62 /4/2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS

WARSAW 2024, QUARTERLY, VOLUME 62, INDEX 365238, ISSN 1429-2955

POLISH SOCIETY OF THEORETICAL AND APPLIED MECHANICS ptmts

POLISH SOCIETY OF THEORETICAL AND APPLIED MECHANICS

JOURNAL OF THEORETICAL AND APPLIED MECHANICS

No. 4 • Vol. 62

Quarterly

eISSN: 2543-6308

ISSN: 1429-2955

WARSAW, November 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
(until 1997 Mechanika Teoretyczna i Stosowana, ISSN 0079-3701)

Beginning with Vol. 45, No. 1, 2007, *Journal of Theoretical and Applied Mechanics* (JTAM) has been selected for coverage in Thomson Reuters products and custom information services. Now it is indexed and abstracted in the following:

- **Science Citation Index Expanded** (also known as SciSearch®)
- **Journal Citation Reports/Science Edition**

* * * * *

Editorial Office

Al. Armii Ludowej 16, room 650; 00-637 Warsaw, Poland
phone (+48) 664 099 345, e-mail: biuro@ptmts.org.pl

www.jtam.pl

* * * * *

Rozwój kwartalnika naukowego *Journal of Theoretical and Applied Mechanics*, ISSN 1429-2955, jest dofinansowany ze środków Ministra Edukacji i Nauki przyznanych z pomocy *de minimis* w ramach programu „Rozwój czasopism naukowych", umowa RCN/SN/0056/2021/1. Niniejszy numer został sfinansowany przez ministerstwo w ramach projektu: Doskonała nauka – Wsparcie konferencji naukowych. V Polski Kongres Mechaniki-25 Międzynarodowa Konferencja Metod Komputerowych Mechaniki, umowa DNK/SN/548136/2022.

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 655-665, Warsaw 2024
https://doi.org/10.15632/jtam-pl/193379

EXPERIMENTAL STUDIES OF THE FLAPPING MOTION OF A BUTTERFLY WING MODEL

Zuzanna Kunicka-Kowalska
Kazimierz Wielki University, Faculty of Mechatronics, Bydgoszcz, Poland
e-mail: zkunicka@ukw.edu.pl
Krzysztof Sibilski
Warsaw University of Technology, Faculty of Power and Aeronautical Engineering, Warsaw, Poland
e-mail: krzysztof.sibilski@pw.edu.pl

The article describes an experiment of the movement of a butterfly wing model in glycerin to obtain a relationship between rotation functions and aerodynamic forces. Measurements were taken during the movement of an artificial wing, modelled on a real one, in a dense viscous medium with proportional reduction of the movement frequency. Reducing the frequency, i.e. slowing down the wing, makes the observation easier and produces more reliable results. The tests started with the movement observed in the living butterfly and then, in each subsequent step, the angles of inclination were modified in such a way as to describe it in terms of simple functions, without losing their physical properties, i.e. due to aerodynamic forces. According to the literature, the insect flight can be divided into three phases of stroke in flapping flight. In each, aerodynamic forces are generated or aerodynamic drag is minimised. The experimental tests were filmed, and the pressure differences were measured. Based on the data analysis, the functions of wing inclination angles in time for the characteristic flight moments – acceleration, hover, and take-off – were specified.

Keywords: flapping flight, butterfly flight, experimental mechanics, verification of numerical calculations

1. Introduction

In literature, the flight of living organisms is often discussed, but there appears a varied phenomenon making its individual aspects still not well understood (Shyy *et al.*, 2007). The aim of this article is to understand deformations of a butterfly wing and their impact on aerodynamic forces. One of the first significant publications that attempted to explain the key issues related to aerodynamics of flapping insect wings was a series of six articles by Ellington (1984). The article (Reade and Jankauski, 2020) is an example of a paper devoted to the issue of butterfly wing kinematics and aerodynamics. Its authors describe wing motion using two vectors (for the front and rear wing). The description did not cover wing deformations or insect in-flight position, as in most studies. Fry *et al.* (2005) present insect flight research results (on the example of a Drosophila Melanogaster fruit fly). The authors distinguish three flapping wing motion phases – beat, break-off and rotation, while noting that the wing motion in each insect is a species feature. The article (Steppan, 2000) presents a series of studies of displacements as well as forces and torques acting on the wings of Idea leuconoe butterfly in flapping flight. The analysis discussed in (Piechna, 1997) indicates that the lift force in a fly is generated mainly in the course of downwards wing motion, and the thrust during upwards wing motion. Furthermore, in butterflies, the wings meet in the upper position first at the front, followed by the other parts of the wing, from front to back – generating something called the "pumping effect". The wings start spreading from the front with the largest driving force generated at that moment. This produces

