design of square tubes with graded thickness. Thin-Walled Structures 2014. doi:10.1016/j.tws.2014.07.004, https://doi.org/10.1016/j.tws.2014.07.004.

49. Crashworthiness of vehicles. Londin, Mechanical Engineering Publications Limited: 1978.

50. Structural crashworthiness. London, Butterworths: 1983.

51. Structural impact and crashworthiness. New York, Elsevier Applied Science Publishers: 1984.

# INFORMATION FOR AUTHORS

## Terms and Conditions of Publication

1. The quarterly „Eksploatacja i Niezawodnosc - 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.