LATVIAN JOURNAL of PHYSICS and TECHNICAL SCIENCES

ISSN 0868 - 8257

5

(Vol. 58)

2021

CONTENTS

LATVIAN JOURNAL of PHYSICS and TECHNICAL SCIENCES

LATVIJAS FIZIKAS un TEHNISKO ZINĀTŅU ŽURNĀLS

ЛАТВИЙСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Published six times a year since February 1964
Iznāk sešas reizes gadā kopš 1964. gada februāra
Выходит шесть раз в год с февраля 1964 года

5 (Vol. 58) • 2021

RĪGA

Language Editor: O. Ivanova
Computer Designer: I. Begicevs

INDEXED (PUBLISHED) IN

www.scopus.com
www.sciendo.com
EBSCO (Academic Search Complete, www.epnet.com), INSPEC (www.iee.org.com).
VINITI (www.viniti.ru), Begell House Inc/ (EDC, www.edata-center.com).

Issuers: Institute of Physical Energetics,
Institute of Solid State Physics, University of Latvia
Registration Certificate Number: 000700221

Editorial Contacts:
11 Krivu Street, Riga, LV - 1006
Ph.: + 371 67551732
E-mail: leo@lza.lv
www.fei-web.lv

sciendo

LATVIAN JOURNAL OF PHYSICS
AND TECHNICAL SCIENCES
2021, N 5

DOI: 10.2478/lpts-2021-0035

MODELLING OF DRAG FORCE REDUCTION FOR A WATERJET PROPULSION SYSTEM

M. Cerpinska*, M. Irbe, A. Pupurs, K. Burbeckis

Laboratory of Experimental Mechanics of Materials, Riga Technical University,
6B-525 Kipsalas Str., Riga, LV-1048, LATVIA
*e-mail: marina.cerpinska@rtu.lv

The paper provides simulation results for SUP (Stand Up Paddle) board appendage resistance. Additional propulsion is added to the SUP board. It is equipped with a waterjet. The waterjet is attached to the board rudder. This increases the drag coefficient for rudder five times. To reduce the drag variable, design options for the waterjet duct were proposed. The simulation tests were performed using SolidWorks Flow software using two types of simulations, namely, the pressure on the body and the flow around the body. The objective was to streamline the bluff duct of the waterjet and thus to create the appendage design with minimum drag force from fluid flow and possibly greater Inlet Velocity Ratio. Calculations showed that rounding-off the edges of waterjet duct resulted in 35 % of drag coefficient reduction, while further streamlining reduced it by additional 10 %.

Keywords: *Drag coefficient, propulsion, resistance, SUP board, waterjet.*

1. INTRODUCTION

This study is the first part of a project aimed at designing the optimal waterjet to motorise the SUP (Stand Up Paddle) board.

The idea of motorising SUP boards for leisure activities is lucrative for two reasons. First, it provides existing users with new experience. It allows relaxing on the water and organising adventurous board trips for longer distances. Second, it attracts new users previously not interested in SUP boarding due to the slow motion of the board or the necessity to paddle actively to get propulsion.

Motorising the leisure activity brings in the dilemma, as discussed in a book by Jennings devoted to Motorised Water Sports [1].

It is necessary to consider damage to the local environment. Yet, the author stresses that motorised boards provide a lot of positive emotions, broadens tourism offerings and, as a result, provide economic benefits for local and regional community [1].

The goal of the project is to attach a waterjet to the SUP fin (rudder). There could be three fins installed on one board, but the design with one fin is more usual. This project involves two teams working together. The team of electrical engineers is responsible for the selection of motor and control system for propulsion. The team of mechanical engineers is responsible for the design of propulsion components. The first challenge for the mechanical engineers is to evaluate the hydrodynamic resistance of the attached object and to choose a design with least resistance because unmodified rudder is a very slender object with low frontal resistance. On the other hand, resistance should be evaluated together with the propulsion capacity.

The handbook on practical ship hydrodynamics states that propulsion and resistance are complex problems, and to solve them a CFD analysis is preferred [2]. In ship hydrodynamics theory, rudder is one of the appendages. The resistance of the appendages under service conditions is tested separately or not tested at all. If the appendage size is small compared to the ship, one would expect an error in resistance calculation to be small. The chapter "Resistance and Propulsion" of the book [2] states that the significance of the resistance created by appendages depends on their size compared to the ship size. Thus, the more complex appendages are installed on the floating object, the more important it is to evaluate their resistance. For the CFD study on submarines, Shariati et al. showed that the total resistance of a submarine increased by fitting appendages on it [3]. The authors used Star CCM+ software, developed using a finite volume method. In that study, the major influence of appendage resistance was detected on the viscous part of the total resistance, while wave making resistance was not affected significantly [3].

The SUP board is a small craft compared to ships. As noted by Carlton [4], for small craft propellers are usually considered a simpler and lighter propulsor. However, in this project a waterjet propulsor was chosen as a safer option compared to an open shaft propeller. Since a SUP board for tourism is mainly used on the rivers, a propulsor needs to be protected from the water plants. Sometimes the bottom of the board is floating very close to the underwater stones. Thus, the case or duct of the waterjet is supposed to be the barrier, which saves propulsion from potential damage. Propeller could still be used as a SUP board propulsor; however, a duct is needed to protect it. In fact, Cooper et al. [5] suggest that a waterjet could be considered a special type of ducted propeller. Majority of test procedures for waterjet propulsion and propellers are currently the same [6].

The typical SUP board weights around 13–14 kg. Depending on the style and size, there could be lighter versions starting from 7 kg and some as heavy as 18 kg. The added mass of a propulsion system should be small enough to ensure that buoyancy is not disrupted.

The waterjet is attached to the SUP fin (rudder) in the following way: the fin is cut into two parts, and the waterjet is in between – far enough from the SUP board to provide space for the charging battery. The waterjet case is a barrier for water plants, fish and stones. As a negative aspect, it is a barrier to the water as well. As noted in the review paper [7], efficient inlets for the waterjet are crucial for the successful performance of the system. The task is, therefore, to choose

the design option balancing between the ease of manufacturing and efficiency.

In this study, we investigated the drag force exerted on the appendage using SolidWorks Flow Simulation software. Two types of simulations (type 1 – the pressure on the body; type 2 – the flow around the body) were tested and compared for accuracy and computational time. Albeit computer simulations are time consuming, the importance of CFD modelling specifically for waterjets is stressed in the Procedures of the Specialist Committee on Validation of Waterjet Test [7]. The Specialist Committee published a review on updated literature about Intake Design and Analysis [7]. The authors in [7] quoted the analysed waterjet inlets for large watercrafts, yet their research stressed the importance of CFD simulations and provided guidance on proper boundary conditions to be applied to a model.

Bulten in his PhD thesis [9], quoted in reference book [4], compared numerical methods used by waterjet industry leaders Wärtsilä Propulsion Netherlands BV and he found that the simplified method to describe waterjet installations was not correct because existing computer programs did not calculate the influence of the hull in the vicinity of the waterjet inlet and the pressure distribution acting on the stream tube [9].

The hypothesis of this study stated that optimising the shape of the waterjet would improve the efficiency of the proposed propulsion solution. The hypothesis was based on the theory provided by Hoerner [10]. Hoerner showed that rounding-off edges in blunt shapes was not perfect (compared to a complete streamlining procedure) but still efficient method to reduce a drag coefficient. Results section showed that this hypothesis was true. Rounding-off edges provided 35 % of drag coefficient reduction, and additional reduction of 10 % was obtained after applying a streamlining solution.

Waterjet in operation induces reaction forces and moment and waterjet/hull interaction stern force [11]. These forces are out of scope of this research paper, but the velocity effect will be discussed in this paper.

2. THEORY

Forces Acting on the Body

According to the goal of this study, the hydrodynamic resistance of the modified rudder should be minimised to ensure propulsion efficiency. In this study, we minimise it by reducing the drag force.

Following the equation from fluid mechanics fundamentals on the external flow [12], the drag force acting on a body is given by Eq. (1):

$$F_D = \frac{1}{2}\rho V^2 C_D A, \quad (1)$$

where F_D – drag force, N or $kg \cdot m/s^2$; ρ – fluid density, kg/m^3; V – relative velocity, m/s; C_D – a drag coefficient, dimensionless; A – a frontal area of the body, m^2. The term $\frac{1}{2}\rho V^2$ is called dynamic pressure [12].

Figure 1 demonstrates the forces acting on the SUP board and a person standing on it. The SUP board, a person standing on it and a paddle a person holds are moving in two media – the water and the air, respectively.

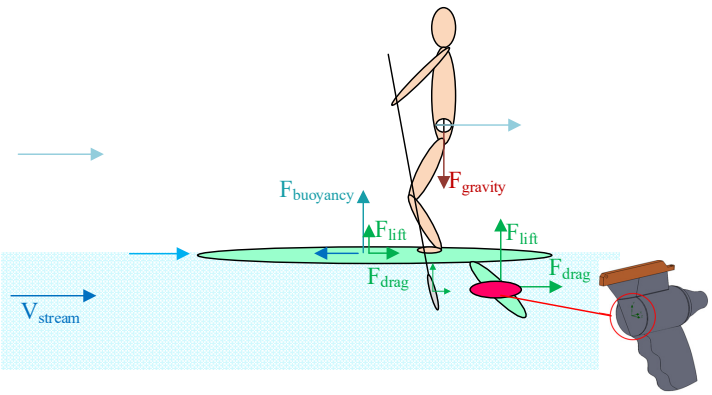


Fig. 1. The forces acting on the SUP board (green colour), the person driving the board (skin colour) and SUP board appendage, including the waterjet (red colour) attached.

According to Eq. (1), the drag force depends on the frontal area of the body. Figure 1 suggests that the area of the modified rudder (red coloured area) would be small compared to the area of the person (skin coloured area). On the other hand, water density is more than 800 times greater than the air density [12]. Therefore, the appendage under water increases the total drag force of a SUP board more than the appendage above the water would do. It is also desirable to minimixe the drag force for the appendage because it would lessen the stresses on the attachment of the SUP fin to the board. We acknowledge that manufacturers of SUP boards do not design the attachment of the fin to experience great frontal drag force, and therefore knowledge about the drag force experienced by appendage is of special importance.

In the water medium, the hydrostatic pressure and resulting buoyancy force are dominant. The hydrodynamic drag and lift forces are acting on a board, but for such a slender body as a SUP board the shear force acts nearly parallel to the flow direction, and thus their contribution to the lift is small [12]. The drag force for such a slender body is mostly due to the skin friction drag, which is tangential force [10] and is also relatively small. Therefore, for the streamlining purposes we will analyse the hydrodynamic pressure locally on the fin, where the waterjet is attached (see the red circle on Fig. 1). Results show that adding a waterjet duct to the fin increases fivefold the drag force a flowing fluid exerts on a body in the flow direction.

The part of resistance force that is related to pressure is called the pressure drag or form drag because of its strong dependence on the form or shape of the body [12]. The pressure field of the SUP board causes deformation of water surface, and a wave system originates behind the board. The wave making resistance force is acting normal to the body as shown in Fig. 1. Since we focus on the drag force of the immersed part, wave making resistance is out of scope in calculations.

Waterjet Design

According to the handbook on propulsion systems, the waterjet design could be simplified to three main components: an inlet ducting, a pump and an outlet or nozzle [4].

The physics of waterjet propulsion is described in reference book [13]. It states that water is drawn through a ducting system by an internal pump. The pump adds

energy to water and it is expelled from the ducting system at high velocity. The unit thrust is primarily generated as a result of the momentum increase imparted to water [13]. The mathematical expression for the description above is provided by Carlton [4] as shown in Eq. (2):

$$T = \rho A_2 V_2 (V_2 - V_1), \quad (2)$$

where T – propulsor thrust produced by the system, N; ρ – density of the water, kg/m^3; A_2 – a cross-sectional area of the nozzle, m^2; V_2 – velocity of the water leaving the system, m/s; V_1 – velocity of the water entering the system, m/s.

No matter how efficient the internal pump is, there are going to be some performance losses caused by flow distortion on the intake. To minimise inlet losses, this paper focuses on the inlet design. The inlet ducting consists of tunnel, inlet and supporting structures [4]. Equation (2) suggests that achieving more thrust requires greater difference between the water inlet and outlet speed. Thus, outlet nozzle should be designed to produce a greater speed of the water leaving the system, while the inlet design, on the contrary, should ensure the water entering speed to be possibly close to the floating object speed. To describe the inlet efficiency, the term "inlet velocity ratio" is commonly used.

Inlet Velocity Ratio (IVR)

The Specialist Committee on Validation of Waterjet Test Procedures [7] suggests using abbreviation IVR (Inlet Velocity Ratio) when describing intake working point.

The IVR is calculated according to Eq. (3) [4]:

$$\text{IVR} = \frac{V_1}{V_s}, \quad (3)$$

where V_1 – the water inlet velocity, m/s; V_s – the craft velocity, m/s.

Following Eq. (3), as the water inlet velocity decreases due to some constructive barrier, the IVR decreases and the flow rate through the waterjet decreases.

Boundary Layers

A boundary layer is a thin layer of viscous fluid close to the solid surface of a wall in contact with a moving stream. The flow velocity at the wall is zero [10], since the flow "sticks" to the wall because of its viscosity.

According to Ledoux et al. [14], jets form "a special group of mixing boundary layers" because there are an outer boundary layer and an inner boundary layer, and therefore Ledoux et al. [14] suggest using a numerical approach for jet problems.

The behaviour of differently shaped objects in the flow was presented by Vutukuru et al. [15]. The authors used ANSYS software to study the interaction forces for the five different forms of the prism, and "boundary layers set to stair stepping with maximum of 7 layers" were used for each shape. The results of the study suggested that the pressure in the suction zone along the boundary layer was independent of the shape of the object with sufficient accuracy (about 2 %), and one could assume "a constant pressure in the boundary layer of the flow suction zone at low speeds" [15]. Considering the results above, in this study we focus on the resistance independent of the boundary layer.

3. NUMERICAL MODELLING AND SIMULATION

One of the goals of the project is to develop a full-scale prototype of the waterjet using available 3-dimensional (3D) printing techniques and to test it in the water. Meanwhile, before the optimal design parameters are known, the tests were performed using SolidWorks Flow simulation software. The 3D simulation, performed in SolidWorks Flow Simulation, discretizes the time-dependent Navier-Stokes equations and solves them on the computational mesh. More details are provided in the section "Mesh Size".

The simulation to obtain a drag coefficient for the rudder with the waterjet attached was performed twice; namely, the simulations of two types were completed. The first type (further referred to as Type 1) was the simulation of the pressure on the body. This simulation focused on the pressure zone on the appendage. The pressure is simulated as a vector acting on the frontal area. The second type of the simulation (further referred to as Type 2) was simulation of the flow around the body, and it focused on the pressure distribution on the whole surface of the object (the uniform body). The convergence goal (a variable to calculate from the fluid flow) was the force (see Eq. (1)). The initial condition of fluid flow inlet velocity was specified to be 2.5 m/s. It was expected that the simulation of Type 2 would provide higher precision. However, we found out that for the specific study of a drag coefficient, results of both simulations were very close, while simulation of Type 1 provided results faster, and the results were easier to analyse. Detailed comparison is provided in the Results section.

Rounding-off Edges and Streamlining

The waterjet cover and rudder are optimised in SolidWorks software to reduce resistance force from the fluid flow. The part subjected to analysis was an intake chamber – motor cover as shown in Fig. 2.

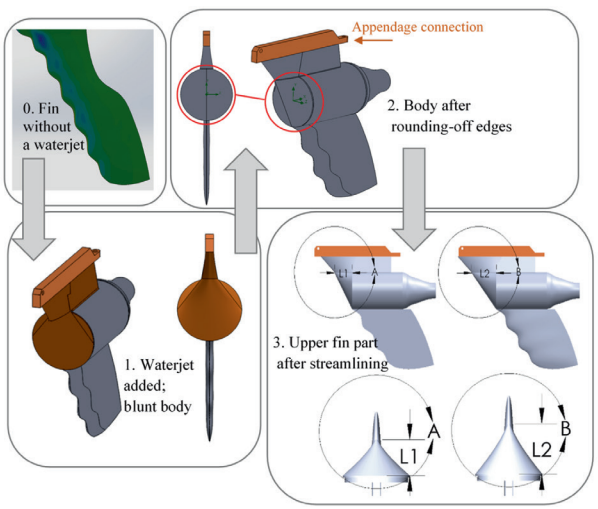


Fig. 2. The fin and waterjet view before modifications and after rounding-off edges and streamlining in SolidWorks.

Figure 2 shows the sequence of the modifications made to the fin and the optimised design of the waterjet. The area marked by red circle shown on the appendage is further streamlined – for the chosen optimal design of the fin, the drag coefficient for different fin lengths (from L1 to L2) is studied. The difference in fin length is demonstrated in Fig. 2 as circle A and circle B. In circle A, the fin is shorter than appendage connection (L1), in circle B the length of the fin is prolonged to the maximum value that appendage connection permitted (L2 = L1 + 30 mm). For the streamlining study, the simulation of flow pressure was performed with a step (difference in fin length) of 5 mm, starting with ~0 mm (L1) and up to 30 mm (L2). The results of the drag coefficient dependency on rudder length are shown in Fig. 7 (simulation of Type 1).

Mesh Size

For meshing the model, SolidWorks Flow Simulation uses a technique called the Finite Volume Approach. For the simulation of fin without a waterjet, the generated mesh consisted of $2.76 \cdot 10^5$ cells, for a simulation with a waterjet – it contained more than $7 \cdot 10^5$ cells. The local mesh is intended for resolving the mesh around the highest pressure in the frontal cylindrical region. An example of the mesh is shown in Fig. 3.

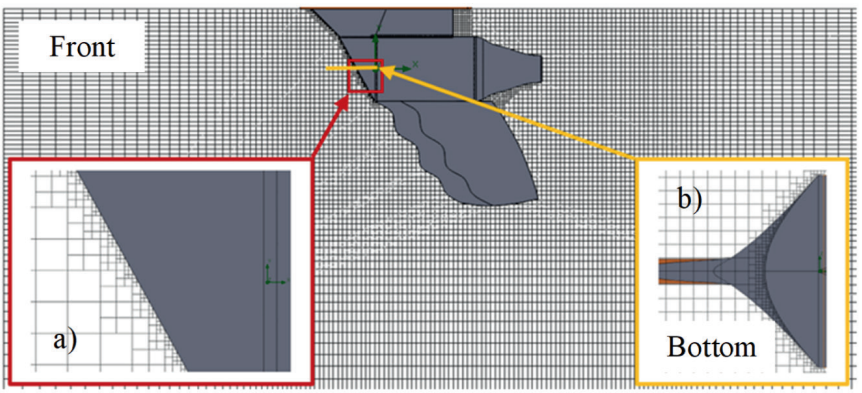


Fig. 3. The illustration of the mesh size.

Figure 3 provides the example of the front cut plot of initial global mesh domain with refinement level four (4), zoomed view for local mesh at cylindrical part in front (red frame–a)) and bottom (yellow frame–b)) planes.

4. RESULTS AND DISCUSSION

The three-dimensional finite element simulation results are given in this section to illustrate calculation of drag coefficient. Figure 4 shows the selected results from the simulation of Type 1 (the pressure on the body).

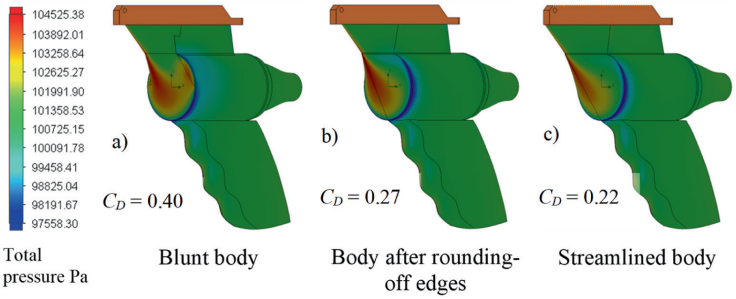


Fig. 4. The pressure on the front surface and resulting drag coefficient for original blunt body, body after round-off edges and streamlined body.

The example of results for the simulation of Type 2 (the flow around the body) is given in Fig. 5. The complete animation for the flow around the body is available contacting the corresponding author.

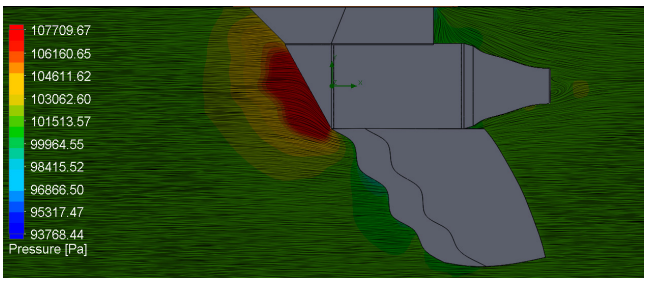


Fig. 5. The example of the simulation of the flow around the body. Frontal pressure distribution and streamlines.

Drag Coefficient and Resulting Force

Simulation of the pressure on the body showed that the drag coefficient for the fin without a waterjet attached was only 0.078. Drag coefficient increased to 0.4 as the waterjet was added.

The results obtained from simulations for the fin with a waterjet attached are summarised in Fig. 6, where the design of blunt body (Fig. 4a) is indicated with orange markers, and the improved design of streamlined body is indicated with blue markers. The second-order polynomial approximation is used for the streamlined body results.

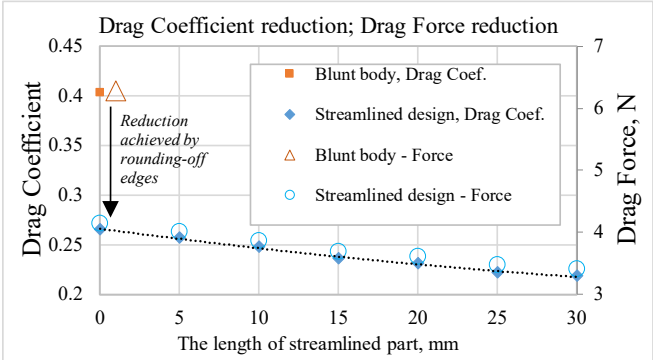


Fig. 6. Drag coefficient and drag force reduction after rounding-off the edges of the waterjet duct.

Figure 6 shows that rounding-off the edges of waterjet duct reduces the drag coefficient from 0.4 to 0.27–0.22 depending on the length of the appendage. It means that the reduction of drag from 35 % to 45 % is achieved, so rounding-off edges is strongly suggested during manufacturing. Meanwhile, further streamlining of design results in 10 % reduction of drag coefficient (see Fig. 2 for illustration of difference between the appendage length from L1 to L2 = 30 mm), so it is recommended to apply it, if manufacturing capacity permits. Figure 6 shows a summary of the average drag coefficient, obtained from the simulation of 200 iterations in SolidWorks software. Please refer to Fig. 7 for a complete illustration of drag coefficient simulation.

According to Fig. 6, the average drag coefficient and drag force decrease as the fin length increases, but for the fin length, which equals 30 mm (this limit is set by the size of the attachment), the gained decrease is almost equal to design of 25 mm, so the further streamlining attempts would be less efficient.

Comparing the Drag Coefficient Results from Two Simulation Methods

In this study, two types of simulations were performed to predict the drag coefficient. The first type of the simulation was for the pressure on the body. The second type of the simulation was for the flow around the body. The results slightly differed for both methods because for the simulation of Type 2 on the flow around the body the mesh size was smaller. Figure 7 illustrates the difference of simulation results for the drag coefficient change during the simulation of the pressure and simulation of the flow.