an increase in thrust by up to 40%. It should be noted that butterflies are classified as insects with a low beat frequency, large area and span of the wings (Sun and Bhushan, 2012). Based on the analysis of source literature data, the authors of (Sun and Bhushan, 2012) state that a butterfly flying forwards always flaps, while combining flapping with gliding to improve flight efficiency. Experimental studies of vortex structures generated by the wings of a Papilio Ulysses were also reviewed in (Hu *et al.*, 2009). The authors of (Okamoto *et al.*, 2009; Hu and Wang, 2010) discuss the gliding flight. Senda *et al.* (2012) calculate the aerodynamic force by applying a model composed of numerous rigid elements, and compare the results with a force measured in the course of the experiment. A two-dimensional analysis of flexible wing flow is presented in (Bluman and Kang, 2017), where the authors attempt to show that wing flexibility leads to an increase in its lift force. Biej-Bijenko (1976) considers butterflies as belonging to a group of functionally double-winged insects, since during flight, the rear pair remains connected with the front pair and they work together as a single surface. Essentially, the information required for this work is that the wings consist of venation that acts as the structure while membranes stretch in between. Such veins are also arranged in an irregular pattern (not along, not across, sometimes almost radially) (Ha *et al.*, 2011).

In order for the research described in this article to be carried out, the investigation and calculations constituting the introduction herein were previously performed: Numerical calculations of the flow around the rigid wing have been carried out, which was necessary (Kunicka-Kowalska *et al.*, 2022b) to start the research. Then, by increasingly complicating the modelling, the model became more realistic by taking into account the flexibility of the wing structure (Kunicka-Kowalska and Sibilski, 2018; Kunicka-Kowalska *et al.*, 2022a). The mechanical properties of this butterfly wings were studied in (Landowski *et al.*, 2010) and taken from the literature (Sun and Bhushan, 2012). The input data for calculations were obtained from an experiment with a live Attacus Atlas insect (Kunicka-Kowalska and Sibilski, 2018). Numerical calculations were performed in the commercial Ansys Fluent software, the algorithms of which are based on the Navier-Stokes equation. In order to validate the deformation and experimentally demonstrate the influence of the modification of the wing rotation on pressure differences of the medium resulting in aerodynamic forces, the tests described in this article were carried out.

2. Materials and methods

Because the frequency of flapping in the air is high and makes experimental work and observation impossible, it was decided to use technical glycerin as the medium. It is a thick and colourless liquid. It was assumed that deformations are caused by propulsive forces of the medium in which the object is moving, so to obtain the same deformations, these forces had to be compared.

$$\varepsilon_p = \varepsilon_g - F_{rp} = F_{rg} \tag{2.1}$$

where: ε_p – deformation in the air, ε_g – deformation in the glycerin, F_{rp} – resistance force in the air, F_{rg} – resistance force in the glycerin.

Following the above reasoning, the frequency that would give analogous deformations was determined by the formula

$$C\frac{\rho_p v_p^2}{2}S = C\frac{\rho_g v_g^2}{2}S \tag{2.2}$$

where: ρ_p – air density, ρ_g – glycerin density, v_p – object velocity in the air, v_g – object velocity in the glycerin, S – area of projected object on a plane perpendicular to the velocity vector.

Since in both cases the object is the same and the frequency is directly proportional to the velocity, the equation can be reduced to a simple proportion

$$\frac{v_g}{v_p} = \sqrt{\frac{\rho_p}{\rho_g}} = \frac{f_g}{f_p} \quad \rightarrow \quad f_g = \sqrt{\frac{\rho_p}{\rho_g}} f_p \quad \rightarrow \quad f_g = \sqrt{\frac{0.00119}{1.26}} 5.48 = 0.168 \text{ Hz} \tag{2.3}$$

where: f_p – frequency in the air, f_g – frequency in the glycerin.

Based on previous research (Kunicka-Kowalska *et al.*, 2022b), the flutter frequency was set to 5.48 Hz. This means that in one minute, the wing will beat 10.08 times, and one stroke will take about 5.95 s. This is a speed suitable for observing deformations, so glycerin is the right medium for this experiment. It can also be easily concluded that with the application of water, this speed would be higher

$$\frac{v_w}{v_p} = \sqrt{\frac{\rho_p}{\rho_w}} = \frac{f_w}{f_p} \quad \rightarrow \quad f_w = \sqrt{\frac{\rho_p}{\rho_g}} f_p \quad \rightarrow \quad f_g = \sqrt{\frac{0.00119}{1}} 5.48 = 0.189 \text{ Hz} \tag{2.4}$$

where: f_w – frequency in water.

Thus, it can be seen that in one minute, the wing will perform 11.34 strokes in water while maintaining the same deformations. However, the Reynolds number, which depends on the fluid viscosity and the remaining dimensionless numbers, was also important for the experiment. Table 1 shows the dimensionless numbers for the flapping wing in water and glycerin while maintaining the deformations observed in the air. The biggest difference is in the Reynolds number. The use of glycerin instead of water allows for any orientation of the wing in space and easier interpretation of the pressure reading in the liquid, which indicates the initial direction and returns of the fluid velocity vector at the wing surface. The authors of the article [19] demonstrate the relationship between the Strouhal number and aerodynamic forces in a flapping flight. In the case in question, however, the movement lasts a short time: the pressure is measured during one stroke. In this situation and with high fluid viscosity, which in turn significantly affects the Reynolds number, the fluid dynamics is reduced and, consequently, the velocity has a limit at zero. This means that regardless of the time instant, the dimensionless Strouhal number will tend to infinity. This allows for a qualitative comparison of pressures and a quantitative comparison of deformations as a simplification to solid mechanics. Table 1 contains dimensionless numbers for continuous motion at the same wing speed.

Table 1. Dimensionless number for the flapping wing in water and glycerin while maintaining the deformations observed in the air

Dimensionless number	Value in water	Value in glycerine
Reynolds number, Re	6.6715	0.0057
Non-dimensional frequency, k	1.0000	1.0000
Strouhal number, St	0.5000	0.5000

To obtain deformations similar to those in nature and to correctly conduct the experiment, it was necessary to choose the right material. The most important properties of the material sought were:

- high flexibility,
- no or negligible hygroscopicity – constant parameters after immersion in glycerin,
- isotropy (different flexibility is to be obtained by varying the thickness of the model),
- ease of processing and forming.

It was decided that the best solution would be to use silicone. Young's modulus of moulding silicone is 0.0015 GPa. From the formula for theoretical Young's modulus, the thickness that

should be kept so the elasticity constant E_c does not change was calculated (Kunicka-Kowalska„ 2020). The model was made in two stages: the veins system and membrane. First, a container was prepared and raised on one side to a given height. There is a fluid level mark in the container. Blue moulding silicone was poured into the prepared container. Thanks to the prepared form, a variable thickness of the initial element was obtained. After concentration, the silicone was deformed, a vein template was applied (Kunicka-Kowalska *et al.*, 2022a) and the lines were cut manually. Next, the container was prepared again (this time by setting it horizontally), and a line pattern was placed on the bottom. A thin layer of transparent moulding silicone was then poured over it. After obtaining a uniform thickness of the layer, the blue trough was arranged according to the diagram at the bottom of the container. After binding, the whole was deformed, and the wing was cut along the contour, applying the template again. Through this procedure, a uniform material was obtained (with the same properties), with no gluing, which made it possible to avoid the additional stiffness of the adhesive. In addition, the veining is highlighted in blue. The wing membrane remained transparent allowing the phenomena to be seen on the other side of the wing surface. Therefore, it was possible to create a real model characterized by the same susceptibility to deformation as the real wing. Additionally, the model made it possible to observe the phenomena in front and behind it due to its transparency. The difference between the real wing and the model was visible in the thickness of the element, which was a necessary simplification to obtain the same deformations. The wing was seated with a sleeve, with the mount in the air, not in the liquid. It was not possible to fully immerse the arm because the servomotors could not be submerged – their sealing would be insufficient under the high pressure of the column of a thick liquid (real model see Fig. 1).