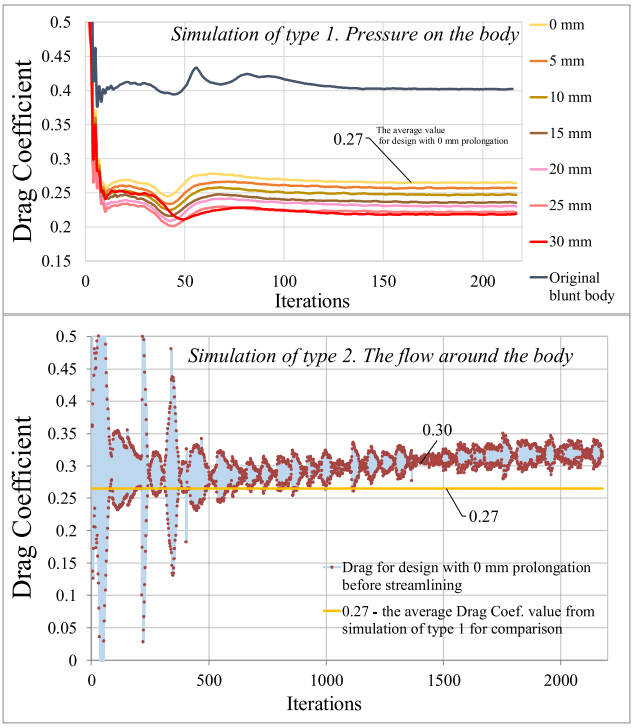


Fig. 7. Drag coefficient values from simulation results for each iteration in SolidWorks – comparison of the drag coefficient values calculated from pressure and from the flow around the body.

To compare measurement results from the two simulations, zeta-scores test [16] for the design of ~0 mm prolongation (no streamlining) was performed. For the simulation of Type 1 (the pressure on the body), the mean value (from all iterations of the simulation, including instability region) of the drag coefficient was 0.27, while for the simulation of Type 2 (the flow around the body) the mean value was 0.3. Uncertainty at standard deviation level, , was calculated to be 0.08 and 0.11 for simulation of Type 1 and Type 2, respectively. The absolute value of the zeta was calculated to be 0.22. It is below 2, which means that the results are not in disagreement. From this analysis we concluded that the simulation of Type 1 should be preferably used to explore different design options because it required less time to complete. For comparison, the simulation of Type 2 takes around 17 hours on a standard laptop computer, while simulation of Type 1 could be completed in a couple of hours. Thus, for the study of the drag coefficient the simulation of pressure is explicit enough. The simulation of Type 2 on the flow around the body should be used when more information about the changes of the drag coefficient in time is needed.

The data in Fig. 7 suggest that during the simulation of the flow around the body the following phenomena occur: as the flow first meets the rudder, the drag coefficient is larger, then it drops to its lowest value, but then it remains stable.

The simulation of the flow around the body could later be used for further studies of vortexes around the body. The flow simulation is also helpful to describe the cavitation effects that may occur in the pump or in the intake during operation of the vessel. Cavitation does not affect the powering characteristics, but it causes possible erosion effects [7].

Inlet Location

After the study of uniform body is completed, the analysis of the best inlet position could be discussed. According to Eq. (3), the inlet should be located to allow for maximum water inlet velocity. The design that is easy to manufacture requires water entering the jet through the grooves cut on the side of the tunnel. This design has a disadvantage – the flow from the sides is small compared to the flow from the front plane. The dark blue colour on the sides of the tunnel in Fig. 4 indicate that the pressure on the sides is the smallest; therefore, the amount of water entering the inlet from the side will be much smaller than the amount of water entering the inlet from the front.

Therefore, we propose creating the waterjet inlet at the front of the tunnel, and in order to protect the various internal waterjet components, to fit it with a guarding mesh, keeping in mind the following principle – the finer the guarding mesh of the inlet, the better protection.

The Velocity Effect on the Drag Coefficient

For the purpose of simulation, the velocity was estimated to be 2.5 m/s or 9 km/h, which resulted from the estimation of SUP board moving speed. According to the external flow theory, the drag coefficient depends on the Reynolds number, specifically below 10^4. At higher Reynolds numbers, the drag coefficients is expected to remain constant [12]. In this study, Reynolds number was calculated to be Re ~475000; therefore, the obtained drag coeffcient would be similar for moving speeds that differed from the assumed speed. When the waterjet is in operation, the speed of the

board increases, but given the Reynolds number calculated the drag coefficient will not differ significantly.

5. CONCLUSIONS

1. Simulation results showed that waterjet added to the rudder accounted for significant frontal resistance. The drag coefficient of the rudder without a waterjet was only 0.078. It increased to 0.4 as the waterjet was added. Drag coefficient was reduced to 0.23 by choosing the optimal design of the waterjet duct.
2. Rounding-off the edges of waterjet duct accounted for 35 % of drag coefficient reduction compared to the initial blunt design, so this improvement was worth applying the necessary manufacturing effort. The further streamlining accounted for additional 10 % decrease in drag coefficient. In total, 45 % of drag coefficient could be obtained by working on the design of the waterjet frontal plane.
3. The study showed that predicting the drag coefficient for the appendage in water could be done via two types of the simulation in SolidWorks Flow. The first type of simulation was the pressure on the body. The second type of simulation was the flow around the body. Both simulations provided clear results and the trends were in a good agreement. The results slightly differed because for the simulation of Type 2 the mesh size was smaller. Simulation of Type 1 required less computational time and was therefore preferable for the study of the drag coefficient. Simulation of Type 2 is preferred when the drag change in time and vortices around the body should be studied.

ACKNOWLEDGEMENTS

The research has been conducted within the project "Development of Technical Solutions for Water Transport Transmission System", grant number ZI-2020/3, Framework of Science and Innovation within research platforms at Riga Technical University, funded by Riga Technical University.

REFERENCES

1. Jennings, G. (2007). *Motorized water sports, water-based tourism, sport, leisure, and recreation experiences*. USA: Elsevier.
2. Bertram, V. (2000). *Practical ship hydrodynamics*. UK: Butterworth-Heinemann.
3. Shariati, S. K., & Mousavizadegan, S. H. (2017). The Effect of Appendages on the Hydrodynamic Characteristics of an Underwater Vehicle near the Free Surface. *Applied Ocean Research, 67*, 31–43. DOI: 10.1016/j.apor.2017.07.001
4. Carlton, J. (2018). *Marine propellers and propulsion*. USA: Butterworth-Heinemann. DOI: 10.1016/B978-0-08-100366-4.00016-X
5. Cooper, R. D., & Doroff, S. W. (1971). *Unsteady propeller forces, fundamental hydrodynamics and unconventional propulsion*. Rome, Italy: Office of Naval Research.

6. Gong, J., Guo, C. Y., Wang, C., Wu, T. C., & Song, K. W. (2019). Analysis of Waterjet-Hull Interaction and its Impact on the Propulsion Performance of a Four-Waterjet-Propelled Ship. *Ocean Engineering*, *180*, 211–222. DOI: 10.1016/j.oceaneng.2019.04.002
7. Brandau, J. H. (1968). Performance of Waterjet Propulsion Systems – A Review of the State-of-the-Art. *Journal of Hydronautics*, *2* (2), 61–73. DOI: 10.1109/ULTSYM.2013.0335
8. Specialist Committee on Validation of Waterjet Test Procedures. (2005). Final Report and Recommendations to the 23rd ITTC. In *Proceedings of 24th International Towing Tank Conference*, (pp. 387–415), 26 March 2002, Edinburgh, Scotland: The University of Newcastle.
9. Bulten, N. W. H. (2006). *Numerical analysis of a waterjet propulsion system*. Eindhoven, the Netherlands: Technische Universiteit Eindhoven. DOI: 10.6100/IR614907
10. Hoerner, S. F. (1965). *Fluid dynamic drag: Practical information on aerodynamic drag and hydrodynamic resistance.* Washington, D.C., USA: Hoerner Fluid Dynamics.
11. Kandasamy, M., Ooi, S. K., Carrica, P., & Stern, F. (2010). Integral Force/Moment Waterjet Model for CFD Simulations. *Journal of fluids engineering*, *132* (10), 101103–101112. DOI: 10.1115/1.4002573
12. Cenqel, Y. A., & Cimbala, J. M. (2017). *Fluid mechanics: Fundamentals and applications* (4th ed). New York, USA: McGraw-Hill Education.
13. Molland, A. F. (2011). *The maritime engineering reference book: A guide to ship design, construction and operation*. Hungary: Elsevier.
14. Ledoux, M., & Hami, A. E. (2017). *Compressible flow propulsion and digital approaches in fluid mechanics*. UK: Wiley-ISTE.
15. Vutukuru, S. K., Tipans, I., Viba, J., & Irbe, M. (2020). Form optimization and interaction analysis of plane symmetry prism in air. In the 19th International Scientific Conference "Engineering for Rural Development" (pp. 739–746), 20–22 May 2020, Jelgava, Latvia: Latvia University of Life Sciences and Technologies. DOI: 10.22616/ERDev2020.19.TF170
16. University of Tartu. (2013). *Lecture 12. Comparing Measurement Results Using Measurement Uncertainty Estimates*. Available at https://sisu.ut.ee/measurement/12-using-measurement-uncertainty-estimates-decision-making

sciendo

LATVIAN JOURNAL OF PHYSICS
AND TECHNICAL SCIENCES
2021, N 5

DOI: 10.2478/lpts-2021-0036

METAL OXIDE NANOSTRUCTURE-BASED GAS SENSOR FOR CARBON DIOXIDE DETECTION

V. Gerbreders, M. Krasovska, I. Mihailova*, J. Kostjukevics,
E. Sledevskis, A. Ogurcovs, A. Gerbreders, A. Bulanovs

Institute of Life Sciences and Technology, Daugavpils University,
1 Parades Str., Daugavpils, LV-5401, LATVIA
*e-mail: irena.mihailova@du.lv

To increase the sensitivity and efficiency of a gas sensor, nanostructured ZnO and Co_3O_4 layers were obtained by hydrothermal synthesis directly on the electrode surface, eliminating the use of binders. Scanning electron microscope images showed that the resulting nanostructured coatings were characterised by good adhesion to the surface and high porosity, which opened up the possibility of their further use in the process of developing a gas sensor. The efficiency of the obtained nanostructured coatings and their sensitivity at room temperature to various concentrations of CO_2 were determined. The resistance curves of the samples were obtained as a function of gas concentration in the chamber, for Co_3O_4 and ZnO nanostructures.

Keywords: *Cobalt oxide, gas sensors, metal oxides, nanowires.*

1. INTRODUCTION

Increasing demands for environmental protection and requirements for more efficient production management have promoted the development of various types of gas sensors [1], the most popular of which are chemical sensors [2]–[4]. The most important function of these sensors for human health is monitoring the quality of atmospheric air, and measuring air composition changes [5]–[8]. Due to spatial constraints of the measurements, portable gadgets can quickly and most accurately define air composition and measure its change [9], [10]. For this purpose, gadgets based on metal oxides are widely used [11]–[15]. These sensors are resistive, wherein the electrical conductivity of the polycrystalline semiconductor film changes over time, according to gas concentration. Depending on the type of gas and the type of sensor

conductivity (p-type or n-type), the resistance of the sensitive layer changes [16]–[23].

In some cases, for the development of this type of sensor, it is necessary to use additional heaters, as the process of chemisorption on a semiconductor surface can be dependent on temperature [24]–[27].

The signal of a resistive gas sensor is the integral magnitude, which corresponds to the content of different gases in the atmosphere, either oxidants or reductants. The process of interaction of sensing materials with gases includes: adsorption; surface chemical reactions; mutual diffusion of the components into the crystallites of the semiconductor; and desorption of the reaction products [28]. Chemisorption and reaction on semiconducting metal oxide surface lead to a change in concentration of charge carriers on the surface layer of the semiconductor. The transformation of the chemical signal (charge carrier concentration change on the surface) into integral conductivity of the sensing material is provided by transport of charge carriers [29]–[31]. The microstructure of the sensitive layer (particle size, existence and size of intracrystalline bridges, diameter and value of the pores, unit surface area) affects both processes on the surface and transport of charge carriers. It is expected that semiconducting nanostructured metal oxides show highly sensitive electrophysical properties and reactivity. This occurs due to much more developed surface area and larger number of active bonds of nanostructured coatings compared to smooth surfaces [32].

Recently, screen printing has been one of the most common methods of preparation of nanostructured electrodes [33]–[37]. This method involves the nanostructure preparation as powder, then mixing the powder with a binder that provides adhesion to the surface. The obtained mixture is coated on the electrode surfaces via a specific sieve. This method is simple and allows for the production of electrodes with different sizes and shapes; however, it has some disadvantages, such as the necessity to use polymers as binders. The polymer structures prevent the formation of highly porous films and limit the diffusion of the detected gas in the sensing layer.

Limitation of diffusion substantially decreases the working area of the electrode, and consequently leads to a reduction in sensitivity. This problem may be resolved by the use of the hydrothermal synthesis method, wherein the nanostructures grow directly on the surface of the electrode [38]. The growth of the nanostructures occurs via physical and chemical processes in aqueous solutions, at temperatures lower than 100 °C. Due to the relatively low temperature of the process, there are many options for substrate materials: any chemically inactive surface can be coated by the nanostructures, including textiles and plastics. The hydrothermal method surpasses other methods in terms of diversity of the obtained morphologies: shape, size and nanostructure orientation can be changed by variation of the growing conditions, such as concentration and compound of the reagents, pH of the reaction media, and temperature and time of growth [39], [40].

Cobalt oxide (Co_3O_4) is a p-type semiconductor. Generally, this type of conductivity in metal oxides exists due to excess oxygen in the crystal structure (oxygen atoms can substitute some metal atoms, or may be situated in internodes). Thus, there are inner accepting states which form positive charge carriers (holes) after capturing electrons. In an air environment, the surface of the semiconductor can adsorb additional particles of oxygen (localised external acceptor states), and this process increases the concentration of charge carriers (holes) near the sur-

face, i.e., it forms a layer where the holes are accumulated. Gaseous oxidants, such as O_2, O_3 or NO_2, enhance this effect, whereas gaseous reductants, such as CO, H_2 or CH_4, affect the system by decreasing the amount of negative charge carriers, leading to a decrease in conductivity [41], [42].

It is known that the chemisorption and desorption of oxygen significantly affect the conductivity of sensing nanomaterials. In air, there is a large concentration of O_2 molecules, which adsorb on the metal oxide (n-type) semiconductor surface. The O_2 molecules capture electrons from the conduction band and become O^{2-} ions. As a result of this process, an electric field develops near the semiconductor surface. This field reduces the number of free electrons, forming a near-surface impoverished layer. In the case of a reduction in O_2 (e.g., air pumping out of the system), the concentration gradient rises near the n-type semiconductor surface. In this case, the concentration of chemisorbed oxygen on the semiconductor surface layer becomes higher than in the outside atmosphere, and the O_2 molecules would desorb from the semiconductor surface layer into the outside atmosphere. The captured electrons would, thus, be released from the O^{2-} ions and return to the conduction band, increasing the density of charge carriers. Under a continuous decrease in pressure, more O_2 molecules will be released, decreasing the thickness of the impoverished surface layer. As a result, the resistance of the sensor element will decrease, and conductivity will grow, accordingly [42], [43]. The opposite effect would be observed if a p-type semiconductor were used as the sensor. During the pumping out of air, the released electrons would recombine with holes, decreasing the density of charge carriers. As the pressure decreases, the conductivity of the sample increases, accordingly.

Resistive sensors with p-type conductivity can have much more favourable characteristics, particularly degree of sensitivity. For n-type semiconductors, the surface and volume components of the conductivity are defined by the same charge carriers (electrons). In the case of p-type semiconductors, the surface and volume components of the conductivity are defined by different type of charge carriers (electrons and holes), and, as a result of their recombination, the change in conductivity is more significant [44]. The interaction between CO_2 and nanostructured surfaces allegedly occurs by a different mechanism than for O_2 [30]. A CO_2 molecule has more stable atomic structure, and does not have a free electron pair for bonding. At low temperatures, water molecules react with CO_2 to form carbonate (CO_3^{2-}) ions on the metal oxide surface via several intermediate products, including bicarbonate ions (Eqs. (1)–(4)) [45], [46]. At high temperatures, CO_2 molecules react with surface O^{2-} ions directly to form CO_3^{2-} ions without intermediate products (Eq. (5)) [47]. Thus, the consumption of electrons by CO_2 molecules during interaction with metal oxide surfaces leads to a decrease in film conductivity, and can be used for CO_2 detection.

$$CO_{2(gas)} + OH^-_{(adsorbed)} \leftrightarrow HCO^-_{3(surface)}. \quad (1)$$

$$CO_{2(gas)} + H^+_{(adsorbed)} + 2e^-_{(surface)} \rightarrow HCO^-_{2(surface)}. \quad (2)$$

$$HCO^-_{2(surface)} \rightarrow CO\uparrow + OH^-_{(surface)}. \quad (3)$$

$$HCO_{3(surface)}^{-} + OH^{-}{}_{(adsorbed)} \leftrightarrow CO_{3}^{2-}{}_{(bulk)} + H_2O_{(gas)}. \quad (4)$$

$$CO_{2(gas)} + O^{2-}{}_{(adsorbed)} \rightarrow CO_{3}^{2-}. \quad (5)$$

The aim of this study was to obtain porous and nanostructured films of cobalt and zinc oxides via hydrothermal methods, and to examine the film efficiencies as working electrodes in CO_2 gas sensors, at room temperature.

2. METHODS AND MATERIALS

Pectinate-type electrodes were used as a base. The electrodes were deposited on glass surfaces via magnetron spattering though a thin chrome mask (Kurt J. Lesker). The electrode dimensions were 1.0 x 1.5 cm.

For increasing nanostructured film adhesion and ordering in the coating, the chrome electrodes were first coated with a seed layer. For the Co_3O_4 seed layer formation, a 0.005 M solution of $(CH_3COO)_2Co{\bullet}4H_2O$ (98 % pure, purchased from Sigma-Aldrich) in ethanol (99.5 % pure, purchased from Chempur) was used as the precursor, and coated onto the electrode by a dipping method. Excess solution was removed by flowing nitrogen stream. Afterwards, the samples were placed in an oven at 350 ºC for 20 min, where thermal decomposition of the $(CH_3COO)_2Co{\bullet}4H_2O$ layer to form Co_3O_4 occurred. The seed layer of ZnO was prepared using the same method. A 0.005 M solution of $Zn(CH_3COO)_2{\bullet}2H_2O$ (98 % pure, purchased from Sigma-Aldrich) in ethanol was used as the precursor. Annealing was carried out at 250 ºC for 20 min.

For the growth of nanostructures, the samples were placed in a stainless steel holder, with the seed layer downside. This positioning was used in order to protect the working surfaces from interaction with collateral nanoparticles, which fall as precipitates during the chemical reaction. The holder with the samples was deposited in heat-resistant glass with a cover, and filled with the working solution. The growth process was carried out in an oven under the following conditions:

- For Co_3O_4 nanostructures, an equimolar aqueous solution 0.1 M in $Co(NO_3)_2{\bullet}6H_2O$ (98 % pure, purchased from Sigma-Aldrich) and 0.1 M in carbamide (CH_4N_2O) (98 % pure, purchased from Sigma-Aldrich) was used. The oven temperature was 95 ºC, and exposure time was 5 h. Afterwards, the samples were washed several times with distilled water, and dried under a flowing nitrogen stream. The samples were then annealed at 450 ºC for 1 h. The annealing process could be verified visually by the colour change of the films, from pink to black.
- For ZnO nanostructures, an equimolar aqueous solution 0.1 M in $Zn(NO_3)_2$ (98 % pure, purchased from Sigma-Aldrich) and 0.1 M in hexamethylenetetramine $(CH_2)_6N_4$ (98 % pure, purchased from Sigma-Aldrich) was used. The oven temperature was 90 ºC and exposure time was 3 h. Afterwards, the samples were washed several times with distilled water, and dried under a flowing nitrogen stream. As a result, white matt films were observed.

Smooth thin films of cobalt oxides were deposited on glass substrate via magnetron sputtering (Kurt J. Lesker). As a target,

99.9 % purity Co was used. The process was conducted in oxygen-argon atmosphere (20 % O_2/80 % Ar), using 300 W discharge power and a sputtering time of 15 min. As a result, the thickness of the films was 500 nm.

The surface morphology of the samples was studied using a scanning electron microscope (TESCAN-Maya3). The chemical composition of the samples was determined using an INCA x-act energy dispersive spectrometer (Oxford Instruments). To determine the sample structure and phase composition, X-Ray Diffraction (XRD) spectra were recorded using a SmartLab Cu Kα (λ = 1.543Å) diffractometer (RIGAKU) with parallel beam geometry, using an additional Ge(220)×2 monochromator.

Measurements of the dependence of electric resistance of the Co_3O_4 film samples on gas mixture composition (and pressure) were conducted in a specialised chamber (Kurt J. Lesker), at room temperature (21 °C) and relative humidity of 26 %. During the measurement, air was pumped out (pressure reduction from atmospheric [760 Torr] to 0.1 Torr), and then atmospheric air was pumped back into the chamber. The same experiments were performed using other gases, namely CO_2 and O_2 (99 % pure, purchased from Linde). During the pumping processes, measurements of pressure and electrical resistance of all nanostructured films were conducted.

In the following experiments, the dependence of the sample resistance on the CO_2 concentration in the gas mixture was determined. Air from the chamber was pumped out down to 0.1 Torr, and then atmosphere air was pumped into a calculated pressure P_{air}. The pressure P_{air} was lower than atmospheric, and the pressure difference was compensated for by pumping CO_2 into the chamber until pressure reached 760 Torr. Experiments were conducted for air-CO_2 mixtures using CO_2 mass fractions ranging from 0.001 to 10,000, corresponding to molar (volume) concentrations of CO_2 of 650 to 997 000 ppm. During the pumping processes, the resistance of the nanostructured samples of Co_3O_4 and ZnO was measured. To determine the nanostructured film relative efficiency, resistance measurements of the plain Co_3O_4 film were also conducted.

3. RESULT AND DISCUSSION

The SEM images and X-ray diffraction graph of the Co_3O_4 nanostructures are shown in Fig. 1.

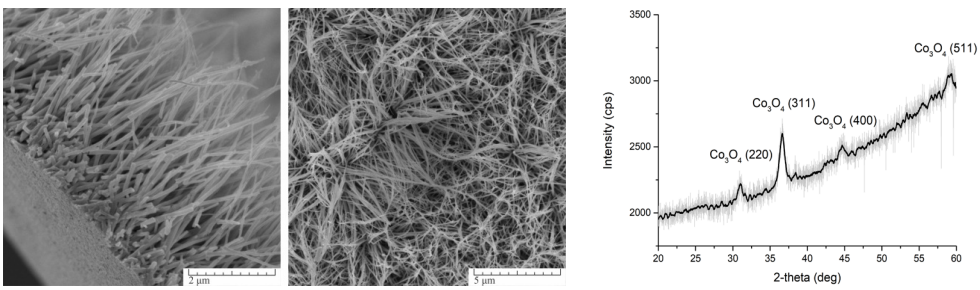


Fig. 1. SEM images and X-ray diffraction graph of Co_3O_4.