Fig. 1. Photo of the actual model

The next step was to model the changes in inclination angles of the artificial wing. The wing kinematics is described in the control program as rotations around the axis: roll around axis X, yaw around axis Y, pitch around axis Z (see Fig. 2). To realize the movement physically as rotation around axes passing through one point, three servomotors were used. Each is responsible

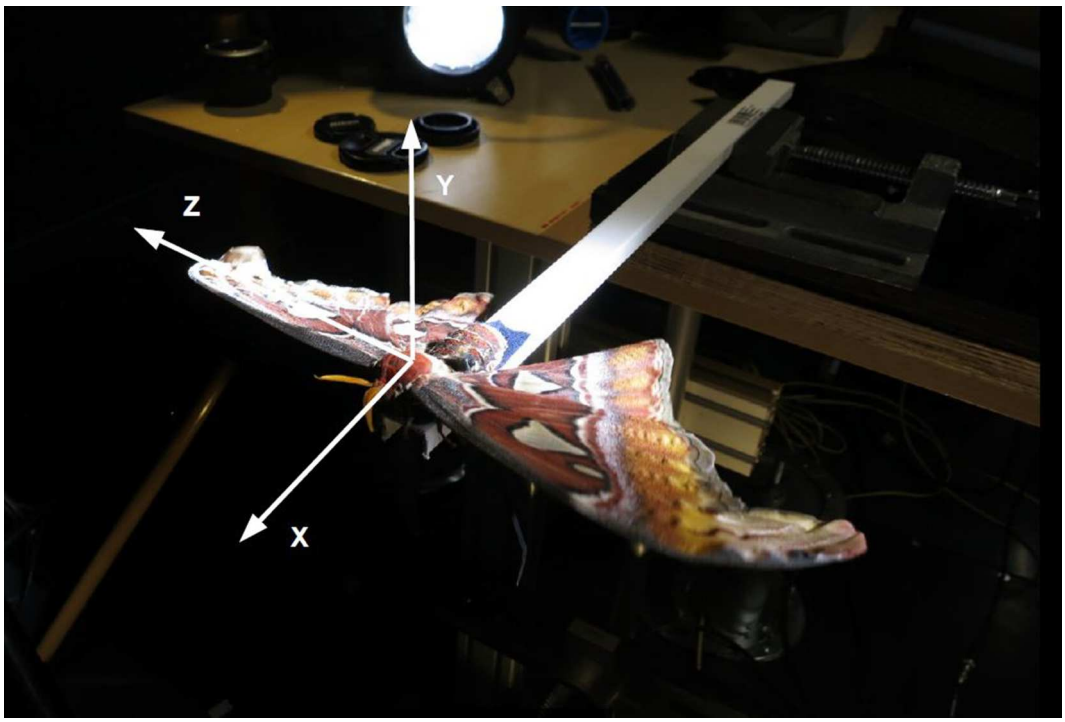


Fig. 2. The adopted coordinate system

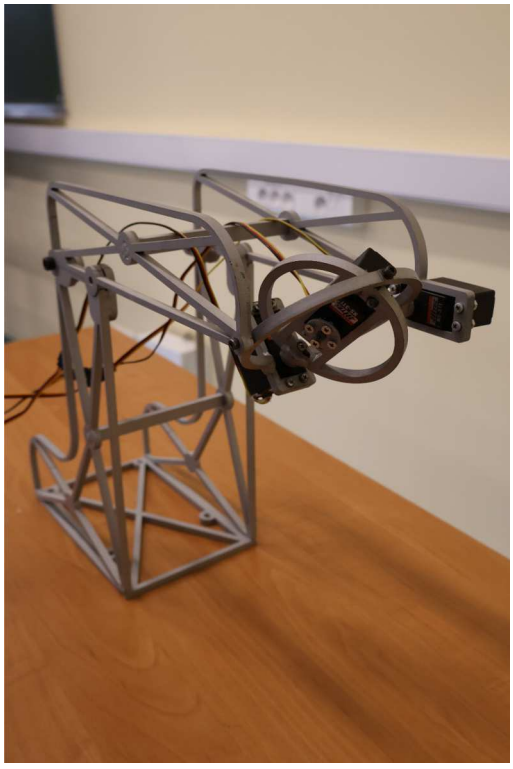


Fig. 3. The wing model control mechanism

for rotation around a different axis. One was located on the axis of rotation and the other two rotated the hoops (see Fig. 3). This solution enabled the implementation of trading at a point. It is possible to tilt the entire arm to temporarily remove the model from the liquid. The whole structure is attached to the plate on which the liquid container is placed (Fig. 4). The places

where the pressure difference is measured are marked with red arrows in Fig. 4. The rotation of the servomotors around the axis was programmed based on the Arduino IDE software.

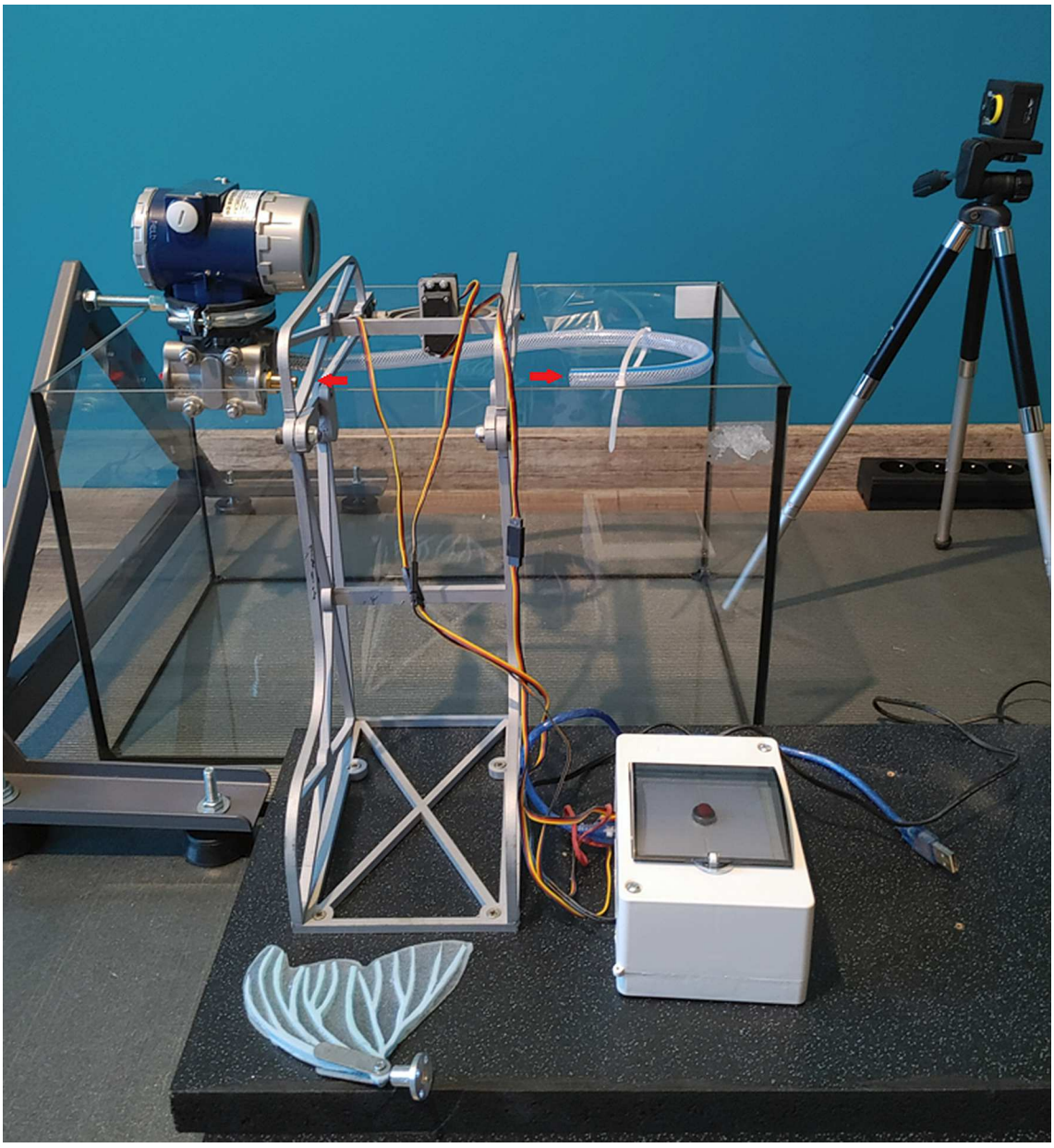


Fig. 4. View of the experimental stand

A single stroke cycle of the wing was initiated by pressing the trigger button. Repetition of the cycle required another button pressing. As the system was located in a very dense and viscous liquid, the influence of gravity on the behaviour of the liquid could be neglected, and thus the wing could be arranged in the liquid in any (from a pragmatic point of view) position. The adopted coordinate system was marked directly on the container and is visible on the recording. The centre of the coordinate system is the same as in the previous numerical analysis (Kunicka-Kowalska *et al.*, 2022a).

3. Results

The wing deflection was measured as the angle of deviation of the plane of the deformed wing from its original position for the same time instants. The wing deflection observed in the original movement (experimental data) was 18 degrees. This shows a significant convergence of the observation of a living object with numerical calculations (19 degrees) and an experiment with the real model (17 degrees). This stage was a form of verification of the correctness of both the calculations and the silicone model itself. After observation and visual analysis, the test stand was expanded by adding a differential pressure sensor (BD Sensors DPT 200 XMD, Warszawa, Poland). The sensor (accuracy: 0.075% of the range) registered the pressure difference in front of and behind the wing in the case of thrust measurement, and above and below the wing in the case of lift measurement. The sensor was placed in the immediate vicinity of the tested object, and a high viscosity of the fluid was a guarantee of moderate diffusion of the phenomenon. All validation measurements, as well as measurements of movement and pressure for subsequent modifications, were recorded with a camera.