Figure 1 shows that the Co_3O_4 film consists of vertical, oriented, long (about 5 μm) strip-type nanostructures. Due to their small thickness relative to their length, the nanostructures lose their vertical alignment during their last stage of growth. Also, possibly due to their weight, the stripes lay chaotically, intertwining and forming a complicated net of porous labyrinths. The X-ray diffraction analysis did not show any other compounds besides Co_3O_4. It can be observed on the diffraction graph that the sample has a high amorphous background and relatively low peak intensity. This is characteristic of nanostructures obtained by the use of carbamide, as observed in previous studies [48].

The SEM images and X-ray diffraction graph of the ZnO structures are shown in Fig. 2.

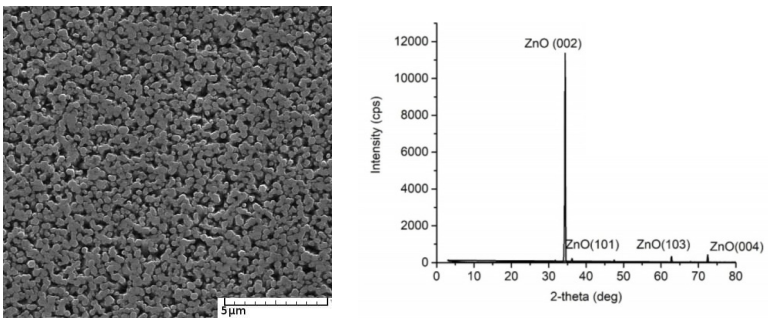


Fig. 2. SEM images and X-ray diffraction graph of ZnO.

The ZnO film consists of strictly ordered, vertically aligned nanorods with a distinctive hexagonal prism form, approximately 2 μm in length. Due to the lower length-to-thickness ratio, relative to the Co_3O_4 nanostructures, the ZnO rods retain their vertical alignment during the entire growth process. However, it is clear that the porosity of the ZnO nanostructures is much lower than for Co_3O_4.

The X-ray diffraction analysis did not show any crystalline compounds besides ZnO. On the diffraction graph, the distinctive peak corresponding to the (002) plane can be seen. The intensities of other peaks are significantly lower than the (002) peak, confirming the dominant orientation of the nanostructures in the vertical direction. The amorphous background of the film is low, indicating a high degree of crystallinity.

Study of the Resistance Dependence of the Co_3O_4 Samples on Gas Mixture Composition and Concentration

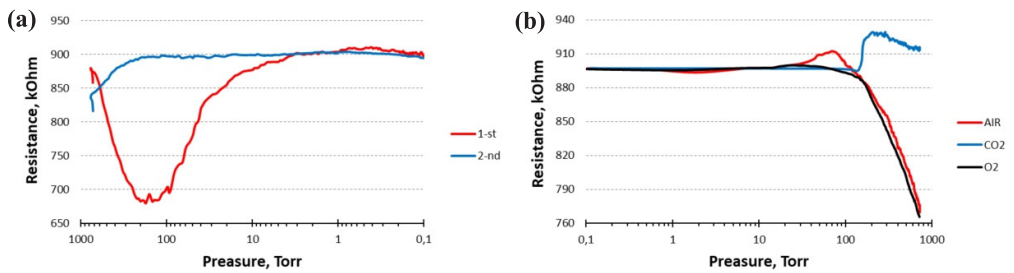


Fig. 3. (a) Dependence of Co_3O_4 sample resistance on gas pressure during air pumping from the chamber; (b) dependence of Co_3O_4 sample resistance on gas pressure during gas (air, CO_2, O_2) pumping into the chamber.

During the first pumping-out of air, down to 175 Torr (Fig. 3(a), red line), the resistance of the sample significantly reduces (approximately 23 %). Afterwards, when the pressure reduced from 175 to 0.1 Torr, an increase in the sample resistance can be observed. The resistance "pit" in the graph may be explained by the release of adsorbed water and gas molecules from the sample surface. The increase in sample resistance as the pressure decreases further may be due to the same mechanism as described above; that is, the sensor sensitivity to the pressure of O_2-containing gas atmospheres. During the pumping-out process, electrons which have been released during oxygen desorption recombine with holes. The amount of free charge carriers decreases, and the resistance grows.

During the second pumping-out process (Fig. 3(a), blue line), the same resistance pit is not observed – the sample does not adsorb atmospheric water, and only the oxygen desorption process influences the change in resistance. The plot shows an 11 % increase in resistance during the first stage of the pumping-out process, and a relatively fast stabilisation of the resistance due to the equilibrium stage.

The pumping-in of CO_2 to 0.1 Torr in the chamber (Fig. 3(b), blue line) leads to an increase in sample resistance up to 4 %. The mechanism of CO_2 adsorption on the metal oxide nanostructures is as described in the introduction.

The pumping-in process of O_2 to 0.1 Torr in the chamber (Fig. 3(b), black line) leads to a sharp decrease in sample resistance. This phenomenon may be explained by the formation of an ionic layer on the sample surface. O_2 plays a major role in the adsorption process, due to its high electronegativity and electron pairs, which help it to adsorb on the surface. Diatomic oxygen molecules act as electron acceptors, going through an ionisation process and forming the ionic layer on the sample surface, assisting in the sorption of other gas molecules.

Electrons on the metal oxide surface act as the active centre for adsorption of atmospheric O_2. O_2 from the environment diffuses to free sectors of O atoms on the metal oxide surface, accepting one or two free electrons. In this way, the chemisorption of O_2 molecules decreases electron mobility, leading to a decrease in conductivity of the n-type materials, and diminishing the adsorption activity centre ability to further bind other particles from the atmosphere.

The pumping-in process of air up to 0.1 Torr in the chamber (Fig. 3(b), red line) leads to an increase in sample resistance up to 3 %, which later decreases to 15 %, relative to the starting resistance of the sample. This may be explained by the fact that air is a mixture of gases, containing both O_2 and CO_2. In this case, the adsorption of CO_2 in the presence of O_2 occurs. As mentioned previously, O_2 additively promotes other molecule sorption into the system, and, thus, the distinctive CO_2 peak can be observed at low concentrations of air during pumping-in. At high pressure, sorption of O_2 becomes the dominant process, and the distinctive resistance decrease because O_2 resembles the O_2 pumping-in plot (Fig. 3(b), black line).

The dependence of relative change of sample resistance on CO_2 amount, for both Co_3O_4 nanostructures and plain films, is shown in Fig. 4.

It can be seen that the nanostructured sample shows a higher relative resistance with higher CO_2 amount than the plain film. The difference becomes especially distinctive at $m_{co2}/m_{air} = 1$ and above, and the maximum ratio of 78 %. Higher performance of the nanostructured sample is explained by the difference in the working surface area. In the case of the plain film, the gas contacted only with the top of the surface, but

in the nanostructured sample, the gas molecules penetrated between the nano-strips, and made contact with the pore surfaces, significantly increasing the amount of active bonds involved in the sorption process.

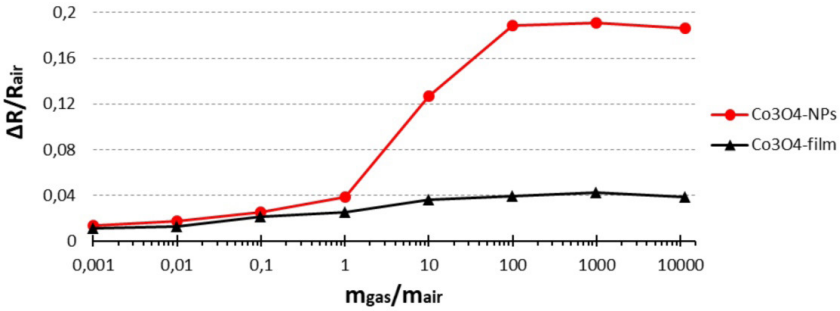


Fig. 4. Dependence of the relative change of different Co_3O_4 sample resistances on CO_2 amount in the chamber.

Figure 5 demonstrates the dependence of sample resistance changes on CO_2 amount in the chamber for metal oxides with different types of conductivity.

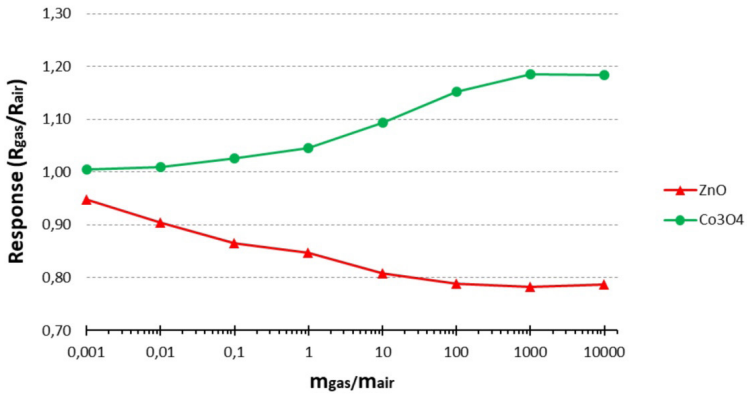


Fig. 5. Dependence of the relative change of the Co_3O_4 and ZnO sample resistances on CO_2 amount in the chamber.

In the case of Co_3O_4 (p-type semiconductor), the resistance grows with an increase in CO_2 amount, with the opposite trend occurring in the case of ZnO (n-type semiconductor). The curves confirm both of the mechanisms described above, namely, the mechanism of adsorption of CO_2 for different type of semiconductors and the mechanism of changes of semiconductor conductivity in the case of varying O_2 concentration in the air-gas mixture. The response magnitudes for both samples are approximately the same. This indirectly proves that the sensor sensitivities do not depend on the material conductivity type, but are defined by the existence of nanostructures (sample porosity and total surface area).

4. CONCLUSIONS

In this study, nanostructured samples of Co_3O_4 and ZnO were obtained by the hydrothermal method. The nanostructured samples showed a response to addition of CO_2 to a gas chamber, and the response was dependent on pressure. These observed effects demonstrate the potential use of these materials as working elements of resistive gas sensors.

The nanostructured samples of Co_3O_4 demonstrate higher resistance changes during increasing CO_2 concentration compared with smooth films of the same material, making the nanostructured material more attractive for sensor preparation.

ACKNOWLEDGEMENTS

This study has been supported by internal research grant No. 14-95/20 of Daugavpils University "Development of a Gas Sensor Based on Cobalt Oxide Nanostructures".

REFERENCES

1. Korotchenkov, G. (2013). *Handbook of gas sensor materials.* New York: Springer. DOI:10.1007/978-1-4614-7165-3
2. Liu, X., Cheng, S, Liu, H., Hu, S., Zhang, D., & Ning, H. (2012). A Survey on Gas Sensing Technology. *Sensors*, *12*, 9635–9665. DOI: 10.3390/s120709635
3. Janata, J. (2009). *Principles of chemical sensors*. New York: Springer.
4. Joshi, N., Hayasaka, T., Liu, Y., Liu, H., Oliveira, O. N., & Lin, L. (2018). A Review on Chemiresistive Room Temperature Gas Sensors Based on Metal Oxide Nanostructures, Graphene and 2D Transition Metal Dichalcogenides. *Microchimica Acta, 185* (4), 185–213. DOI:10.1007/s00604-018-2750-5
5. Kampa, M., & Castanas, E. (2008). Human Health Effects of Air Pollution. *Environmental Pollution*, *151*, 362–367. DOI: 10.1016/j.envpol.2007.06.012
6. Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., … & Preuss, P. W. (2013). The Changing Paradigm of Air Pollution Monitoring, *Environmental Science & Technology*, *47*, 11369−11377. DOI:10.1021/es4022602
7. Babadjouni, R., Hodis, D., Radwanski, R., Durazo, R., Patel, A., Liu, Q., & Mack, W. (2017). Clinical Effects of Air Pollution on the Central Nervous System: A Review. *Journal of Clinical Neuroscience*, *43*, 16–24. DOI:10.1016/j.jocn.2017.04.028
8. Glencross, D., Ho, T., Camica, N., Hawrylowicz, C., & Pfeffer, P. (2020). Air Pollution and its Effects on the Immune System. *Free Radical Biology and Medicine*, *151*, 56–68. DOI:10.1016/j.freeradbiomed.2020.01.179
9. Lobur, M., Korpyljov, D., Jaworski, N., Iwaniec, M., & Marikutsa, U. (2020). Arduino based ambient air pollution sensing system. In *IEEE XVIth International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH),* (pp. 32–35), 22-26 April 2020, Lviv, Ukraine: Lviv Polytechnic National University. DOI: 10.1109/MEMSTECH49584.2020.9109460.

10. Hasenfratz D., Saukh O., Sturzenegger S., &Thiele L. (2012) Participatory air pollution monitoring using smartphones. In *Proceedings of the 2nd International Workshop on Mobile Sensing*, (pp. 1–5), 16–20 April 2012, Beijing, China.
11. Galstyan, V., Comini, E., Baratto, C., Faglia, G., & Sberveglieri, G. (2015). Nanostructured ZnO Chemical Gas Sensors. *Ceramics International*, *41*, 14239–14244. DOI:10.1016/j.ceramint.2015.07.052
12. Zhu, L., & Zeng, W. (2017). Room-Temperature Gas Sensing of ZnO-Based Gas Sensor: A Review. *Sensors and Actuators A*, *267*, 242–261. DOI:10.1016/j.sna.2017.10.021
13. Dey, A. (2018). Semiconductor Metal Oxide Gas Sensors: A Review. *Materials Science & Engineering B*, *229*, 206–217. DOI: 10.1016/j.mseb.2017.12.036
14. Yoon, J., Kim, H., Jeong, H., & Lee, J. (2014). Gas Sensing Characteristics of P-Type Cr2O3 and Co3O4 Nanofibers Depending on Inter-Particle Connectivity. *Sensors and Actuators B*, *202*, 263–271. DOI:10.1016/j.snb.2014.05.081
15. Moseley, P. T. (2017). Progress in the Development of Semiconducting Metal Oxide Gas Sensors: A Review. *Measurement Science and Technology, 28* (8), 082001 (15 p.). DOI:10.1088/1361-6501/aa7443
16. Jing, Z., & Zhan, J. (2008). Fabrication and Gas-Sensing Properties of Porous ZnO Nanoplates. *Advanced Materials, 20* (23), 4547–4551. DOI:10.1002/adma.200800243
17. Liu, C., Zhao, L., Wang, B., Sun, P., Wang, Q., … & Lu, G. (2017). Acetone Gas Sensor Based on NiO/ZnO Hollow Spheres: Fast Response and Recovery, and Low (ppb) Detection Limit. *Journal of Colloid and Interface Science, 495,* 207–215. DOI:10.1016/j.jcis.2017.01.106
18. Sonker, R. K., Sabhajeet, S. R., Singh, S., & Yadav, B. C. (2015). Synthesis of ZnO Nanopetals and its Application as NO2 Gas Sensor. *Materials Letters, 152,* 189–191. DOI:10.1016/j.matlet.2015.03.112
19. Yu, L., Guo, F., Liu, S., Yang, B., Jiang, Y., … & Fan, X. (2016). Both Oxygen Vacancies Defects and Porosity Facilitated NO 2 Gas Sensing Response in 2D ZnO Nanowalls at Room Temperature. *Journal of Alloys and Compounds, 682,* 352–356. DOI:10.1016/j.jallcom.2016.05.053
20. Zhu, L., & Zeng, W. (2017). Room-Temperature Gas Sensing of ZnO-Based Gas Sensor: A Review. *Sensors and Actuators A: Physical, 267,* 242–261. DOI:10.1016/j.sna.2017.10.021
21. Yoon, J.-W., Kim, H.-J., Jeong, H.-M., & Lee, J.-H. (2014). Gas Sensing Characteristics of p-Type Cr2O3 and Co3O4 Nanofibers Depending on Inter-Particle Connectivity. *Sensors and Actuators B: Chemical, 202,* 263–271. DOI:10.1016/j.snb.2014.05.081
22. Deng, S., Chen, N., Deng, D., Li, Y., Xing, X., & Wang, Y. (2015). Meso- and Macroporous Coral-like Co3O4 for VOCs Gas Sensor. *Ceramics International, 41* (9), 11004–11012.DOI:10.1016/j.ceramint.2015.05.045
23. Zoolfakar, A. S., Ahmad, M. Z., Rani, R. A., Ou, J. Z., Balendhran, S., … & Kalantar-zadeh, K. (2013). Nanostructured Copper Oxides as Ethanol Vapour Sensors. *Sensors and Actuators B: Chemical, 185,* 620–627. DOI:10.1016/j.snb.2013.05.042
24. Fine, G. F., Cavanagh, L. M., Afonja, A., & Binions, R. (2010). Metal Oxide Semi-Conductor Gas Sensors in Environmental Monitoring. *Sensors, 10* (6), 5469–5502. DOI:10.3390/s100605469
25. Wang, C., Yin, L., Zhang, L., Xiang, D., & Gao, R. (2010). Metal Oxide Gas Sensors: Sensitivity and Influencing Factors. *Sensors, 10* (3), 2088–2106. DOI:10.3390/s100302088
26. Zhang, J., Liu, X., Neri, G., & Pinna, N. (2015). Nanostructured Materials for Room-Temperature Gas Sensors. *Advanced Materials, 28* (5), 795–831. DOI:10.1002/adma.201503825
27. Zhu, L., & Zeng, W. (2017). A Novel Coral Rock-like ZnO and its Gas Sensing. *Materials Letters, 209,* 244–246. DOI:10.1016/j.matlet.2017.08.020

28. Arregui, F. J. (ed.). (2009). *Sensors based on nanostructured materials.* New York: Springer. DOI:10.1007/978-0-387-77753-5
29. Fort, A., Panzardi, E., Vignoli, V., Hjiri, M., Aida, M., Mugnaini, M., & Addabbo, T. (2019). Co3O4/Al-ZnO Nano-Composites: Gas Sensing Properties. *Sensors, 19* (4), 760. doi:10.3390/s19040760
30. Shankar, P., & Rayappan, J. (2015) Gas Sensing Mechanism of Metal Oxides: The Role of Ambient Atmosphere, Type of Semiconductor and Gases – A Review, *Science Letters Journal, 4,* 126, (18 p.). http://www.cognizure.com/scilett.aspx?p=200638572
31. Yang, S., Jiang, C., & Wei, S. (2017). Gas Sensing in 2D Materials. *Applied Physics Reviews, 4* (2), 021304. DOI:10.1063/1.4983310
32. Krasovska, M., Gerbreders, V., Mihailova, I., Ogurcovs, A., Sledevskis, E., … & Sarajevs, P. (2018). ZnO-Nanostructure-Based Electrochemical Sensor: Effect of Nanostructure Morphology on the Sensing of Heavy Metal Ions. *Beilstein Journal of Nanotechnology, 9,* 2421–2431. DOI:10.3762/bjnano.9.227
33. Azeez, O. A., Sabry, R. S., Hassan, M. A. M., & Madlul, S. F. (2015). Synthesis and Characteristics of Screen Printed ZnO Thick Films Nanostructures Grown Using Different Methods. *Journal of Materials Science: Materials in Electronics, 26* (6), 4051–4061. DOI:10.1007/s10854-015-2944-0
34. Ferraz, H. C., Machado, D. F., & de Resende, N. S. (2017). Nanostructured Screen-Printed Electrodes Based on Titanate Nanowires for Biosensing Applications. *Materials Science and Engineering: C, 70,* 15–20. DOI:10.1016/j.msec.2016.08.046
35. Sarkar, K., Braden, E. V., Bonke, S. A., Bach, U., & Müller-Buschbaum, P. (2015). Screen-Printing of ZnO Nanostructures from Sol-Gel Solutions for their Application in Dye-Sensitized Solar Cells. *ChemSusChem, 8* (16), 2696–2704. DOI:10.1002/cssc.201500450
36. Carotta, M. C., Martinelli, G., Crema, L., Malagù, C., Merli, M., … &Traversa, E. (2001). Nanostructured Thick-Film Gas Sensors for Atmospheric Pollutant Monitoring: Quantitative Analysis on Field Tests. *Sensors and Actuators B: Chemical, 76* (1–3), 336–342. DOI:10.1016/s0925-4005(01)00620-7
37. Solis, J. ., Saukko, S., Kish, L., Granqvist, C., & Lantto, V. (2001). Semiconductor Gas Sensors Based on Nanostructured Tungsten Oxide. *Thin Solid Films, 391* (2), 255–260. DOI:10.1016/s0040-6090(01)00991-9
38. Zhou, Q., Chen, W., Xu, L., & Peng, S. (2013). Hydrothermal Synthesis of Various Hierarchical ZnO Nanostructures and their Methane Sensing Properties. *Sensors, 13* (5), 6171–6182. DOI:10.3390/s130506171
39. Mihailova, I., Gerbreders, V., Bulanovs, A., Tamanis, E., Sledevskis, E., … & Sarajevs, P. (2014). Controlled growth of well-aligned ZnO nanorod arrays by hydrothermal method. In *the 8th International Conference on Advanced Optical Materials and Devices (AOMD-8),* (9421–23), 25–27 August 2014, Riga, Latvia.
40. Krasovska, M., Gerbreders, V., Paskevics, V., Ogurcovs, A., & Mihailova, I. (2015). Obtaining a Well-Aligned ZnO Nanotube Array Using the Hydrothermal Growth Method. *Latvian Journal of Physics and Technical Sciences, 52* (5), 28–40. DOI:10.1515/lpts-2015-0026
41. Mokoena, T. P., Swart, H. C., & Motaung, D. E. (2019). A Review on Recent Progress of p-Type Nickel Oxide Based Gas Sensors: Future Perspectives. *Journal of Alloys and Compounds,* 267–294. DOI:10.1016/j.jallcom.2019.06.329
42. Kim, H.-J., & Lee, J.-H. (2014). Highly Sensitive and Selective Gas Sensors Using p-Type Oxide Semiconductors: Overview. *Sensors and Actuators B: Chemical, 192,* 607–627. DOI:10.1016/j.snb.2013.11.005
43. Ji, H., Zeng, W., & Li, Y. (2019). Gas Sensing Mechanisms of Metal Oxide Semiconductors: A Focus Review. *Nanoscale, 11,* 22664–22684. DOI:10.1039/c9nr07699a

44. Barsan, N., Simion, C., Heine, T., Pokhrel, S., & Weimar, U. (2010). Modelling of Sensing and Transduction for p-Type Semiconducting Metal Oxide Based Gas Sensors. *Journal of Electroceramics, 25* (1), 11–19. DOI:10.1007/s10832-009-9583-x

45. Wurzinger, O., & Reinhardt, G. (2004). CO-sensing Properties of Doped SnO2 Sensors in H2-rich Gases. *Sensors and Actuators B: Chemical, 103* (1–2), 104–110. DOI:10.1016/j.snb.2004.04.041

46. Ostrick, B., Fleischer, M., Meixner, H., & Kohl, C.-D. (2000). Investigation of the Reaction Mechanisms in Work Function Type Sensors at Room Temperature by Studies of the Cross-Sensitivity to Oxygen and Water: The Carbonate–Carbon Dioxide System. *Sensors and Actuators B: Chemical, 68* (1–3), 197–202. DOI:10.1016/s0925-4005(00)00429-9

47. Madou, M. J., & Morrison, S. R. (1989). *Chemical sensing with solid state devices*. San Diego: Academic Press.

48. Krasovska, M., Gerbreders, V., Sledevskis, E., Gerbreders, A., Mihailova, I., … & Ogurcovs, A. (2020). Hydrothermal Synthesis of ZnO Nanostructures with Controllable Morphology Change. *CrystEngComm, 22* (8), 1346–1358. DOI:10.1039/c9ce01556f

sciendo

LATVIAN JOURNAL OF PHYSICS
AND TECHNICAL SCIENCES
2021, N 5

DOI: 10.2478/lpts-2021-0037

APPLICATION OF 3D ROUGHNESS PARAMETERS FOR WEAR INTENSITY CALCULATIONS

N. Bulaha*, O. Linins, A. Avisane

Institute of Mechanics and Mechanical Engineering
Riga Technical University,
6B Kipsalas Str., LV-1048, Riga, LATVIA
*e-mail: mmk@rtu.lv

In this paper, calculations of 3D parameter *Vm* (material volume) of surfaces with irregular roughness and comparison with experimental data were performed, with further application of this parameter in calculations of wear intensity. First, using Mountains Map software for profilometric measurements, 3D roughness processing and determination of material volume *Vm* at specific relative levels γ were performed. The next step was an additional analysis of the distribution of surface ordinates using a theoretical and experimental Laplace function. The given check confirmed that for mostly surfaces with irregular roughness the ordinate distribution corresponds to the normal Gaussian distribution law, but in cases when the asymmetry of the ordinate distribution function goes outside the permissible limits ($|\Delta Ssk|$> 10%), errors> 10 % occur. On this basis, the mathematical formula of the material volume *Vm* was derived, and the obtained calculations were compared with the measured values. The results showed that the calculated values of the parameter *Vm* were very close to the experimental data ($|\Delta Vm|$<10 %), while at the relative level γ=+3, errors occurred that was related to the deviation from the normal distribution law. It was concluded that the given parameter could be used in the calculations of linear wear intensity, knowing the relative level γ.

Keywords: *3D roughness parameters, friction, material volume, wear.*

1. INTRODUCTION

Nowadays, the issue of surface wear does not lose its relevance because all the time designers and technologists are trying to find effective solutions to increase the wear resistance of surfaces. As it is known, the wear intensity is influenced by several factors / parameters: the material and its properties, load, lubrication conditions, and roughness. The surface quality is very important in this regard because the type of surface treatment and the roughness parameters largely determine the service life of the friction pair.

Surfaces for the friction pairs of the machine apparatus mainly are treated with abrasive materials, which ensure the required surface smoothness and accuracy. Surfaces, after such treatment, are characterised by irregular arrangement of micro-irregularities, and roughness parameters for such surfaces are calculated using normal random field theory [1].

At this time, Standard ISO 25178-2, which defines 3D surface roughness parameters, does not give a complete picture of the physical meaning of these parameters. No relationships are shown between the surface roughness parameters, which would help determine the factors that affect a particular parameter and the way they can be changed / improved.

In the calculations of linear wear intensity, one of the important roughness parameters is the volume of the deformed material. To determine it, the following tasks have been performed in the given study:

1. Compliance check of the ordinate distribution function of surfaces with irregular roughness to the normal distribution law using an additional method – Laplace function analysis.
2. Determination of roughness parameter *Vm* at different relative levels γ in the Mountains Map software.
3. Output of parameter *Vm*, comparing the calculated and experimental values.
4. Calculations of linear wear intensity using the roughness parameter *Vm*.

2. SURFACE ORDINATE DISTRIBUTION

In the given study, six different surfaces were studied: flat ground, polished, lapped, shot peened, sandblasted and after electro-erosion.

As it was mentioned above, surface roughness after treatment with abrasive materials was irregular. For this type of surface, it is important to check compliance with the normal distribution law in order to make further calculations. In scientific works [2], the compliance of surface ordinate distribution function to the normal distribution law was substantiated using such criteria as:

1. asymmetry *Ssk* of the surface ordinate distribution function;
2. the excess *Sku* of the surface ordinate distribution function;
3. Pearson criterion;
4. correlation function.

An additional criterion that can be used to solve this issue is the Laplace function, which represents the area of a material at a given level *c* (in percent).

In the general case, the Laplace function is expressed by Eq. (1):

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-z^2}{2}} dz. \quad (1)$$

The parameter given in the standard ISO 25178-2 is defined as a ratio of the area of the material at a height c to the evaluation area (Fig. 1).

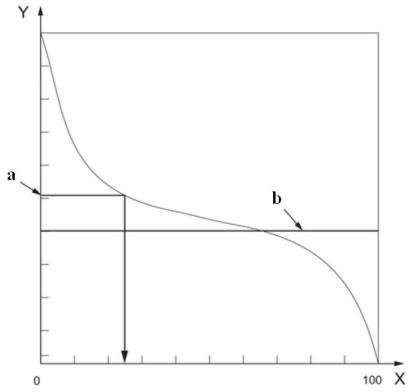


Fig. 1. Areal material ratio:
X – areal material ratio *Smr(c)*, in percent; Y – height; a – level *c*; b – reference plane [3].

The graph of the Laplace function can be constructed for any of the studied surfaces at the levels c, using Eq. (2):

$$E\{Smr(c)\} = \frac{1}{A}\iint_{\Omega} \xi\{r(z,c)\} = 1 - \Phi\left(\frac{c}{Sq}\right), \quad (2)$$

where A – nominal surface area, mm^2; c – level, μm; Sq – root mean square height, μm.

$$c = \gamma \cdot Sq, \quad (3)$$

where γ – relative level. $\gamma_{standart}$=3, 2, 1, 0, -1, -2, -3.

Using the Mountains Map software [4], the area of the material at the relative levels γ = +3, +2, +1, 0, -1, -2, -3 was determined for the studied surfaces. In Table 1, the calculated and measured values of the parameter *Smr(c)* for a flat ground surface were summarised and the absolute error was calculated. The calculated values of this parameter are fully compliant with the Laplace function indicators.

Table 1. Comparison of Theoretical and Experimental Values of Roughness Parameter *Smr(c)* for Flat Ground Surface

γ	$Smr(c)_{calc}$, %	$Smr(c)_{measured}$, %	$\lvert\Delta Smr(c)\rvert$, %
+3	0.13	0.12	8.30
+2	2.28	2.20	3.30
+1	15.87	17.30	9.04
0	50.00	47.90	4.20
-1	84.13	85.40	1.50
-2	97.72	97.80	0.08
-3	99.87	100.00	0.14

In Fig. 2, the graphical representation of the experimental values of the parameter *Smr(c)* and the Laplace function are shown. It can be concluded that for the given surface the distribution of ordinates very exactly corresponds to the normal distribution law and falls within the deviation range of +/- 10 %.

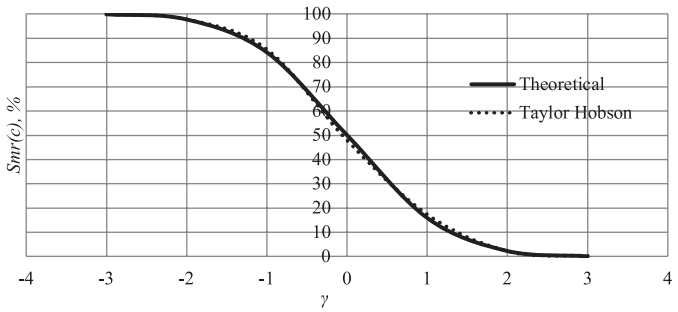


Fig. 2. Graphical representation of the calculated and experimental values of the parameter *Smr(c)*.

In turn, for the other studied surfaces there are deviations from the standard values of the Laplace function at the upper relative levels γ= +2 and +3 (Fig. 3). The value of the absolute error Δ*Smr(c)* is greater than 10 % in several cases. This result is influenced by the asymmetry *Ssk* of the surface ordinate distribution function. The experimental values of *Smr(c)* at *Ssk*>0.05 and *Ssk* <-0.05 differ from the theoretical ones by up to 60 %.

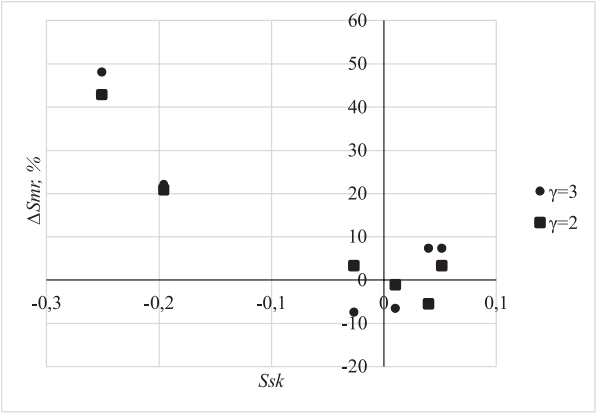


Fig. 3. Influence of asymmetry *Ssk* of ordinate distribution function on absolute error |Δ*Smr*|.

3. CALCULATION OF MATERIAL VOLUME *Vm*

The volume of the micro-irregularities of the rough surface above the level *c* is a very important feature in the calculations of the contact deformation. The volume of surface material at a given level expresses the amount of deformed material in the friction process.

The average volume of material can be obtained analogously to the surface profile area by integrating the surface cross-sectional area *A(c)* at levels *c* [5].

$$Vm = \int_c^{\infty} A(c)dc. \tag{4}$$

Due to the fact that

$$A(c) = Smr(c) \cdot Aa, \tag{5}$$

the mathematically expected value of the material volume, taking into account Eq. (2), will be equal to:

$$E\{Vm\} = Aa \int_c^{\infty} E\{Smr(c)\}dc, \tag{6}$$

$$E\{Vm\} = Aa \int_c^{\infty} \left[1 - \Phi\left(\frac{c}{Sq}\right)\right] dc, \tag{7}$$

$$E\{Vm\} = Sq \left\{\frac{1}{\sqrt{2\pi}} exp\left(-\frac{c^2}{2Sq^2}\right) - \frac{c}{Sq}\left[1 - \Phi\left(\frac{c}{Sq}\right)\right]\right\}. \tag{8}$$

Expression (8) indicates that the material volume *Vm* varies depending on the relative level γ. As γ changes from -∞ to + ∞, the material volume increases. At γ values -∞ <γ <-1, it can be considered that the volume of deformed material practically does not change according to the linear law.

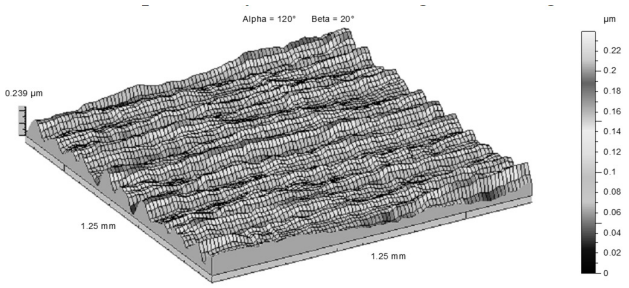


Fig. 4. Topography of a flat ground surface.

In Fig. 4, a flat ground surface is visible, the ordinate distribution of which corresponds to the normal Gaussian distribution law, which was justified in the previous paragraph. For a given surface *Vm* calculations were performed at different levels *c* subtracted from the mean plane. Theoretical values of material volume were calculated according to Eq. (8), while the experimental data were obtained by Mountains Map software, using a function "slices" to generate a surface divided into two parts – material and voids. In Fig. 5, the material volume is marked in grey colour, and void volume – in black. By selecting the required level, it is possible to obtain the required *Vm* values. In addition, this function represents the area of the material at a given level, which corresponds to the values of the parameter *Smr(c)*.

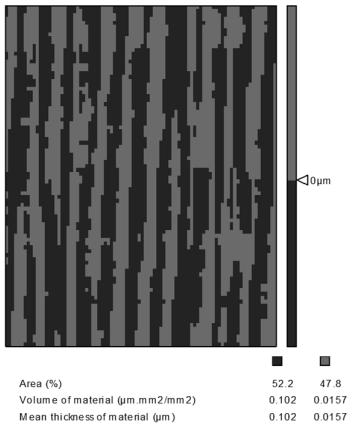


Fig. 5. Representation of surface material and void volume at level *c*.

Table 2 shows the calculated and experimental values of parameter *Vm* and their comparison in the form of absolute error for the flat ground surface. At all relative levels, except the highest (γ =+3), the calculated values coincide very precisely with the experimental data, and the absolute error falls within the range of +/-10 %. The discrepancy at the relative level γ=+3 arises because the *Ssk* value of the ordinate distribution function has a large deviation (> 10 %) from the standard values of normal distribution law.

Table 2. Comparison of Theoretical and Experimental Values of Roughness Parameter *Vm* for Flat Ground Surface

γ	Vm_{calc}, mm^3/mm^2	$Vm_{measured}$ mm^3/mm^2	$\|\Delta Vm\|$, %
+3	1.50E-08	1.08E-08	28.03
+2	3.33E-07	3.24E-07	2.74
+1	3.27E-06	3.46E-06	5.90
0	1.56E-05	1.57E-05	0.38
-1	4.25E-05	4.23E-05	0.39
-2	7.87E-05	7.87E-05	0.04
-3	1.18E-04	1.18E-04	0.33

Graphically, the experimental and theoretical values of the parameter *Vm* are shown in Fig. 6. In the graph, we can see only those levels, at which deformations during friction can actually occur depending on the applied load [6].

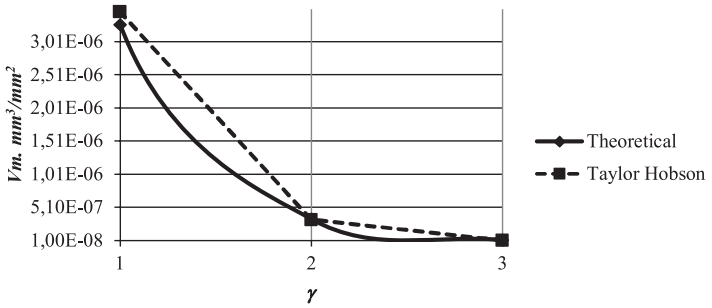


Fig. 6. Graphical representation of the calculated and experimental values of the parameter *Vm*.

Calculations of material volume *Vm* and comparison with experimental data were performed for all six studied surfaces to understand the extent to which the asymmetry of the ordinate distribution function affected the results. From the data in Table 3, it can be concluded that very large errors between theoretical and experimental values occur at the relative levels γ = +2 and +3; it is also related to the deviations of the asymmetry *Ssk*.

Table 3. Influence of Asymmetry *Ssk* of Ordinate Distribution Function on Absolute Error |Δ*Vm*|

No. Nr.	Type of surface treatment	*Ssk*	|Δ*Vm*|, %		
			γ=+3	γ=+2	γ=+1
1	Flat grinding	0.051	28.03	2.74	5.9
2	Lapping	0.039	14.50	0.28	1.16
3	Sand blasting	-0.196	2.53	23.55	8.95
4	Electro erosion	-0.176	53.79	40.79	9.26
5	Shot peening	-0.026	17.42	1.31	1.31
6	Polishing	-0.247	98.41	86.76	3.83

4. CALCULATION OF WEAR INTENSITY J_H

In the given study, the analysis of the surface contacting process was performed using the equivalent surface. The application of such a method for solving engineering tasks has been justified in several studies [5]–[7].

In general, there is a direct deformation of the material volume *Vm* in the contact zones, but it will not localize in a specific volume, but will spread deeper in the body of the micro-irregularity. In turn, it is substantiated that the deformation spread factor at the deeper layers of the material can be disregarded.

According to [8], the linear wear intensity J_h is determined by Eq. (9):

$$E\{J_h\} = \frac{Vm}{n \cdot d \cdot Aa}, \quad (9)$$

where *Vm* – material volume, mm³/mm²;
n – a number of cycles to failure;
d – contact length, mm;
Aa – nominal contact area, mm².

In Eq. (10), the roughness parameter *Vm*, which expresses the amount of deformed material, is figured out. Knowing the relative level γ, this parameter can be easily determined for any surface. In turn, at the beginning of solving an engineering task, it is important to determine how the micro-irregularities of the studied surface will be deformed.

The contact area to a greater extent is characterised by the ability of surface micro-irregularities to be elastically deformed, what can be described by the contact type criterion *CC* [6]. The calculation formula of *CC* includes roughness parameters – *Sa*, *RSm* and material mechanical properties – micro-hardness and elastic constant:

$$CC = \frac{RSm \cdot \theta \cdot H_\mu}{Sa}, \quad (10)$$

where *RSm* – an average step of micro-irregularities, mm;
θ – elastic constant of material, m²/MN;
H_μ – surface microhardness, N/mm²;
Sa – arithmetic mean height, μm;

$$\theta = \frac{1-\mu_2^2}{E_2}, \quad (11)$$

where μ – Poisson's ratio;
E – elastic modulus, MN/m².

If the value of the contact type criterion is >1.7, then the condition of elastic deformation is fulfilled, but if *CC* <0.7, the contact is plastic.

Accordingly, depending on the type of contact, the level of deformation can be expressed by the following equations:

$$F_1\{\gamma\} = \left(\frac{q \cdot \theta_{sum}}{k_{elast}}\right) \cdot \left(\frac{RSm}{Sa}\right), \quad (12)$$

$$F_2\{\gamma\} = \frac{q}{H \cdot k_{plast}}, \quad (13)$$

where $F_{1,2}(\gamma)$ – tabulated functions [8];
q – load, MN/m^2;
$k_{elast,\ plast}$ – a coefficient that depends on the anisotropy of the surface (the parameter *Str*);
θ_{sum} – total elastic constant of material, m^2/MN.

The total elastic constant is calculated according to the following equation:

$$\theta_{sum} = \frac{1-\mu_1^2}{E_1} + \frac{1-\mu_2^2}{E_2}. \quad (14)$$

The number of cycles to failure also depends on the type of deformation – elastic or plastic, and n_{el} and n_{plast} can be expressed by the equations:

$$E\{n_{el}\} = \left(\frac{3\pi\theta\sigma_0}{4kf} \cdot \sqrt{\frac{r}{Sz \cdot \varepsilon}}\right)^{t_{el}}, \quad (15)$$

$$\mathrm{E}\{n_{plast}\} = \left(\frac{e_0}{\sqrt{2}} \sqrt{\frac{\sigma_T - 2fH_\mu}{\sigma_T + 2fH_\mu}} \cdot \frac{r}{Sz \cdot \varepsilon}\right)^{t_{plast}}, \quad (16)$$

where e_0, σ_0 – extrapolated values of the fatigue curve, MN/m^2;
k – constant depending on friction-fatigue characteristics, k=3(1+μ);
f – coefficient of friction;
$t_{el,\ plast}$ – fatigue curve parameters;
σ_T – material yield strength, N/mm^2;
r – mean summit curvature [1], mm^{-1};
Sz – max height of surface, μm;
ε – relative convergence, mm^{-2}.

$$r = \frac{1}{2}\pi^2 \cdot Sq \cdot \left(\frac{4}{RSm^2}\right)(1 + Str^2)\gamma\,, \quad (17)$$

$$\varepsilon = 1{,}54 \cdot \left(\frac{r}{Sz}\right)^{\frac{1}{5}} (q \cdot \theta)^{\frac{2}{5}}. \quad (18)$$

Determination of the wear intensity according to Eq. (9) requires the study of the contact size, which is expressed as the length of the actual contact area in the direction of friction. According to [8], the mathematically expected value of the contact length $E\{d\}$ is determined by the equation:

$$E\{d\} = RSm[1 - \Phi(\gamma)]e^{-\frac{\gamma^2}{2}}. \quad (19)$$

Linear wear is expressed by the following formula:

$$h = J_h \cdot L_{tr}, \quad (20)$$

where L_{tr} – a friction path, μm;

$$L_{tr} = V \cdot t, \quad (21)$$

where V – traverse speed, m/min;
t – working time, st.

Thus, knowing the output calculation data: roughness parameters, physical-mechanical, friction-fatigue and constructive characteristics of friction pair materials, it is possible to calculate the linear wear intensity according to Eq. (9). In this study, wear calculations for a lathe friction pair guide-table (Fig. 7) were made.

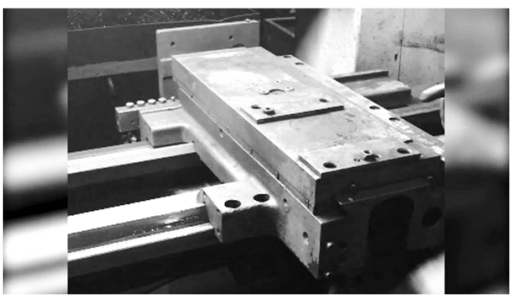


Fig. 7. Friction pair guide – table.

Table 4 summarises the data on the material characteristics of the friction pair,

as well as tabulated values of different coefficients. In this example, the wear intensity was calculated for the sliding guide – the part that absorbs the load during contact.

Table 4. Output Data of Friction-Pair Guide-Table for Wear Intensity Determination

Parameters Cast iron 300 ISO 185		Guide Steel C45 EN 10083-2	Table	Units
Roughness parameters	*Sa*	0.24	0.24	μm
	Sq	0.3	0.3	μm
	RSm_1	0.12	0.12	mm
	RSm_2	0.6	0.6	mm
	Str	0.22	0.24	-
	Vm	$1.19 \cdot 10^{-6}$	-	mm^3/mm^2
	Sz	2.25	-	μm
Physical-mechanical characteristics	*E*	$1.3 \cdot 10^5$	$2.1 \cdot 10^5$	MN/m^2
	μ	0.25	0.3	-
	H_μ	1800	1960	N/mm^2
Friction-fatigue characteristics	σ_0	1600	-	MN/m^2
	t_{el}	6.45	-	-
	f	0.1		-
	k	3.75		-
Constructive characteristics	*q*	0.5		MN/m^2
	V	0.5		m/min
	t	5000		st.

1. First, the elastic constant of the material was determined according to Eq. (11):

$$\theta = \frac{1-0.25^2}{1.3 \cdot 10^5} = 7.21 \cdot 10^{-6} (\text{m}^2/\text{MN}).$$

2. The contact type criterion *CC* according to Eq. (10):

$$CC = \frac{0.12 \cdot 7.21 \cdot 10^{-6} \cdot 1800}{0.24 \cdot 10^{-3}} = 6.49.$$

It can be assumed that in this case the elastic contact is provided (*CC*> 1.7).

3. The relative level γ according to Eq. (12) in the case of elastic contact is determined as follows:

$$F_1\{\gamma\} = \left(\frac{0.5 \cdot 1.15 \cdot 10^{-5}}{0.4562}\right) \cdot \left(\frac{0.12 \cdot 10^3}{0.24}\right) = 0.06 \cdot 10^{-2},$$

where k_{elast} = 0.4562 – a tabulated value obtained at *Str* = 0.12 / 0.6 = 0.2 [8].

In turn, the total elastic constant of the contact is equal to:

$\theta_{sum} = \frac{1-0.3^2}{2.1\cdot10^5} + \frac{1-0.25^2}{1.3\cdot10^5} = 1.15 \cdot 10^{-5}$(m^2/MN).

4. In this case, the calculated relative level γ = 2.3 [8].

5. The mean summit curvature is equal to:

$r = \frac{1}{2}\pi^2 \cdot 0.3 \cdot \left(\frac{4}{0.12^2}\right)(1 + 0.2^2) \cdot 2.3 = 0.99(mm^{-1})$.

6. The relative convergence was determined by Eq. (18):

$\varepsilon = 1.54 \cdot \left(\frac{0.99}{2.25\cdot10^{-3}}\right)^{\frac{1}{5}} (0.5 \cdot 7.21 \cdot 10^{-6})^{\frac{2}{5}} = 2.18 \cdot 10^{-3}$(mm^{-2}).

7. The number of cycles to failure was determined by Eq. (15):

$E\{n_{el}\} = \left(\frac{3\pi\cdot7.21\cdot10^{-6}\cdot1600}{4\cdot3.75\cdot0.1} \cdot \sqrt{\frac{0.99}{2.25\cdot10^{-3}\cdot2.18\cdot10^{-3}}}\right)^{6,45}$ =5693844682.

8. The value of the contact length is equal to:

$d = 0.12 \cdot [1 - \Phi(2.3)]e^{-\frac{2.3^2}{2}} = 9.13 \cdot 10^{-5}$(mm).

9. The volume of the deformed material according to theory will be equal to:

$Vm = 0.001 \cdot 0.3\left\{\frac{1}{\sqrt{2\pi}}e^{-\frac{2.3^2}{2}} - 2.3[1 - \Phi(2.3)]\right\} = 1.1 \cdot 10^{-6}$(mm^3/mm^2).

10. The linear wear intensity was determined by Eq. (9):

$J_h = \frac{1.1 \cdot 10^{-6}}{5693844682 \cdot 9.13 \cdot 10^{-5}} = 2.12 \cdot 10^{-12}$.

11. The friction path was determined by the traverse speed and working time:

$L_{tr} = 0.5 \cdot 5000 * 60 = 15 \cdot 10^{10}$ (μm).

12. The linear wear was determined as follows:

$h = 2.12 \cdot 10^{-12} \cdot 15 \cdot 10^{10}$ =0.32 (μm).

In addition, calculations were performed using the measured values of roughness parameters *Vm* and *Str* (Table 4). The linear wear intensity J_h=$2.3\cdot10^{-12}$ and linear wear *h*=0.35 μm. The absolute error in both cases is less than 10 %.

5. CONCLUSIONS

The study investigated the dependence of linear wear intensity J_h of the surface on the volume of the material *Vm*, the surface texture index *Str* and the arithmetic mean height *Sa*. In addition, the obtained equation of the material volume *Vm* contains such parameters as *Sq* – the root mean square height and the level *c*. To check the theoretical relationships of the parameter *Vm*, roughness measurements were performed using a 3D measuring equipment Taylor Hobson Talysurf Intra 50. The Mountains Map software was used to process the obtained data. A comparison of the theoretical values and experimental data shows that the absolute error of the surface material volume $|\Delta Vm|$ at the relative level γ=+3 in some cases reaches 100 %, at the level γ=+2 - $|\Delta Vm|$ <45 %, at lower levels - $|\Delta Vm|$ <10 %, what can be explained by the asymmetry of the surface ordinate distribution function.

Calculations of linear wear intensity J_h show that the offered roughness parameters *Vm*, *Str* and *Sa* simplify the calculation of wear because using measured values, it is not necessary to calculate each parameter separately.

REFERENCES

1. Рудзитис, Я. (2007). Контактная механика поверхностей, 2 – ая часть - Микротопография шероховатости поверхности (Surface Contact Mechanics. Part 2 – Surface Roughness Microtopography). Riga: Riga Technical University. (In Russian).
2. Bulaha, N. (2018). Calculations of surface roughness 3D parameters for surfaces with irregular roughness. In *17th International Scientific Conference "Engineering for Rural Development"* (1437–1444), 23–25 May 2018, Jelgava, Latvia.
3. ISO 25178-2:2012, Geometrical Product Specifications (GPS) – Surface Texture: Areal – Part 2: Terms, Definitions and Surface Texture Parameters.
4. Exploring Surface Texture. (2011). 7th edition. Great Britain: Taylor Hobson Limited. Available at http://www.taylor-hobson.com/uploads/learningzone/metrologybooks/Exploring%20Surface%20Texture%202014.pdf
5. Рудзитис, Я. А. (2007). Контактная механика поверхностей. Часть 1. (Mechanics of Surface Contact. Part 1). Riga: Riga Technical University. (in Russian).
6. Демкин, Н. Б., Рыжов, Э. В. (1981). Качество поверхности и контакт деталей машин (Surface Quality and Contact of Machine Parts). М: Машиностроение. (in Russian).
7. Longuet-Higgins, M. S. (1957). Statistical Properties of an Isotropic Random Surface. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *250* (975), 157–174.
8. Крагельский, И. В. (1968). Трение и износ. 2-е изд. (Friction and Wear, 2nd ed.) М.: Машиностроение. (in Russian).

sciendo

LATVIAN JOURNAL OF PHYSICS
AND TECHNICAL SCIENCES
2021, N 5

DOI: 10.2478/lpts-2021-0038

FORECASTING ALGORITHM BASED ON TEMPERATURE ERROR PREDICTION USING KALMAN FILTER FOR MANAGEMENT SYSTEM DEVELOPMENT

N. Bogdanovs, R. Belinskis, V. Bistrovs,
E. Petersons, A. Ipatovs

Department of Transport Electronics and Telematics,
Faculty of Electronics and Telecommunications,
Riga Technical University,
LV-1658, Riga, LATVIA

The study offers a new method of collection and processing of meteorological data from the meteorological service based on observations and correction of numerical weather forecast errors using a new prediction algorithm. This algorithm vastly increases the accuracy of the short-term forecast of outdoor air temperature, which is subject to uncertainty due to the stochastic nature of atmospheric processes. Processing of temperature data using Kalman filter provides the decrease in predicted temperature errors. The main setup methods of Kalman filter have been examined. The article also describes the implementation of accuracy improving algorithm of predicted temperature using Python.

Keywords: *Algorithm, forecast, Kalman filter, prediction, temperature.*

1. INTRODUCTION

Temperature forecast data are of importance in different fields, such as agriculture, tourism, hydrology, and construction for heating, ventilation, and air conditioning system operation in the buildings.

Many meteorological services provide information about forecast temperatures for exact place or region. However, this does not mean that temperature forecast is available for each point on the Earth. In some cases, the temperature information is used for the nearest available place.

The main problem from the perspective of meteorological data consists in their

inherent uncertainty due to the stochastic nature of atmospheric processes, imperfect knowledge of initial conditions of weather model and modelling errors. To solve this problem, the authors of the study propose using Kalman filter (KF), which will gather the data from the meteorology service. Temperature information is based on the outputs of special weather models or weather data numeric analysis. It is impossible to forecast the temperature values that would be equal to the observed ones using weather models. Therefore, there is a certain level of deviation from the observed temperature. This will be demonstrated in Section II. Section III will give possible analytical presentation of temperature forecast errors. Section IV defines KF based algorithm for temperature data processing. The results of temperature data processing will be discussed in Section V.

2. RELATED WORKS

Perera discussed the energy saving potential in residential buildings in Norway using the developed mathematical model [1]. Research did not use Kalman filtering, but the author's method obtained a good approximation to model parameters.

One of the widely used techniques for improving numerical weather forecast is model output statistics (MOS). It works well but cannot account for every local effect and does not correct systematic errors from numeric weather prediction (NWP) model [2]-[4].

Another technique for improving weather forecast is based on regression equation development. This algorithm is called PERFECT PROG. The drawback of perfect program is its inability to correct biases of forecast data [5]-[7].

Kalman filter based algorithm is naturally designed to eliminate systematic errors, and it easy integrates forecast data based on NWP model, which considerably simplifies the core of the algorithm. The last leads to a decrease in the data processing time.

Despite a large number of valuable research papers describing different types of weather forecast algorithm, there is no weather forecast system that could be integrated into existing buildings using predictions and ensuring corrections in air temperature forecast within next three hours.

3. TEMPERATURE FORECAST ERRORS

The temperature forecasts for Riga, Latvia, were provided by Eurometeo or Gismeteo. Local temperatures measurements were obtained from meteorological station Riga (University).

Figure 1 presents daily evolution of the measured and forecast (Eurometeo) temperatures in January 2018 for Riga. The difference between the observed and forecast temperatures is well observable. This difference can be interpreted as a temperature forecast error. Distribution of this forecast error is shown in Fig. 2.

Figure 3 presents distribution of the temperature forecast errors for March. The bias and the standard deviation of temperature forecast errors for January, February, and March are shown in Table 1. The fore-

cast errors values are changing randomly within a certain range of temperatures. In addition, there is non-zero bias for error values.

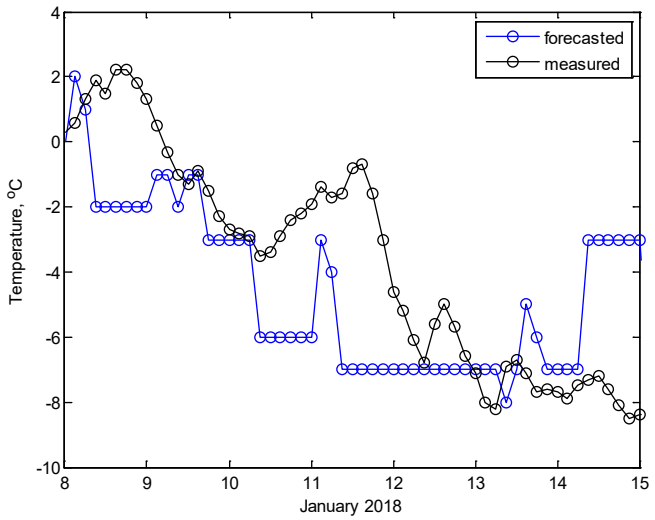


Fig. 1. The temperature observations and forecasts (www.eurometeo.ru) for Riga from 8 to 15 January 2018.

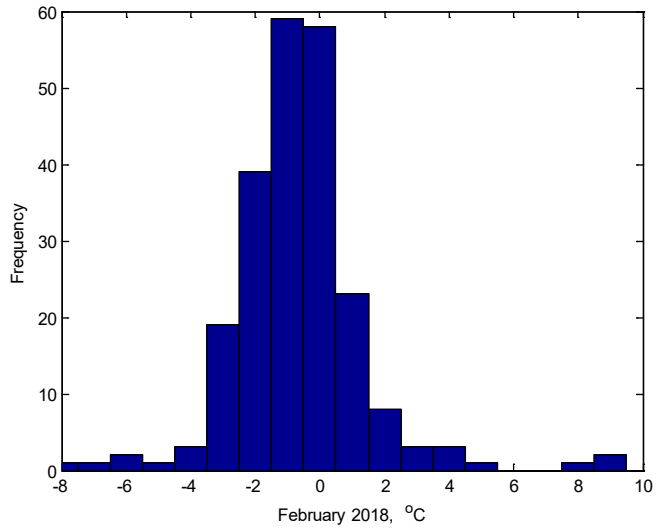


Fig. 2. The distribution of the temperature forecasting error for Riga.

Kalman filter (KF) can be suggested to process forecast temperature data for decreasing a systematic forecast error [1]–[5]. As an output of temperature model (temperature prediction) is available from various meteorological services, there is no need to develop temperature model for KF. This simplifies the algorithm development process and increases the data processing speed. The measured temperatures are available in time series data for the needed period. Thus, the advances in weather forecast processes allow considering the KF algorithm for data processing from another perspective and obtaining valuable results.

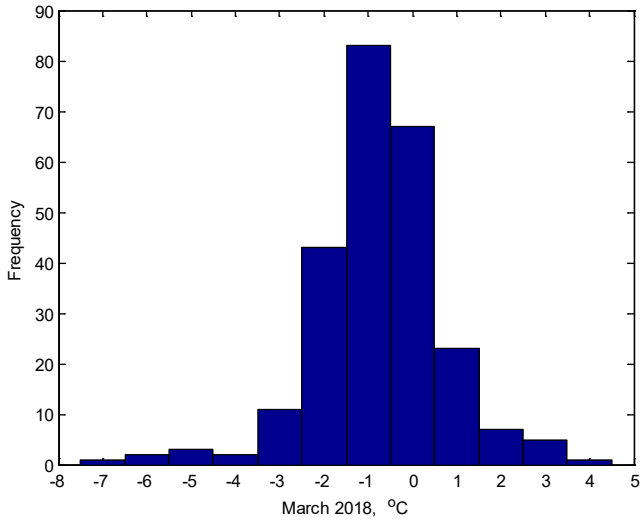


Fig. 3. The distribution of the temperature forecasting error for Riga.

Table 1. Statistical Parameters of the Temperature Forecasting Error

Errors	January	February	March
Bias of the forecast error, °C	0.7520	0.5960	0.6859
Std. deviation of the forecast error, °C	3.2512	1.9958	1.4479

4. ANALYTICAL DEFINITION OF THE FORECAST ERROR

It was shown in [1], [4] that the forecasting error can be represented in function of the forecast temperatures:

$$\varepsilon = \alpha_0 + \alpha_1 T_{m0} + \alpha_2 T_{m0}^2 + \cdots,$$

where α_n – corresponding regression coefficients, ε – a temperature forecast error.

The advantages of Eq. (1) are possible forecast error non-linearity and integration of such a measurement model in the updating stage of linear KF. The regression coefficients are the system states of KF.

5. KALMAN FILTER DEFINITION

The Kalman filter is an estimation algorithm that provides optimal estimates for the system state. The algorithm uses information about deterministic and statistical properties of the system and direct and indirect measurements.

The basic elements of the KF algorithm are system state vector, X_k; error covariance matrix, P_k; system noise covariance matrix, Q_k; system model, Φ_k; measurement vector, Z_k; measurement noise covariance matrix, R_k; and measurement model, H_k.

The system noise covariance matrix describes uncertainty of the true values of the system states. The error covariance matrix describes uncertainty of the state

estimates. The system model describes evolution of the states and error covariance matrix during discrete time step, k, of KF operating. In the present case, there is no knowledge how system states are changing over the time; therefore, the state transition matrix is a simple identity matrix. The system noise covariance matrix will be a diagonal matrix, as it is assumed that there is no correlation between system noises of the states. The values of such a diagonal covariance matrix can be increased to account for possible unknown dynamics of the analysed system. The measurement vector consists of only one element – difference between locally observed or measured temperature and temperature forecast. If the quality of the data measurement is unknown, the diagonal elements of the measurement noise covariance matrix can be increased. Normally, this matrix describes statistical properties of the temperature sensor. In the present case, this property is unknown; therefore, the values were defined empirically. The measurement model describes relation between the system state and the measurement vector. Measurement vector is defined as difference (forecast error) between forecast temperature T_{mo} and observed temperature T_{me}. The polynomial of n degree describes the relation between the forecast error and forecast temperature using regression coefficients α_n [1].

The KF c algorithm has two steps – prediction and updating (or correction). During the prediction step, the algorithm estimates the error covariance matrix and system states. As the state transition matrix is an identity matrix, the predicted states for next time step are equal to the corrected states from the previous time step [7], [8]. During updating steps, the system states are corrected using the observed temperature information. The error covariance matrix is updated as well. The set of the equations of KF is given below:

Updating :

$$\mathbf{K}_k = \mathbf{P}_{k(predicted)} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k(predicted)} \mathbf{H}_k^T + \mathbf{R}_k)^{-1},$$

$$\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_{k(predicted)} + \mathbf{K}_k (\mathbf{Z}_k - \mathbf{H}_k \hat{\mathbf{X}}_{k(predicted)}),$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k(predicted)},$$

Predicting :

$$\hat{\mathbf{X}}_{k+1(predicted)} = \mathbf{\Phi}_k \hat{\mathbf{X}}_k,$$

$$\mathbf{P}_{k+1(predicted)} = \mathbf{\Phi}_k \mathbf{P}_k \mathbf{\Phi}_k^T + \mathbf{Q}_k,$$

where K_k – Kalman gain matrix, P_k – state uncertainty covariance matrix, H_k – measurement sensitivity matrix, R_k – measurement noise covariance matrix, Φ_k – state transition matrix, Q_k – system noise covariance matrix, $\hat{X}_k$ – state estimates.

The corresponding state vector is defined as $X_k=[\alpha_{0,k}\ \alpha_{1,k}\ \alpha_{2,k}]$, where α – a regression coefficient. The degree of polynomial equals two. It was observed that when using higher degrees (three and more) of polynomials, considerable estimation error deviation appears. A similar conclusion was obtained in [1]. Initial values for regression coefficients were set to zero.

Taking into account Eq. (1), the measurement sensitivity matrix was defined as:

$$H_k = [T_{mo,k}\ T_{mo,k}^2]$$

The diagonal values for covariance matrices were defined during data processing trials by KF in order to achieve decrease bias and standard deviation for the temperature forecast error. The corresponding diagonal elements of the covariance matrices were defined within a value range from 0.4 to 1.4 for the measurement noise and from 0.01 to 0.1 for the system noise.

Algorithm steps are shown in Fig. 4. Three-hourly (at 2:00, 5:00, 8:00, 11:00, 14:00, 17:00, 20:00, 23:00) temperature forecast (external data) for Riga was provided by Eurometeo (www.eurometeo.ru) or Gismeteo (www.gismeteo.ru). The corresponding local temperature measure-

ments were obtained from the meteorological station in Riga for the same hours as the forecast. These data were processed by KF and the regression coefficients, α_0, α_1 α_2, were evaluated as a result. The time of the processing step for prediction and updating stages was equal to 3 hours. Correction error (ε) was calculated by the algorithm and added to the temperature forecast value for the next time step. Then, cubic interpolation for corrected temperature data series was made in order to obtain temperature forecast for each hour (output data).

The functionality of this algorithm was also checked when the observed temperature was not available for 1–5 days. During this situation, the algorithm continued to provide forecast at an acceptable quality level (better than the temperature forecast provided by the model of the considered meteorological services).

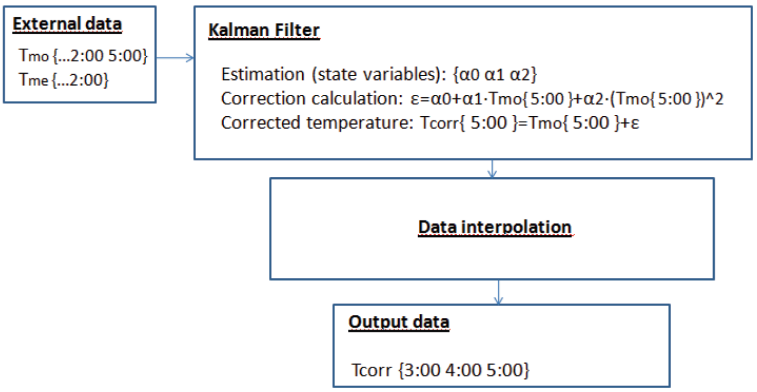


Fig. 4. Algorithm steps.

6. TEMPERATURE DATA PROCESSING RESULTS

Results of data processing, using algorithm described in Section IV, will be discussed in this section. The temperature data were analysed for three months in order to calculate bias and standard deviation.

The evaluation of statistical parameters of corrected forecast error is given in Table 2.

Table 2. Statistical Parameters of Error for Corrected Temperature Forecast (for each 3 h) Using KF

	January		February		March	
	model	KF	model	KF	model	KF
Bias, °C	0.752	0.025	0.5960	0.0007	0.686	0.002
Std. dev., °C	3.251	2.599	1.9958	1.3057	1.448	1.221

Results in Table 2 show a decrease in systematic bias for corrected forecast temperature. Standard deviation slightly decreased as well. The observed decrease in the bias error is more than 100 times comparing with uncorrected forecast. Standard deviation decreases in average by 20 %.

Figure 5 shows the distribution of temperature error after meteorological data processing by KF.

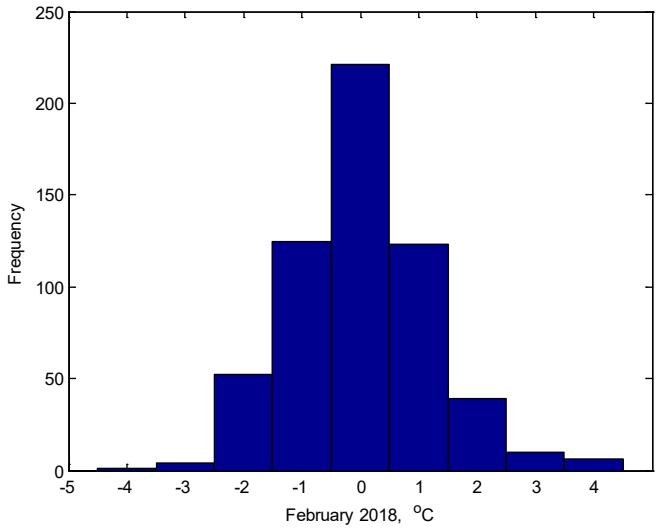


Fig. 5. The distribution of the temperature error after correction using KF for Riga.

Figure 5 shows that error values (difference between the corrected and measured temperature) are concentrated around zero and the shape of histogram is very similar to the normal one. Corrected values by KF follow the measured values more exactly comparing with direct output of temperature model (forecast values). Error distribution (for February 2018) between forecast and measured values is shown in Fig. 2.

Additional experiment was conducted for forecast temperature in summer time. The forecast temperatures using model were taken from the meteorological service gismeteo.ru.

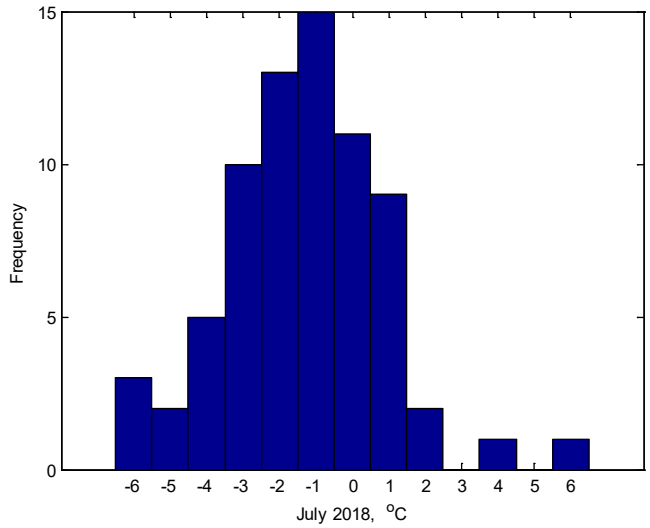


Fig. 6. The distribution of forecast error (www.gismeteo.ru) for Riga from 11 to 19 July 2018.

The temperatures were forecast for the period of 11–19 July 2018 for Riga (Latvia). During this period, the heat wave reached the region of Riga. The distribution of temperature difference between the forecast and measured temperature is shown in Fig. 6. There are significant error components in the range of 1 to 3 degree. This indicates that it was quite complicated to forecast precisely temperature dynamics during this period. One of the reasons can be that the air was warm itself and there was no need for sun to spend energy to maintain it warm. However, sun energy can be spent to increase air temperature at once. Therefore, temperature highly depends (more than usually) on the level of cloudiness. The forecast for cloudiness was quite unstable during this period and in some cases was stable only for next 8 hours.

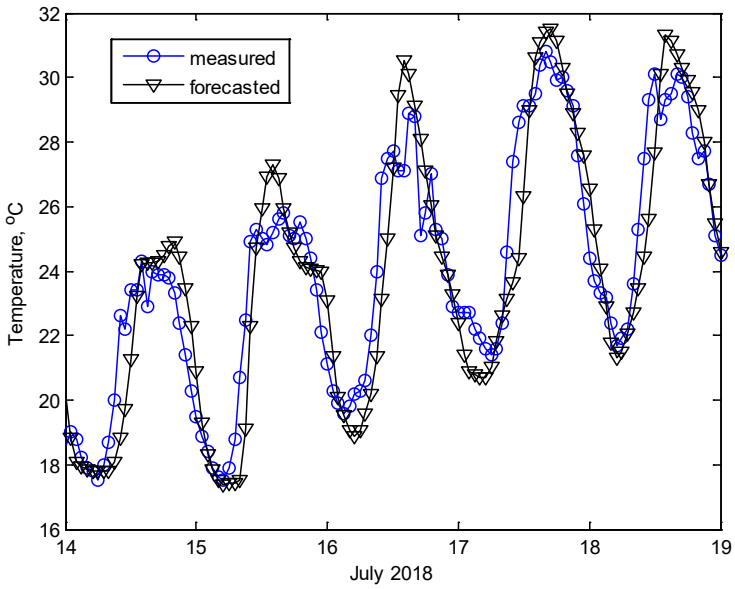


Fig. 7. Measured temperature and its forecast using KF for Riga from 14 to 19 July 2018.

The results of temperature data (the forecast and measurement data) processing are shown in Fig. 7. The temperature data processing algorithm gives the corrected forecast for next three hours based on the observed temperature for the present moment and temperature dynamics from the past period. This temperature dynamics is reflected by values of regression coefficients. Other mode of temperature data processing by KF was checked, when data processing algorithm worked a certain period with availability of the observed temperature data and the algorithm estimated regression coefficients, then the measured temperature was not available and algorithm used the predicted temperature value instead. The results of such processing are shown in Fig. 8.

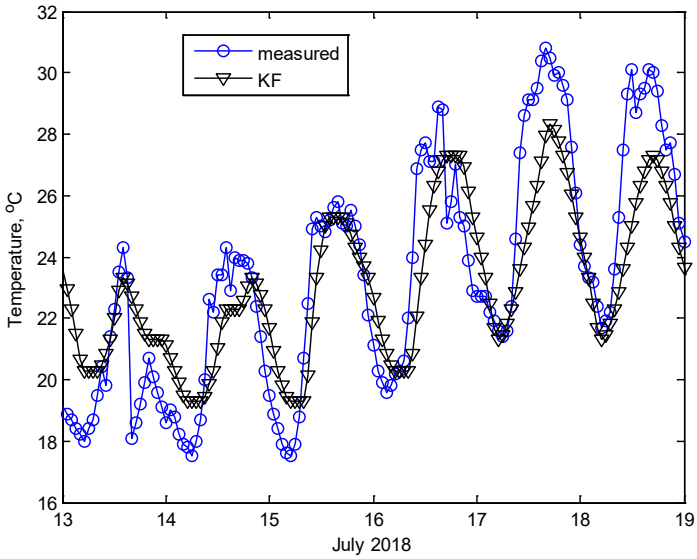


Fig. 8. Measured temperature and its forecast using KF (without the use of the measured temperature data) for Riga from 13 to 19 July 2018.

The measured temperature was available from 11 July, 00-00 till 12 July, 00-00, and temperature was not available from 12-07, 01-00 till 19-07, 00-00. The correction of the forecast was still working but not so good compared with the situation when the observed temperatures were available, but still better than an initial forecast. Figure 8 shows that a forecasting error increases with time. This is logical as KF adjustment became worse, because the measured data were not available for correction.

Additional data processing (by KF) with the use of "past" temperature statistics allows improving precision of temperature forecast. This is especially true when the meteorological service uses an improper model of the temperature forecast for a local area of interest. KF uses real measurement data to provide optimal estimates of the system states [9], [10]. This means that the algorithm can take into account temperature changes due to the natural factors (altitude, type of local area, etc.), and changes in temperatures due to human factors which are difficult to model, but this information can be easily integrated into KF measurements. For example, the temperature sensor (needed for heating system adjustment in the room) is fixed somewhere near the window of the house. In wintertime, the window can be open and local temperature sensor will indicate that it is warmer outside and the heating system will provide less heating, as KF output will react on the obtained sensor measurement by increasing the forecast temperature. This is an advantageous behaviour of such an algorithm.

7. PRACTICAL IMPLEMENTATION OF THE WEATHER PREDICTION ALGORITHM

Great part of general energy consumed by buildings increased interest in optimizing building operation and promoted the development of new approaches to climate

control in premises. The present study uses the algorithm providing correction of air temperature forecast within the next three hours [11], [12].

Figure 9 shows the implementation of the advanced algorithm of temperature prediction using Python.

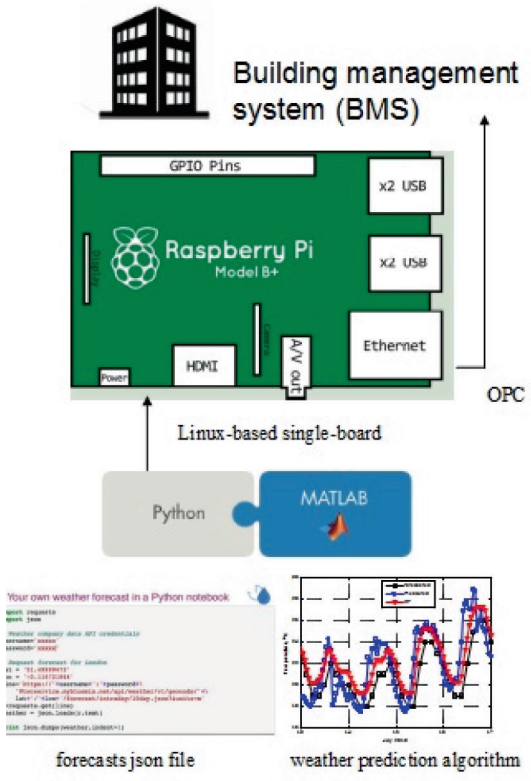


Fig. 9. Control system developed by Python using a weather prediction algorithm.

Control system developed by Python can regulate the curve of building heating based on the external temperature forecast. The idea of the developed management logic consists in adjusting the heating curve depending on a future state [13]-[15]. The research describes the shifted profile of external temperature, and the changed curve of heating is calculated according to a set of parameter settings for the whole heating season. Weather forecast data are based on Eurometeo (www.eurometeo.ru) or Gismeteo (www.gismeteo.ru) databases considering the typical meteorological year. There were a number of tests performed to select the best period for the shift, and, finally, the period of one-hour standby was used. The results during the modelling were obtained in Matlab & Python [16], [17].

We managed to integrate it into the building old automatic equipment using OPC (Open Platform Communications) standard. We used Matlab machine learning toolbox and weather forecast. OPC is the interoperability standard for secure and reliable exchange of data in an industrial automation area [17]. By analysing next day weather forecast and the database with

data from wireless temperature sensors, a self-learning algorithm was created that allowed adjusting automatic equipment of the building.

This program analysis of Matlab helps define energy efficiency of buildings according to the European standard EN 15603. This standard sets rules for energy consumption assessment in the building and methods, which should be used to determine nominal parameters of building engineering systems.

8. CONCLUSIONS

The article proposed a new method of gathering and processing the meteorological data from the meteorological service. This algorithm can significantly increase the accuracy of short-term forecast of external air temperature. The considered algorithm provides temperature forecast correction for the next three hours. The temperature data processing by KF provides bias (decrease100 times) and standard deviation (decrease by 20 %) decreasing forecasted temperature error. The design of Kalman filter can be simplified owing to data availability from different temperature forecast models. Therefore, it can provide the best temperature for enzymatic detection to ensure the accuracy of results. Such an algorithm takes into account unmodelled temperature deviations based on a series of measured temperatures. It was shown that useful temperature correction could be provided when the observed temperature was not available. One of the possible improvements of the analysed algorithm can be automatic tuning implementation.

9. FUTURE RESEARCH

In future, it is planned to address all the above-mentioned problems within the framework of research by providing the following solutions:

Several weathers forecast models used at the prediction stage of KF could potentially decrease a temperature forecast error. Introducing adaptive tuning of KF noise matrix coefficients makes set up of data processing faster. As weather is a nonlinear dynamic system, another type of KF – unscented Kalman filter – can be applied for better tackling with system nonlinearity.

Next research of our control algorithm will include human occupancy in the building. New algorithm using machine learning and building new data will upgrade our model results.

REFERENCES

1. Perera, D.W., & Skeie, N. (2016). Comparison of Space Heating Energy Consumption of Residential Buildings Based on Traditional and Model-Based Techniques. *Buildings, 7* (2), 27.
2. Galanis, G., Louka, P., Katsafados, P., Pytharoulis, I., & Kallos, G. (2006). Applications of Kalman Filters Based on Non-Linear Functions to Numerical Weather Predictions. *Ann. Geophys*, 24, 2451–2460.

3. Anadranistakis, M., Lagouvardos, K., Kotroni, V., & Skouras, K. (2002). Combination of Kalman Filter and an Empirical Method for the Correction of Near-Surface Temperature Forecasts: Application over Greece. *Geophysical Research Letters, 29* (16), 231–234.
4. Libonati, R., Trigo, I., & DaCamara, C. (2008). Correction of 2 m-Temperature Forecasts Using Kalman Filtering Technique. *Atmospheric Research, 87,* 183–197.
5. Homleid, M. (1995). Diurnal Corrections of Short-Term Surface Temperature Forecasts Using the Kalman Filter. *Weather and Forecasting, 10* (4), 689–707.
6. Brunet, N., Verret, R., & Yacowar, N. (1988). An Objective Comparison of Model Output Statistics and "Perfect Prog Systems" in Producing Numerical Weather Element Forecasts. *Weather and Forecasting, 3* (4), 273–281.
7. Bo, P. (2004). A Statistical Method for Forecasting Extreme Daily Temperatures Using ECMWF 2-m Temperatures and Ground Station Measurements. *Meteorol*, 245–251.
8. Chai, B., Tushar, W., Hassan, N. U., Yuen, C., & Yang, Z. (2016). Managing energy consumption in buildings through offline and online control of HVAC systems. In *IEEE Region 10 Conference (TENCON),* (pp. 3368–3373), 22–25 November 2016, Singapore.
9. Hadwan, H., & Reddy, P. (2016). Smart Home Control by Using Raspberry Pi & Arduino. *International Journal of Advanced Research in Computer and Communication Engineering, 5* (4), 283–288.
10. Lynggaard, P. (2014). *Artificial Intelligence and Internet of Things in a "Smart Home" Context: A Distributed System Architecture*. PhD Thesis. Press / Media.
11. Gomez Ortega, L., Han, L., Whittacker, N., & Bowring, N. (2015). A machine-learning based approach to model user occupancy and activity patterns for energy saving in buildings. In *Science and Information Conference*, (pp. 474–482), 10–12 June 2015, London, UK.
12. Zhang, Y., & Handy, I. (2007). Short-Term Prediction of Weather Parameters Using Online Weather Forecasts. *Building Simulation 2007,* 1411 – 1417.
13. Mohinder, S., Grewal, & Angus, P. (2008). *Kalman filtering: Theory and practice using MATLAB* (3rd ed.). New Jersey: John Wiley & Sons.
14. Rolando, D., Madani, H., Braida, G., Tomasetig, R., & Mohammadi, Z. (2017). Heat pump system control: The potential improvement based on perfect prediction of weather forecast and user occupancy. In *12th IEA Heat Pump Conference 2017,* (pp. 1–9), 15–18 May 2017, Rotterdam, Netherlands.
15. Bogdanovs, N., Bistrovs, V., Ipatovs, A., & Beļinskis, R. (2018). Weather prediction algorithm based on historical data using Kalman filter. In *2018 Advances in Wireless and Optical Communications (RTUWO)*, (pp. 94–99), 15–16 November 2018, Riga, Latvia.
16. Siroky, J., Oldewurtel, F., Cigler, J., & Prívara, S. (2011). Experimental Analysis of Model Predictive Control for an Energy Efficient Building Heating System. *Applied Energy, 88* (9), 3079 –3087.
17. Brown, R.G., & Hwang, P.Y.C. (1997). *Introduction to random signals and applied Kalman filtering.* Wiley.

sciendo

LATVIAN JOURNAL OF PHYSICS
AND TECHNICAL SCIENCES
2021, N 5

DOI: 10.2478/lpts-2021-0039

DAMAGE DETECTION OF THE ROD IN THE CROSSFLOW USING SURROGATE-BASED MODELLING

S. Upnere[1,2,*], J. Auzins[1]

[1]Institute of Mechanics and Mechanical Engineering,
Riga Technical University, 6B Kipsalas Str., Riga, LATVIA
[2]Ventspils University of Applied Sciences,
101 Inzenieru Str., Ventspils, LATVIA
*e-mail: upnere@protonmail.com

An effective and accurate methodology is developed to create an inverse surrogate model for the mass reduction analysis of the rod in the rod bundle inserted in the crossflow. The performance of two surrogate modelling approaches has been evaluated. These models are the Response Surface Method and Legendre polynomial approximations. The relationship between dominant frequencies, support stiffness and rod mass derived from Computational Fluid Dynamics simulations is used as input data for approximations. The selection of sample points is implemented with a new type of orthogonal design. The results have shown that the proposed methodology can reliably replace the finite volume model and drastically reduce computational time.

Keywords: *CFD, inverse model, rod bundle, surrogate model.*

1. INTRODUCTION

For the safe operation of engineering equipment, it is necessary to continuously monitor whether the system does not exceed its operational limits and does not compromise its structural integrity. Relatively frequently it is not possible to measure changes in the internal elements of the system directly during the operation of the equipment, for example, in heat exchangers or nuclear reactor assembly. For complex systems, their behaviour as a function of the input parameters is generally not known. In such situations, physical and numerical experiments can often be the only way to

determine the relevant relationships. Nevertheless, full-model simulations would require long calculations from several hours to several days or weeks, even using high-performance computing resources. That approach is not applicable to the monitoring of the system during its operation. Therefore, to reduce computing time, simplified models – surrogate models or metamodels – are used that approximate the original model. Adequate mathematical models can be used to detect as early as possible changes in the typical behaviour of the system that could indicate its malfunction and structural integrity problems.

This study focuses on heat exchangers consisting of a bundle of rods placed in a highly turbulent crossflow. Flow induces rod vibrations with a typical spectrum that can be used to indirectly detect changes in system parameters. The decrease in the mass of the rod may be one of the parameters indicating that damage has occurred in the installation. Therefore, Computational Fluid Dynamics (CFD) simulations and surrogate modelling approaches are used to find the relationship between the mass of the rod and the dominant frequencies.

In the literature, various types of engineering problems are solved through surrogate models such as design optimization, uncertainty quantification, parameter identification, sensitivity analysis or inverse problems. Available surrogate models include, e.g., Response Surface Method (RSM), Kriging (or Gaussian process modelling), radial basis functions and artificial neural networks. RSM is predicated on the assumption that many physical systems are smooth and continuous and can be well approximated, at least in the region of interest, by low-order polynomials [1]. As follows from [2], the polynomial approximation is not applicable to highly non-linear functions, but in cases of low non-linearity and low dimensions, it is an effective approach. The Kriging method is of most use when the true function is particularly computationally intensive, e.g., a CFD-based calculation [3]. As an example, we can mention the minimization of vortex-induced vibrations using Kriging described by Filho et al. [4]. Banyay et al. [5], [6] used Kriging surrogate for global sensitivity analysis of flow-induced vibrations. In the present paper, low-order (less than 4th order) polynomials and Legendre polynomial were investigated. Two-factor forward and inverse surrogate models were created.

The training data set for the surrogate model development is generated by exercising complex high accuracy models such as finite volume or finite element models. The selection of sample points can be implemented through the design of experiments. The design of experiments is very important for a high-quality surrogate model. There are several metrics to determine the adequacy and goodness of fit of the model, such as a root mean square error, cross-validation error, T-tests or Pearson's χ^2 test, etc. As mentioned by Greenwood and Nikulin [7], when the amount of data is small, no test is very good, and when there are a lot of high-quality data almost any test, properly designed and applied, will give good results. However, when addressing problems with high computational costs, an effective sampling plan must be sought, which means a minimum number of points that provide a surrogate model with good precision. Techniques that are commonly considered are Latin Hypercube Sampling, Central Composite Design (CCD), orthogonal arrays, factorial designs. In this study, a new type of orthogonal design [8] is used. Based on the obtained sample points, the selection of initial conditions for the full-order model is carried out.

The paper is structured as follows. In

Section 2, the CFD approach applied to obtain the results of the full model is formulated. Section 3 describes the types of employed surrogate models, the sampling method and the model accuracy metrics. Verification results are presented in Section 4. Finally, concluding remarks are given in Section 5.

2. HIGH-FIDELITY COMPUTATION

The full model uses the configuration of the rod bundle described in Upnere et al. [9], in which laboratory experiments were carried out to examine the vibrations of single flexible-mounted rod caused by the crossflow in the rigid rod array. The configuration has a triangular array with pitch-to-diameter ratio, P/d = 1.1. The rod diameter is 8 mm and it is made of steel (in the laboratory experiment).

The open-source CFD tool OpenFOAM 2.4.x was used to solve differential equations of mass and momentum. Unsteady Reynolds-Averaged Navier–Stokes (URANS) equations are solved using the k-ω SST turbulence model to ensure the closure of the equation system. A lowReynolds turbulence model approach was used. The 2D problem was addressed to simplify numerical calculations.

2.1 Description of the Numerical Model

The investigation of an appropriate computational domain was achieved through a decoupling approach. The number of cylinders was reduced by columns in the in-flow direction and by rows in the crossflow direction. Several computational domains were analysed to find an optimal case given both the accuracy of the results and the cost of computing. The columns and rows of the cylinder arrays were decreased step by step with a decrement 1. Other conditions remained the same. The drag force F_x was used to compare the impact of array rows and columns.

The largest analysed domain was a 4×5 cylinder array, see the domain A in Fig. 1. Decreasing rows from four to two changes the force F_x by less than 1 %. The boundary condition at the top and bottom is simplified as a symmetry condition. The reduction of columns in the direction of flow by removing one column behind the test cylinder causes a non-significant change in the calculated force. In contrast, the reduction of one column in the direction of upstream causes a change of approximately 4 %. Because of the above, the domain B (see Fig. 1) can be considered an optimal computational domain for the analysed Reynolds numbers and array configuration.

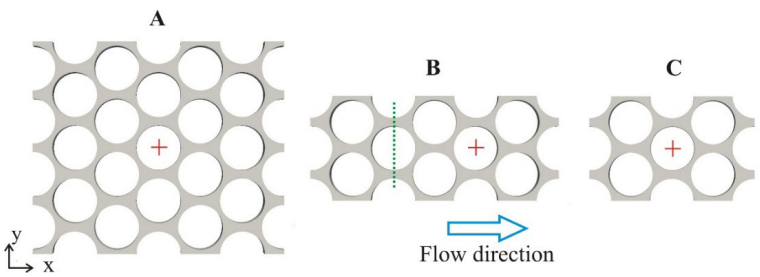


Fig. 1. Analysed computational domains. The centre of the test cylinder is marked with the red cross.

As RANS-type calculations are performed that are time-averaged, it is possible to determine a typical distribution of flow between two adjacent cylinders, see the green dotted line in the domain B. Using the found flow profile as an inflow boundary condition, the computational domain can be reduced to a 2×2 array of cylinders, see the domain C in Fig. 1. This approach is useful because the developed surrogate models require a set of calculations that change only the mass and support stiffness of the test cylinder and not the characteristic of the inlet flow. Operating pressure p = 0 Pa was defined on the outflow boundary.

Different mesh refinements were studied to avoid the influence of the cell size on the numerical simulations. As a low-Reynolds turbulence model was used, particular attention was paid to the correct description of the boundary layer around cylinders. The final version of the mesh contains 476 000 cell elements. Around cylinders, it contains 25 layers with an expansion ratio of 1.05, providing y^+ less than 1.

Based on studies available in the literature, it is known that the largest oscillations are in a direction perpendicular to the flow and therefore the movement of the test cylinder is only allowed in the transverse direction. The rotation of the test cylinder is also not allowed. The movement of the test cylinder in crossflow was studied by mounting it as a massspring system, where m is the mass of the oscillating part and k is the stiffness coefficient of the mounting.

2.2 Verification and Validation of the Model

According to Roache [10], the quantification of uncertainty in numerical simulations contains the three most important items: verification of codes, verification of calculations and validation. Since standard OpenFOAM solvers are used in the modelling process, the verification of the code is not carried out in this case. Solution verification is carried out using flow simulation around a single cylinder at Re = 11000 and comparing integral parameters with data available in the literature. Table 1 shows a summary of the drag coefficient, $C_d = 2F_x/(u^2S)$, where u is flow velocity, S is an area that interacts with the flow and Strouhal number, St, from the various studies.

Table 1. Comparison of the Modelled and Experimental Results of the Drag Coefficient C_d and Strouhal Number St

The study	C_d	St
Khan et al. [11], RANS, 2D	1.150	0.201
Khan et al. [11], RANS, 3D	1.210	0.203
Dong et al. [12], DNS	1.208	0.209
Wornom et al. [13], LES	1.22	0.20
Nguyen and Nguyen [14], DES	1.133	0.200
Gopalkrishnan [15], Exp	1.186	0.193
Norberg [16], Exp	-	0.202
Present study, URANS, 2D	1.19	0.21

The analysis of the test case allows selecting the optimal turbulence model, boundary conditions and computational cell size.

The model validation was implemented by comparing the numerical results with the experimental drag force values, C_d, the natural frequency of the test rod in the flow, f_n and the flow-induced frequency f_t. The results obtained are summarised in Table 2.

Table 2. Validation of the Full Numerical Model

The study	f_n	f_t	C_d
Experiment (Upnere *et al.*, 2020)	58.50	27.30	0.346
CFD, 2D	59.32	28.66	0.359
CFD, 3D	-	28.56	0.352

As shown in Table 2, the difference between the results of CFD and the experimental measurements obtained is less than 5 %.

2.3 Dominant Frequencies

The dominant frequencies were determined using the movement of the test cylinder and the calculated lift coefficient. The fast Fourier transform (FFT) and inverse FFT method were applied to time series obtained from numerical simulations. The FFT approach was implemented in Matlab.

3. SURROGATE MODELLING

3.1 Design of Experiments

The selection of sampling points has a major effect on building a reliable surrogate model. Space-filling designs become a primary consideration for the design of computer simulation experiments [1]. Space-filling design allows selecting those sample points that are distributed within the entire domain.

In this study, the new class of experimental designs described in [8] is used. Unlike classical CCD designs, in this case, all design points are within the research domain. Orthogonal designs of this type are particularly suitable for use with orthogonal Legendre polynomials detailed in the next subsection. This approach allows avoiding problems related to the optimal selection of significant terms, for example, in cases where the number of points is not much larger than the total number of possible terms. The following constraints were used to develop the experimental design [8]: 1) all experimental points were located in the unit cube $[-1, 1]^l$; 2) designs had central symmetry, axial symmetry and 90-degree rotation symmetry properties; 3) designs were invariant to the permutation of input variables; 4) designs gave replications of centre point; 5) designs were orthogonal; 6) designs were D-optimal.

Figure 2 shows the 13-point orthogonal design used in the study to create a surrogate model with mass m = [0.2765; 0.3253] and stiffness coefficient k = [15725; 18500] as factors.

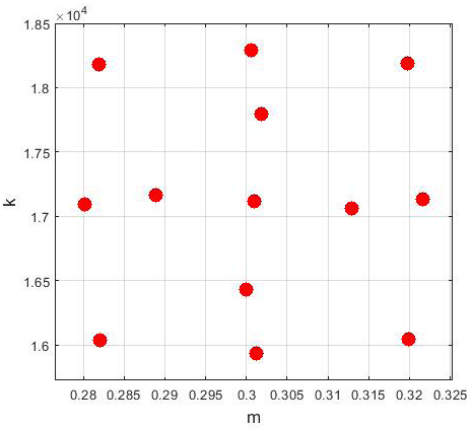


Fig. 2. The two factors (mass m and stiffness coefficient k) orthogonal design with 13 points.

3.2 Response Surface Method

Polynomial approximations produce surrogate models using low-order polynomials in a relatively small region of parameter space [2]. If the response function to be studied is expected to be linear, the first-order polynomial approximation can be used, defined as [17]:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon, \tag{1}$$

where $\hat{y}$ is the system response, x_i, $i = 0, 1, \ldots, n$, are independent variables, the parameters β_i are called the regression coefficients, ε represents the noise or error observed in the response y assumed to be independent and normally distributed with mean zero and constant variance σ^2.

If the system under investigation is nonlinear, a higher-order polynomial should be used. For example, in the general case, a second-order polynomial regression model with n simulation inputs is as shown below:

$$\hat{y} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} {x_i}^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \varepsilon. \tag{2}$$

Regression coefficients β in Eq. (2) are estimated applying the least-squares criterion

$$\hat{y} = (X'X)^{-1} X'w, \tag{3}$$

where w denotes the n-dimensional vector with the simulation outputs. The least-squares estimator $\hat{\beta}$ exists if and only if the inverse of $X'X$ exists; e.g., $\hat{\beta}$ exists if X is orthogonal [18]. The method of the least squares returns the vector of the regression model parameters, which minimizes the sum of squared errors between the predictions of the regression model and the values of the true function.

3.3 Legendre Polynomial Approximation

The Legendre polynomials $P_i(x)$ are defined over $x \in [-1; 1]$ by the recurrence relation as [19]:

$$\begin{cases} P_0(x) = 1 \\ P_1(x) = x \\ (i+1)P_{i+1}(x) = (2i+1)xP(x) - iP_{i-1}(x), i - 1,2, \ldots \end{cases} \tag{4}$$

The classic and discrete Legendre polynomials are orthogonal. In this study, orthonormal Legendre polynomials are used, which means that all terms must be multiplied by $\sqrt{\frac{2i+1}{2}}$. For onedimensional problem the first 4 orthonormal polynomials are:

$$\begin{cases} P_0(x)=\sqrt{\frac{1}{2}} \\ P_1(x)=\sqrt{\frac{3}{2}}x \\ P_2(x)=\sqrt{\frac{5}{2}}(3x^2-1)/2 \\ P_3(x)=\sqrt{\frac{7}{2}}(5x^3-3x)/2 \end{cases}. \quad (5)$$

The multivariable polynomial terms are the product of single variable polynomials and they are orthonormal. The property of orthonormality allows for optimal selection of significant terms in the approximation polynomial function [8]. This is useful in cases where multivariate high order polynomials are used since in such situations the number of terms is rapidly increased and makes the application more difficult. The cross-validation criterion is used to decide whether to eliminate or keep the term with a small Euclid norm [8]:

$$\|P_i(x)\| = \sqrt{\int_{-1}^{1} P_i^2(x)dx} = \sqrt{\frac{2}{2i+1}}. \quad (6)$$

In surrogate models, assessment methods should be applied to check if the developed models are adequate. The selection of approximations here is based on strategies described in the next subsection.

3.4 Surrogate Model Adequacy and Fitting

Myers et al. [1] define two steps for checking the response surface: 1) to examine the fitted model to ensure that it provides an adequate approximation to the true system and 2) to verify that none of the least-squares regression assumptions is violated.

One of the techniques for checking model adequacy is residual analysis. In the context of surrogate modelling, the residual is defined as

$$e_i = y_i - \hat{y}_i, i = 1,2,\ldots,n. \quad (7)$$

In the plot of residuals e_i versus the predicted response $\hat{y}_i$, the residuals should scatter randomly, suggesting that the variance of the original observations is constant for all values of y.

To estimate the generalization error, leave-k-out cross-validations commonly used. The cross-validation error can be used for surrogate model parameter estimation, model selection and validation when it is too costly to employ a separate validation data set [3]. Based on the study of Meckesheimer et al. [20], the value of $k = 1$ is recommended for providing a prediction error estimate for low-order polynomials, and for Kriging surrogate models it is recommended to select k as a function of the fitting design size, for example, $k = 0.1N$ or $k = \sqrt{N}$. In this study, the relative cross-validation error is expressed as:

$$\sigma_{cr\%} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_{-i}-y_i)^2}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}} 100\ \%, \quad (8)$$

where $\hat{y}_{-i}$ denotes the prediction of the response using the surrogate model created without the point i (i=1, 2, …, n) and $\bar{y}$ is the average value of y.

For regression models, R^2-type metrics (R^2 is the coefficient of determination) are the commonly used goodness-of-fit measure. An adjusted coefficient of determination R^2_{adj} is used to avoid the fact that adding a variable to the model always increases the value of R^2. Pearson's χ^2 test is another commonly used metric. The χ^2 test is a goodness-of-fit test to see whether the data are in concordance with the null hypothesis H_0 [7]. Pearson's χ^2 test indicates whether the frequency distribution of specific events observed in the sample corresponds to a theoretical distribution.

4. RESULTS

CFD simulations are used to obtain the response values in the sampling points. The values from the high-fidelity model are applied to construct an approximation model that would indirectly detect a change in the mass of the rod.

4.1 Forward Model

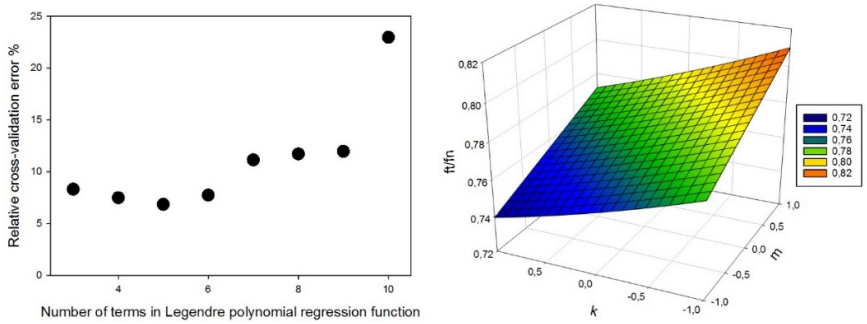
Based on the results of the laboratory experiments described in Upnere et al. [9], the frequency ratio between the natural frequency of the rod in the flow f_n and the flow-induced frequency f_t is selected as a response for the surrogate modelling.

As the design of experiments (see Fig. 2) is orthogonal, Legendre polynomials can be used as an approximation. The transformation from physical units to the unit cube $[-1; 1]^2$ is as follows:

$$\begin{aligned} x_1 &= 40.983896062695\mathrm{m} - 12.332054325265, \\ x_2 &= 0.000720720720720\mathrm{7k} - 12.333333333333, \end{aligned} \tag{9}$$

where $m = [0.2765, 0.3253]$ and $k = [15725, 18500]$.

Analysis of the significance of polynomial terms based on Euclid norm in Eq. (6) shows that the coefficient per term does not depend on the number of terms used for approximation. This is due to the orthogonality of the design of experiments. From the relative cross-validation error depending on the number of terms of the Legendre polynomials presented in Fig. 3a, it follows that half of the polynomial terms can be discarded without affecting the outcome.



(a) (b)

Fig. 3. Legendre polynomials: (a) the relative cross-validation error depending on the number of orthogonal terms; (b) response surface of 5-term Legendre polynomials in the unit cube.

The response surface of the 5-terms polynomial model in the unit cube is shown in Fig. 3b and the corresponding Legendre polynomial can be described as follows:

$$\begin{aligned} \hat{y}(x_1, x_2) &= 1.544039 P_0 P_0 + 0.020323 P_0 P_1(x_1) - 0.020521 P_0 P_1(x_2) + \\ & 0.001028 P_0 P_2(x_2) - 0.000967 P_1(x_1) P_1(x_2). \end{aligned} \tag{10}$$

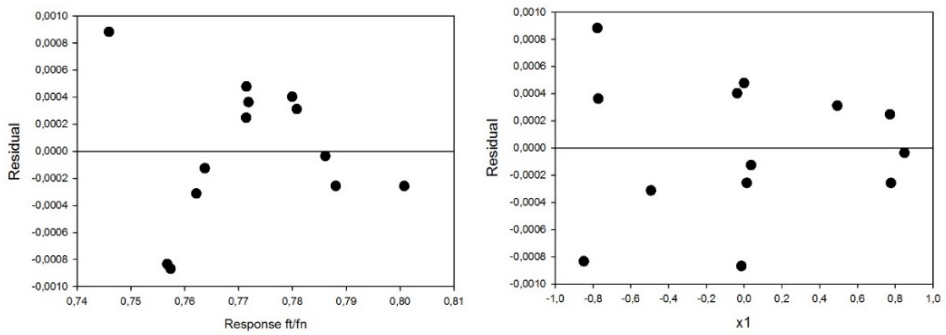
In addition to Legendre polynomials, the first, second and third-order polynomials are investigated. The comparative results of the model quality metrics are summarised in Table 3.

Table 3. Approximation Quality Indicators of the Forward Model

Approximation	$\sigma_{cr\%}$	R^2_{adj}
First-order	8.31	0.996
Second-order	8.16	0.998
Third-order	22.96	0.998
Legendre	6.85	0.998

Table 3 shows that there is a small difference between the first- and second-order polynomial approximation which gives 8.31 % and 8.16 % cross-validation error, respectively. From Fig. 3a and Table 3 it follows that the lowest error – 6.85 % is obtained by 5-term Legendre polynomials.

The adequacy of the model is also confirmed by plots of residuals that are randomly scattered (see Fig. 4). The residual versus response f_t/f_n is shown in Fig. 4a. The residual values are not correlated with factors x_1 and x_2. The example of the residual versus input x_1 is presented in Fig. 4b.



(a) (b)

Fig. 4. The residual and regression line: (a) the residual values vs. response; (b) the residual vs. factor x_1.

4.2 Inverse Model

The inverse surrogate model for predicting changes of the oscillating cylinder mass was created using f_t/f_n and k as factors. The formulas for transformation to unit cube for Legendre polynomials are:

$$x_1 = 37.2562667089699 f_t/f_n - 28{:}8236008539,$$
$$x_2 = 0.0008489240339426\mathrm{k} - 14{:}52721253084, \quad (11)$$

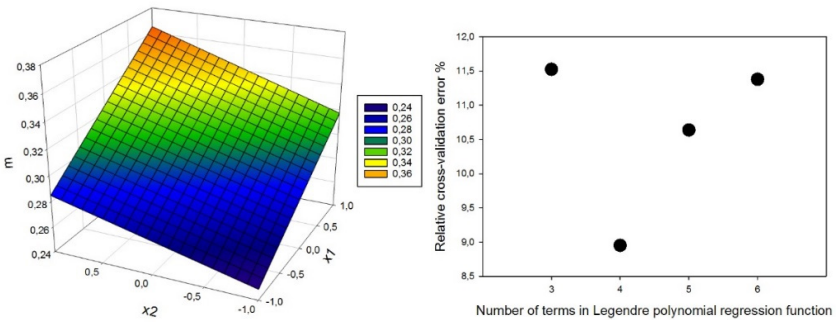
where f_t/f_n = [0.746816665, 0.800498909] and k = [15934.5382978725, 18290.4617021275].

As the design of experiments is not orthogonal, the coefficients for the terms of the Legendre polynomials vary depending on the number of terms. Therefore, the use of Legendre polynomials is not as effective as in the forward model. The comparative results of the model quality metrics are summarised in Table 4.

Table 4. Approximation Quality Indicators of the Inverse Model

Approximation	$\sigma_{cr\%}$	R^2_{adj}
First-order	11.52	0.992
Second-order	11.37	0.996
Third-order	36.41	0.995
Legendre	8.95	0.997

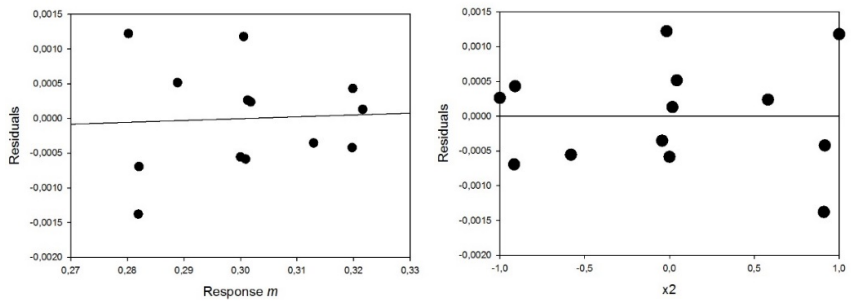
As can be seen from Table 4, there is only one method of approximation, for which the crossvalidation error is less than 10 %, the 4-term Legendre polynomials. Therefore, the Legendre polynomials were selected for building the surrogate model. Figure 5 presents response surfaces of 4-term Legendre polynomials and the relative cross-validation error depending on the number of terms.



(a) (b)

Fig. 5. Legendre polynomials of inverse surrogate model: (a) the response surface designed using 4 terms; (b) the relative cross-validation error vs. terms of the polynomial.

The residual graph shows the absence of correlation between the response and residual values. Examples of residuals are shown in Fig. 6.



(a) (b)

Fig. 6. Residual values and regression line of the inverse surrogate model: (a) the residual versus predicted response; (b) residual values versus input x_2.

The obtained model equation using 4-term Legendre polynomials is as follows:

$$m(f_t/f_n) = -0.11026 f_t/f_n - 4.97899 \cdot 10^{-5} k + 8.74349 \cdot 10^{-5} f_t/f_n \cdot k + 0.083652. \quad (12)$$

The inverse model verification was implemented by selecting specific points and comparing the results of the model and numerical experiments at these points. The comparison graph is shown in Fig. 7. The confidence interval *CI* is determined to measure the uncertainty around the estimated values:

$$CI = \bar{y} \pm z \cdot \frac{\sigma}{\sqrt{n}}, \tag{13}$$

where σ is standard deviation and for 95 % *CI* z = 1.96. The upper and the lower confidence bound can be seen in Fig. 7 with a dashed blue line.

From Fig. 7, it can be concluded that the resulting surrogate model can predict the behaviour of the analysed model system with high precision – when input factor values are changed within 15 %, the difference is less than 1.5 %.

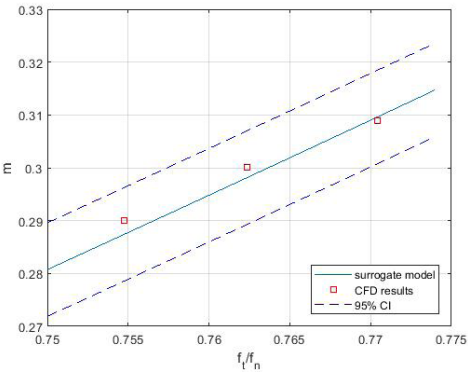


Fig. 7. Verification of the inverse surrogate model if the stiffness coefficient is 17600 N/m.

5. CONCLUSIONS

A methodology was developed for creating a surrogate model for the mass reduction analysis of the rod in the rod bundle inserted in the crossflow. The interaction between mass and stiffness as factors and frequency ratio as the response was applied to build the forward model. The inverse surrogate model was created using mass as a function of stiffness and frequency ratio.

The new third-order orthogonal design of experiments was used to define sample points. Orthogonality of the experimental design allowed for simple elimination of insignificant terms in Legendre polynomial approximations.

Four methods of approximation were investigated: first-, second- and third-order polynomials and Legendre polynomials. The most appropriate approximation for both the direct and inverse models for the problem under study was obtained using Legendre polynomials. The corresponding cross-validation errors were 6.85 % and 8.95 %. Verification of the inverse model showed that the difference between the results of the CFD calculations and the predicted values of the surrogate model was less than 1.5 %.

The analysis carried out showed that the developed method of approximation and the obtained factors-response relationships had a good fit with numerical experiments. The results also demonstrated that the surrogate model could reliably replace the finite volume simulations and drastically reduce computational time. Therefore, the proposed methodology is useful in the monitoring of cooling systems as well as in the design process of such systems.

ACKNOWLEDGEMENTS

The research has been partially conducted with the financial support of the Latvian Council of Science project "Creation of Design of Experiments and Metamodeling Methods for Optimization of Dynamics of Multibody 3D Systems Interacting with Bulk Solids and Fluids" (lzp-2018/2-0281).

REFERENCES

1. Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). *Response surface methodology: Process and product optimization using designed experiments.* New York: John Wiley & Sons.
2. Khaledi, K., Miro, S., Konig, M., & Schanz, T. (2014). Robust and Reliable Metamodels for Mechanized Tunnel Simulations. *Computers and Geotechnics*, *61*, 1–12.
3. Forrester, A., & Keane, A. (2009). Recent Advances in Surrogate-Based Optimization. *Progress in Aerospace Science*, *45* (1–3), 50–79.
4. Filho, U., Antunes, A., Bastos, S., & Lyra, P. (2015). Minimization of Vortex Induced Vibrations Using Surrogate Based Optimization. *Structural and Multidisciplinary Optimization*, *52* (4), 717–735.
5. Banyay, G. A., Shields, M. D., & Brigham, J. C. (2019). Efficient Global Sensitivity Analysis for Flow-Induced Vibration of a Nuclear Reactor Assembly Using Kriging Surrogates. *Nuclear Engineering and Design*, *341*, 1–15.
6. Banyay, G. A., Shields, M. D., & Brigham, J. C. (2020). Efficient Global Sensitivity Analysis of Structural Vibration for a Nuclear Reactor System Subject to Nonstationary Loading. *Nuclear Engineering and Design*, *361*, 1–14.
7. Greenwood, P. E., & Nikulin, M. S. (1996). *A guide to chi-squared testing*. New York: John Wiley & Sons.
8. Auzins, J. (2014). High order orthogonal designs of experiments for metamodeling, identification and optimization of mechanical systems, in Oñate, E., Oliver, J. & Huerta, A. (eds.), *11th World Congress on Computational Mechanics, 5th European Conference on Computational Mechanics, and 6th European Conference on Computational Fluid Dynamics*, 3190–3201.
9. Upnere, S., Jekabsons, N., Dementjevs, S., & Wohlmuther, M. (2020). Effects of Variable Parameters on the Behaviour of the Single Flexibly-Mounted Rod in a Closely-Packed Array. *Journal of Vibroengineering*, *22*, 1–20.
10. Roache, P. J. (1998). *Verification and validation in computational science and engineering*. New Mexico: Hermosa Publishers.
11. Khan, N., Ibrahim, Z., Nguyen, L., Javed, M., & Jameel, M. (2017). Numerical Investigation of the Vortex-Induced Vibration of an Elastically Mounted Circular Cylinder at High Reynolds Number ($Re=10^4$) and Low Mass Ratio Using the RANS Code. *PLoS ONE*, *12* (10), 1–17.
12. Dong, S., & Karniadakis, G. (2005). DNS of Flow Past a Stationary and Oscillating Cylinder at Re = 10000. *Journal of Fluids and Structures*, *20* (4), 519–531.
13. Wornom, S., Ouvrard, H., Salvetti, M., Koobus, B., & Dervieux, A. (2011). Variational Multiscale Large-Eddy Simulations of the Flow past a Circular Cylinder: Reynolds Number Effects. *Computers & Fluids*, *47* (1), 44–50.
14. Nguyen, V.-T., & Nguyen, H. (2016). Detached Eddy Simulations of Flow Induced Vibrations of Circular Cylinders at High Reynolds Numbers. *Journal of Fluids and Structures*, *63*, 103–119.

15. Gopalkrishnan, R. (1993). *Vortex-Induced Forces on Oscillating Bluff Cylinders*. PhD Thesis, Massachusetts Institute of Technology.
16. Norberg, C. (2003). Fluctuating Lift on a Circular Cylinder: Review and New Measurements. *Journal of Fluids and Structures*, *17* (1), 57–96.
17. Montgomery, D. (2017). *Design and analysis of experiments*. New York: John Wiley & Sons.
18. Kleijnen, J. P. C. (2016). Regression and Kriging Metamodels with their Experimental Designs in Simulation: A Review. *European Journal of Operational Research*, *000*, 1–6.
19. Feng, X., Zhang, Y., & Wu, J. (2018). Interval Analysis Method Based on Legendre Polynomial Approximation for Uncertain Multibody Systems. *Advances in Engineering Software*, *121*, 223–234.
20. Meckesheimer, M., Booker, A., Barton, R., & Simpson, T. (2002). Computationally Inexpensive Metamodel Assessment Strategies. *AIAA Journal*, *40* (10), 2053–2060.

sciendo

LATVIAN JOURNAL OF PHYSICS
AND TECHNICAL SCIENCES
2021, N 5

DOI: 10.2478/lpts-2021-0040

RESEARCH OF KINEMATIC STEPPING MECHANISM

I. Grinevich*, V. Nikishin, G. Springis

Riga Technical University, Daugavpils Affiliation,
90 Smilsu Str., Daugavpils, LV-5410, LATVIA
*e-mail: Ivans.Grinevics@rtu.lv

In this paper, a kinematic analysis of Theo Jansen's stepping mechanism has been carried out and an algorithm for finding the output link trajectory from the given dimensions of the stepping mechanism elements, implemented by Mathcad program, has been developed. It is possible to output characterising parameters of all intermediate links with any number of intermediate links of a step cycle of the mechanism. The dimensions of the mechanism elements have been selected so that they provide the optimal smooth trajectory of the stepping point, minimising the mechanism oscillations in the vertical plane. A comparison of the trajectories of the foot in this study with the classical trajectory of Theo Jansen and the trajectory from article [7] has been provided. A minimum swing of the oscillation of the centre of mechanism masses in the vertical plane has been selected as an optimality criterion, combined with the maximum smoothness of the trajectory, provided that maximization of the step height is not required.

Keywords: *Kinematic analysis, step trajectory, Theo Jansen mechanism.*

1. INTRODUCTION

Robotics is one of the fastest growing areas of modern science. A topical direction in this area is the development and study of stepping mechanisms. Theo Jansen's [1], [2] stepping mechanism, simulating the gait of a living organism, is widely-known. Theo Jansen's "beach animals" created on this principle from plastic pipes are set in motion by the force of the wind. In [3], an overview of the work on the applications of the Jansen mechanism in various fields is given. It was stated that the majority of developments were at the research stage. In [4], a detailed study of the Jansen mecha-

nism was carried out, the trajectory of the stepping point was determined, and a modification of the basic mechanism was proposed in order to improve its characteristics. In [5], a method for increasing the maneuverability through a mechanism with 5 links implementing a step height change was proposed. To find the trajectory of the proposed mechanism and the Jansen mechanism, an analytical method was applied using complex number forms. The problem of stability of stepping mechanisms was investigated in [6]. In [7], a structural analysis of the Jansen mechanism was carried out and the trajectory of the output link was determined from 12 intermediate positions.

2. THEORY

The scheme of the stepping mechanism (Theo Jansen's mechanism) is shown in Fig. 1. It consists of two stiff triangles AON, FBT and rods KA = a, KB = z, NF = g, OB = l, O_1K=r. Points A, O, N, B, F, K, O_1 are hinges, with O and O_1 being fixed firmly. By rotating the rod $O_1K(r)$ against the clock, the mechanism is set in motion. The step trajectory is implemented by vertex T of triangle FBT. Let us denote the sides of the triangles as AN=w, OA=v, ON=u , FB=f, BT=p, FT=s, ∠AON=α, ∠FBT=β, distance OO_1 =d. It is necessary to select numerical values for parameters r, a, z, d, g, l, v, u, w, f, p, s, which provide the optimal trajectory of point T. We introduce a rectangular coordinate system xOy (Fig. 1), then the coordinates of point T.

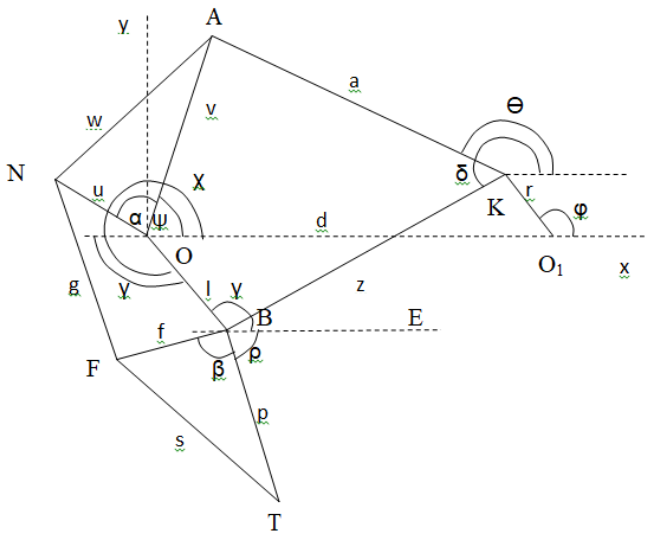


Fig. 1. Scheme of the stepping mechanism.

$$\begin{cases} x_T = x_B + p \cdot \cos\rho \\ y_T = y_B + p \cdot \sin\rho \end{cases}, \quad (1)$$

where x_B, y_B are the coordinates of point B, angle ρ is the angle between horizontal BE and side BT of the lower triangle (Fig. 1). To obtain the coordinates of point B, the vector equality is used:

$$\overrightarrow{OB} = \overrightarrow{OK} + \overrightarrow{KB}, \quad (2)$$

where $\overrightarrow{OK} = \overrightarrow{OO_1} + \overrightarrow{O_1K}$.

Or:

$$\vec{l} = \vec{d} + \vec{r} + \vec{z}, \quad (3)$$

This equality in coordinate form:

$$\begin{cases} x_B = x_d + x_r + x_z \\ y_B = y_d + y_r + y_z \end{cases}, \quad (4)$$

or $\begin{cases} l \cdot \cos\chi = d + r \cdot \cos\phi + z \cdot \cos\delta \\ l \cdot \sin\chi = 0 + r \cdot \sin\phi + z \cdot \sin\delta \end{cases}$

where φ is the angle between the horizontal and O_1K, χ is the angle between Ox and OB, δ is the angle between the horizontal and KB (Fig.1). From (4) it follows:

$$\begin{aligned} z \cdot \sin\delta &= l \cdot \sin\chi - r \cdot \sin\phi \\ z \cdot \cos\delta &= l \cdot \cos\chi - r \cdot \cos\phi - d \end{aligned}, \quad (5)$$

and by the ratio:

$$(z \cdot \sin\delta)^2 + (z \cdot \cos\delta)^2 = z^2, \quad (6)$$

the expression is found:

$$A \cdot \sin\chi + B \cdot \cos\chi = C, \quad (7)$$

where

$$\begin{aligned} A &= 2 \cdot l \cdot r \cdot \sin\phi, \\ B &= 2 \cdot l \cdot r \cdot \cos\phi + 2 \cdot d \cdot l, \\ C &= l^2 + r^2 + d^2 + 2 \cdot r \cdot d \cdot \cos\phi - z^2, \end{aligned} \quad (8)$$

In (7), a replacement is used:

$$\sin\chi = \frac{2 \cdot t}{1 + t^2}, \quad \cos\chi = \frac{1 - t^2}{1 + t^2}, \quad (9)$$

where $\chi \neq \pi + 2\pi n$ и $t = tg\left(\frac{\chi}{2}\right)$.

Resulting in a quadratic equation:

$$(B + C) \cdot t^2 - 2 \cdot A \cdot t + (C - B) = 0, \quad (10)$$

the roots of which will be real at a positive discriminant:

$$A^2 + B^2 - C^2 \geq 0: \quad (11)$$

$$t_{1,2} = \frac{2 \cdot A \pm \sqrt{4 \cdot A^2 - 4 \cdot (B + C) \cdot (C - B)}}{2 \cdot (B + C)} = \frac{A \pm \sqrt{A^2 + B^2 - C^2}}{B + C}, \quad (12)$$

then

$$\chi_{1,2} = 2 \cdot arctg(t) \ . \quad (13)$$

The graph of χ_2 depending on angle φ is shown in Fig. 2, and angle χ_1 corresponds to an unrealizable trajectory (the mechanism is mirrored relative to the horizontal).

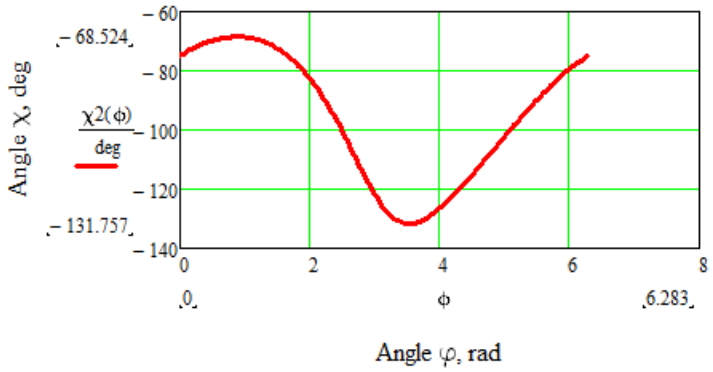


Fig. 2. Dependence of χ_2 on angle φ.

The coordinates of point B (Fig. 3) will be:

$$\begin{cases} x_B = l \cdot \cos(\chi_2(\varphi)) \\ y_B = l \cdot \sin(\chi_2(\varphi)) \end{cases}. \qquad (14)$$

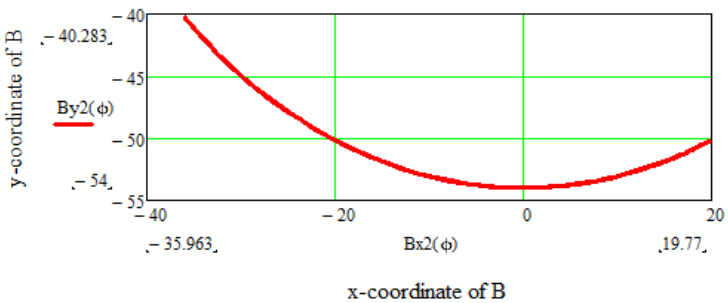


Fig. 3. Trajectory of point.

Similarly, the coordinates of point A are found:

$$\overrightarrow{OA} = \overrightarrow{OK} + \overrightarrow{KA}. \qquad (15)$$

Taking into account that

$\overrightarrow{OA} = \overrightarrow{OK} + \overrightarrow{KA}$. then

$$\vec{v} = \vec{d} + \vec{r} + \vec{a}. \qquad (16)$$

The coordinate form of these relationships is:

$$\begin{cases} x_A = x_d + x_r + x_a \\ y_A = y_d + y_r + y_a \end{cases}, \text{ or}$$

$$\begin{cases} v \cdot \cos\psi = d + r \cdot \cos\varphi + a \cdot \cos\theta \\ v \cdot \sin\psi = 0 + r \cdot \sin\varphi + a \cdot \sin\theta \end{cases}, \qquad (17)$$

where $a \cdot \sin\theta = v \cdot \sin\psi - r \cdot \sin\phi$,
$a \cdot \cos\theta = v \cdot \cos\psi - d - r \cdot \cos\phi$.

From the ratio:

$$(v \cdot \sin\psi - r \cdot \sin\varphi)^2 + \left(v \cdot \cos\psi - \left(r \cdot \cos\varphi + d\right)\right)^2 = a^2, \qquad (18)$$

we find A_1:

$$A_1 \cdot \sin\psi + B_1 \cdot \cos\psi = C_1, \quad (19)$$

where

$$A_1 = 2 \cdot v \cdot r \cdot \sin\phi,$$
$$B_1 = 2 \cdot v \cdot r \cdot \cos\phi + 2 \cdot d \cdot v,$$
$$C_1 = v^2 + r^2 + d^2 + 2rd \cdot cos\varphi - a^2, \quad (20)$$

In B (19) a replacement is used:

$$\sin\psi = \frac{2 \cdot \tau}{1+\tau^2}, \quad \cos\psi = \frac{1-\tau^2}{1+\tau^2}, \quad (21)$$

where $\psi \neq \pi + 2\pi n$ и $\tau = tg\left(\frac{\psi}{2}\right)$, (22)

resulting in a quadratic equation in relation to τ:

$$(B_1 + C_1) \cdot \tau^2 - 2 \cdot A_1 \cdot \tau + (C_1 - B_1) = 0. \quad (23)$$

The roots of this quadratic equation:

$$\tau_{1,2} = \frac{2 \cdot A_1 \pm \sqrt{4 \cdot A_1^2 - 4 \cdot (B_1 + C_1) \cdot (C_1 - B_1)}}{2 \cdot (B_1 + C_1)} = \frac{A_1 \pm \sqrt{A_1^2 + B_1^2 - C_1^2}}{B_1 + C_1}, \quad (24)$$

with the corresponding angles:

$$\psi_{1,2} = 2 \cdot arctg(\tau_{1,2}), \quad (25)$$

Thus, $\psi_{1,2}$ will be real if the discriminant is positive $A_1^2 + B_1^2 - C_1^2 \geq 0$. The dependence of ψ_1 on angle φ is shown in Fig. 4, and the angle ψ_2 corresponds to an unrealizable trajectory.

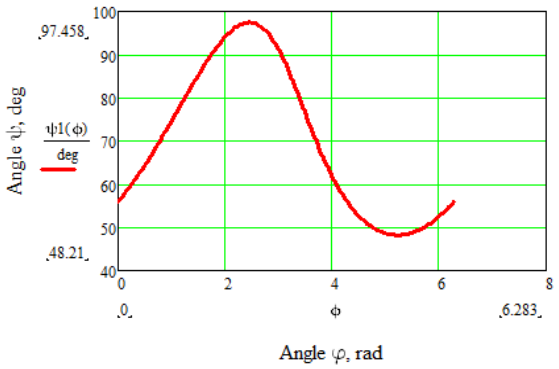


Fig. 4. Dependence of ψ1 on angle φ.

The coordinates of point A (Fig. 5) are found as follows:

$$\begin{cases} x_A = v \cdot \cos(\psi_1(\varphi)) \\ y_A = v \cdot \sin(\psi_1(\varphi)) \end{cases}. \quad (26)$$

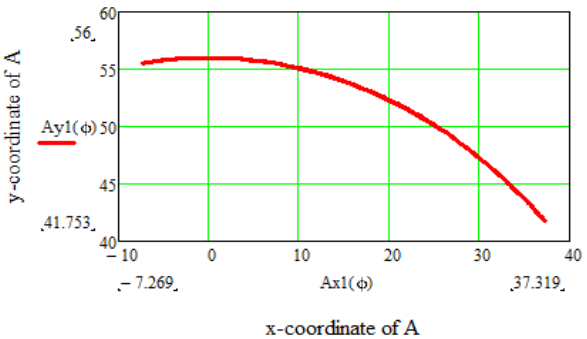


Fig. 5. Trajectory of point A.

From quadrilateral BFNO (Fig. 6) the angle ρ is determined (BE – horizontal):

$$2\pi - \rho(\phi) = \eta(\phi) + \nu(\phi) + \beta + \gamma(\phi)\,. \qquad (27)$$

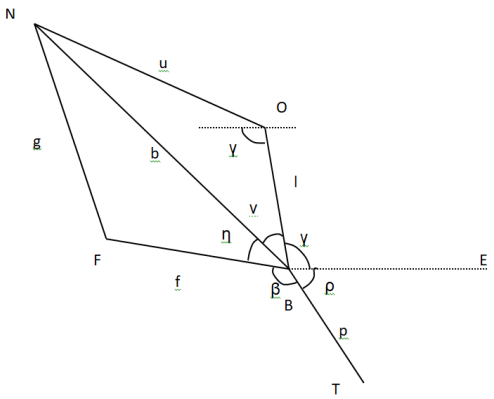


Fig .6. Angles in BFNO (BN=b).

Angles ν and η are determined after dividing the quadrilateral by diagonal b into two triangles (Fig. 6).

According to the cosine theorem:

$$b(\varphi) = \sqrt{u^2 + l^2 - 2\cdot u\cdot l\cdot \cos(\gamma(\varphi)}\,, \qquad (28)$$

with

$$\gamma(\varphi) = 2\cdot\pi - \psi_1(\varphi) - \chi_2(\varphi) - \alpha\,, \qquad (29)$$

where α is found from triangle AON (Fig.1):

$$\alpha = \arccos\left(\frac{v^2 + u^2 - w^2}{2\cdot u\cdot v}\right).$$

The change in the length of diagonal BN = b depending on angle φ is shown in Fig. 7. The graph of Eq. (29) is shown in Fig. 8.

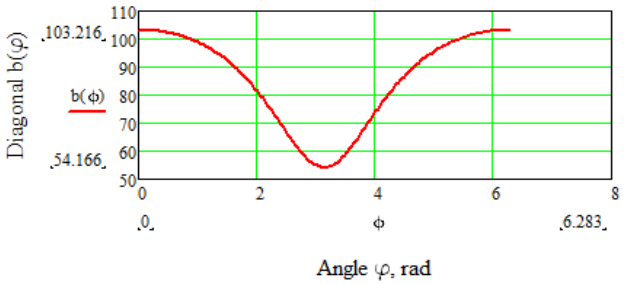


Fig. 7. Dependence of diagonal b on angle φ.

Similarly:

$$u^2 = b^2 + l^2 - 2\cdot l\cdot b\cdot \cos\nu\,, \qquad (30)$$

where

$$\nu(\varphi) = \arccos\left(\frac{b(\varphi)^2 + l^2 - u^2}{2\cdot b(\varphi)\cdot l}\right). \qquad (31)$$

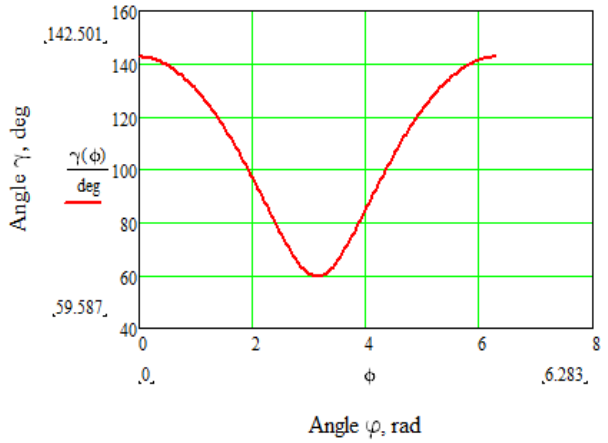


Fig. 8. Dependence of angle γ on φ.

Similarly:

$$g^2 = b^2 + f^2 - 2 \cdot b \cdot f \cdot \cos\eta, \qquad (32)$$

$$\eta(\varphi) = \arccos\left(\frac{b(\varphi)^2 + f^2 - g^2}{2 \cdot b(\varphi) \cdot f}\right). \qquad (33)$$

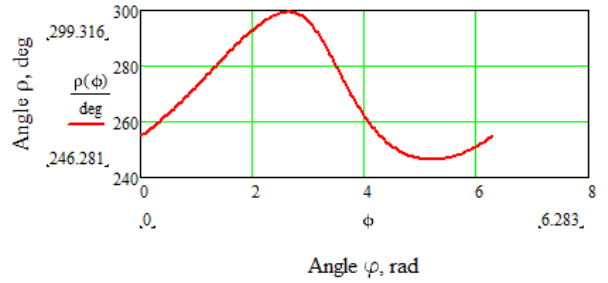


Fig. 9. Dependence of angle ρ on angle φ.

Then we find:

$$\rho(\varphi) = \pi + \beta + \text{í}(\varphi) + \text{ç}(\varphi) - \div(\varphi), \qquad (34)$$

where angle β is defined from triangle FBT (Fig. 1):

$$s^2 = f^2 + p^2 - 2 \cdot f \cdot p \cdot \cos\beta$$
$$\beta = \arccos\left(\frac{f^2 + p^2 - s^2}{2 \cdot f \cdot p}\right). \qquad (35)$$

Figures 10–12 show examples of trajectories of stepping point T defined by (1) for different sizes of the elements of the mechanism.

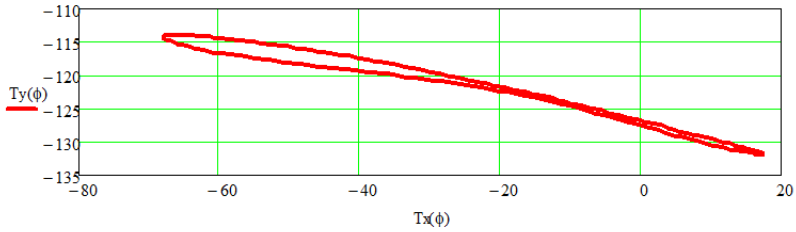


Fig. 10. Invalid trajectory of point T.

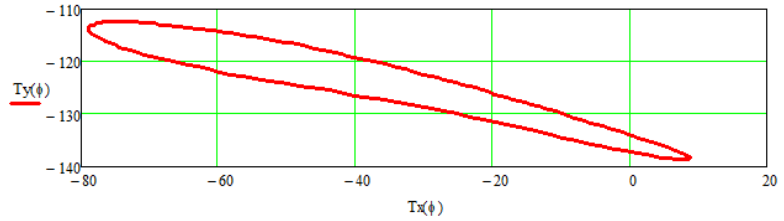


Fig. 11. Acceptable trajectory of point T.

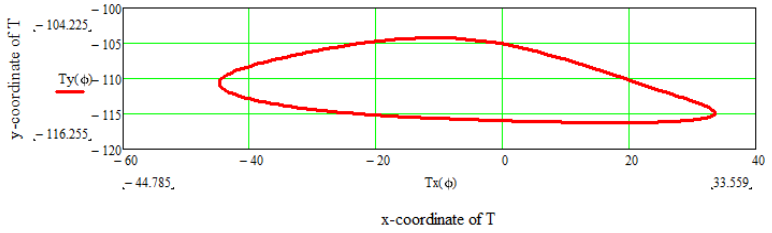


Fig. 12. Optimal trajectory of point T.

RESULTS AND DISCUSSION

The optimal dimensions of the elements of the stepping mechanism were selected (Fig. 1): r=20(O_1K), a=68(KA), z=85(BK), d=61(OO_1), g=65(NF), l=54(OB), v=56(OA), u=55(NO), w=76(AN), f=50(FB), p=66(BT), s=90(FT). In this case, it is necessary to raise point O_1 of the mechanism above the horizontal so that the straight line OO_1 makes the angle of 5° with the horizontal, which brings the lower part of the trajectory (Fig.12) closer to the horizontal. The swing amplitude of the gravity centre of the mechanism in the vertical plane at the given dimensions is approximately 3.2 at the step length of 78.3.

For comparison, Fig. 13 shows the trajectory of the mechanisms created by Theo Jansen, the dimensions of the elements are taken from [4] and are given in the second column of Table 1. Figure 14 shows the tra-

jectory obtained in [7], the corresponding dimensions of the elements taken from the illustration in this work (the dimensions are not indicated by the authors) and are given in the third column of Table 1.

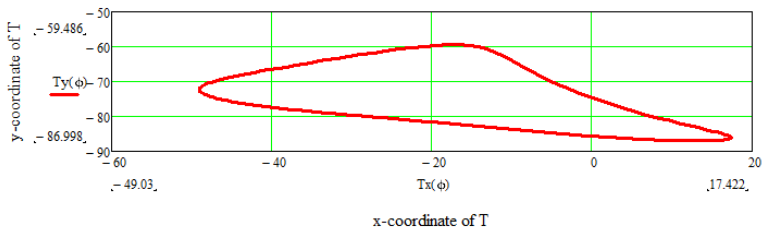


Fig. 13. The trajectory of Theo Jansen's mechanism.

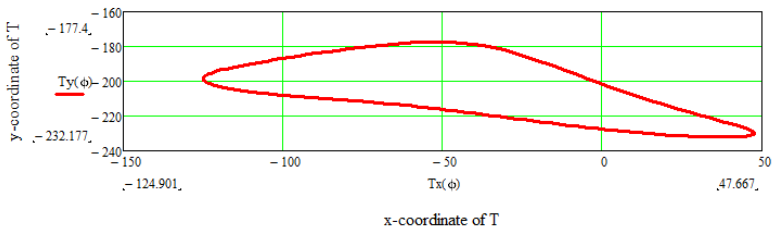


Fig. 14. The trajectory of this work [7].

Table 1. Comparison of Dimensions and Parameters of Mechanisms

Parameter	Theo Jansen	Fomin A., et al. [7]	Present research
r	15	40	20
a	50	137	68
z	61.9	167	85
d	38.79	107	61
g	39.4	106	65
l	39.3	107	54
v	41.5	111	56
u	40.1	108	55
w	55.8	149	76
f	36.7	97	50
p	49	129	66
s	65.7	174	90
Length of the step	67.8	172.5	78.3
Swing of occislations	3.2	6.3	3.2
Ratio	3.2/67.8=0.047	6.3/172.5=0.037	3.2/78.3=0.041

The bottom line of the table shows the main parameter characterising the oscillations of the centre of masses of the mechanism in the vertical plane, which is equal to

the ratio of the swing of oscillations to the step length of the mechanism. The trajectory of this work (Fig. 12) in this parameter exceeds the trajectory of Theo Jansen (Fig. 13), but it is less than the trajectory of [7] (Fig. 14). However, approximately in the centre, the lower part of the trajectory of work [7] has a slight deflection upward from the desired straight line, which generates unwanted oscillations of the mechanism with a higher frequency compared to the main frequency of a full step.

4. CONCLUSIONS

The trajectories of Theo Jansen and work [7] have a common defect – a sharp shape of the rear part. The transition from the bottom of the trajectories to the top part is not smooth. The trajectory proposed in this paper is free from this drawback and is smooth. These rationales allow us to consider the trajectory proposed in this study as optimal.

REFERENCES

1. Jansen, B. L., Doubrovski, E., & Verlinden, J. C. (2014). Animaris Geneticus Parvus: Design of a Complex Multi-Body Walking Mechanism. *Rapid Prototyping Journal, 20* (4), 311–319.
2. Jansen, T. (2007). *The great pretender.* 10 Publishers.
3. Somak, S., & Pramod, B. (2017). Study of Applications of Jansen's Mechanism in Robot. *International Journal of Advance Research and Innovation*, *5* (3), 354–357.
4. Ghassaei, A. (2011). *The Design and Optimization of a Crank-Based Leg Mechanism*. PhD Thesis. Pomona College Department of Physics and Astronomy, Claremont.
5. *Lovasz, E. Ch., Pop, C., Pop, F., & Dolga V. (2014). Novel Solution for Leg Motion with 5-Link Belt Mechanism. International Journal of Applied Mechanics and Engineering, 19 (4), 699–708.*
6. Komoda, K., & Wagatsuma, H. A. (2011). Study of Availability and Extensibility of Theo Jansen Mechanism toward Climbing over Bumps. *The 21st Annual Conference of the Japanese Neural Network Society*, 3 – 28.
7. Fomin, A. S., Kiselev, S. V., & Oleksenko, A. V. (2017). Strukturnij analiz mehanizma Jansena. *Vestnik Sibirskovo gosudarstvennovo industrialnovo universiteta*, *4* (22), 51–53.