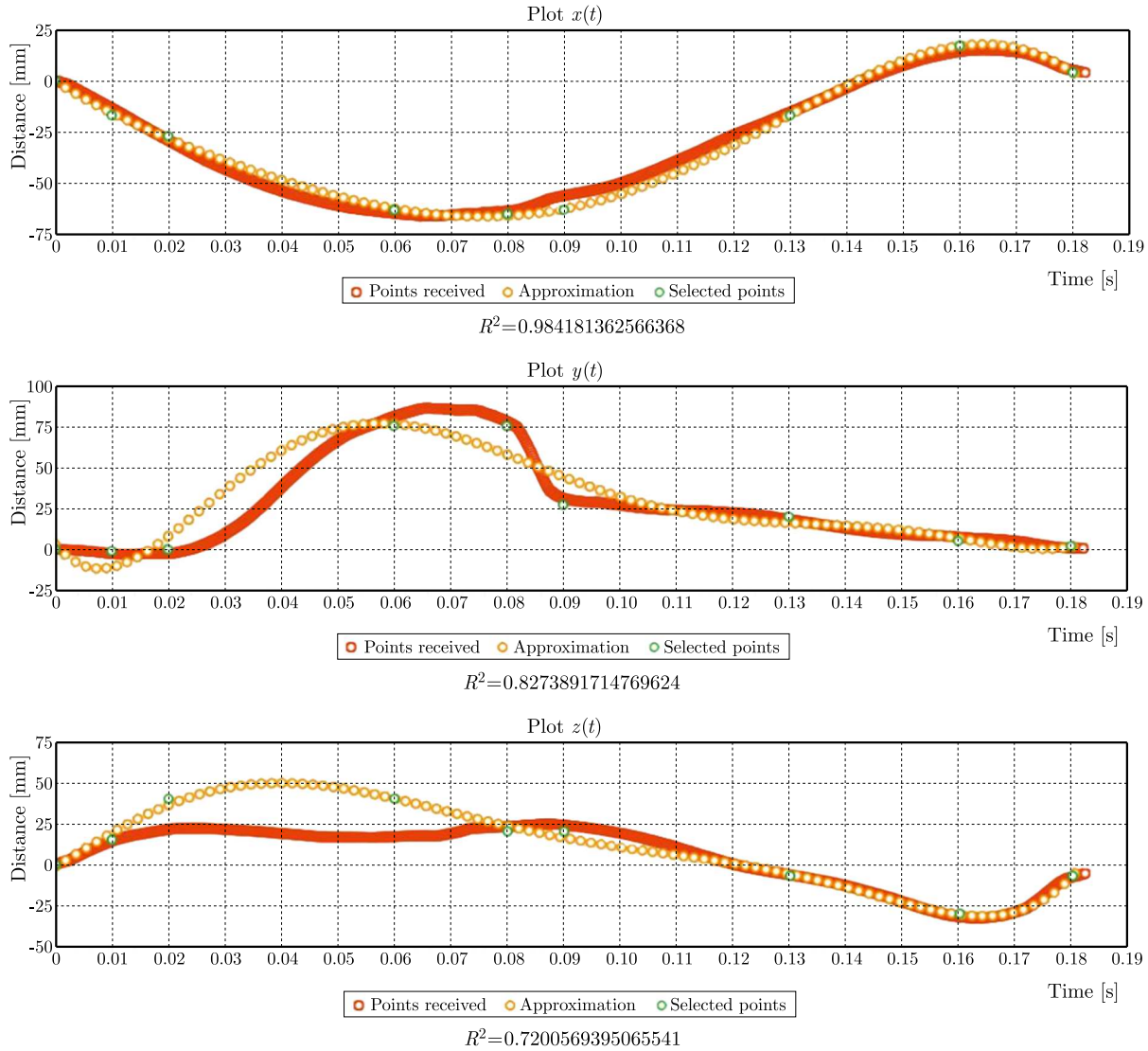


Fig. 5. Modification of wing rotation as a function of time, version A

During the experiment, a number of modifications to the rotational movement of the wing were then introduced. The modifications were introduced to demonstrate how changing rotation at a specific stage of stroke affects the pressure differential. These modifications were introduced by moving the plot points in the Stirling program (approximation software). The algorithm then adjusted the course of polynomial curve, thus correcting the original motion. The most important modifications are listed below:

- modification A – increasing the angle of attack during the climb phase: the potential benefit should be an increase in the lift by obtaining greater dynamics of the leading edge vortex (LEV);
- modification B – reduction of the angle of attack when the wings move backwards: the expected effect would be an increase in the thrust as a result of increasing the wing area during push-back;
- modification AB – combination of modifications A and B, i.e. a simultaneous increase of the angle of attack during the climb phase and a decrease when the wings move backwards: in this case, it was expected to obtain both increased lift and thrust;
- modification C – modification of experimental data by adjusting the function to the literature data (Shyy *et al.*, 2007): the aim was verification;
- modification CB – combination of modification B (reduction of the angle of attack when the wings move backwards) in the first part of the graph, and modification of C (Shyy *et al.*, 2007) in the final part of the function.

For example, the modification of rotation in version A with time is shown in Fig. 5. The orange points show the primary kinematics of the butterfly wing. The measurement of the actual wing rotation was made based on photos from a high-speed camera, the data includes 912 points during one wing stroke (Kunicka-Kowalska, 2020).

4. Discussion

The mathematical description of the wing rotation was implemented in the program controlling the servomotors via Arduino. The obtained results of the pressure difference change over the cycle and are shown in the graphs (Figs. 6 and 7). Table 2 shows the total pressure difference in the cycle. Positive values show positive strength. Due to the high frequency of the wing impact, the total pressure difference in the cycle can be considered as a single impulse. However, this value as a function of time shows the influence of particular phases of wing movement (and therefore also their modification) on the final effect in the form of aerodynamic forces.

Table 2. Total pressure difference in the cycle for various modifications pressure difference [mbar]

Pressure difference [mbar]	Tsst	Mod A	Mod B	Mod AB	Mod C	Mod CB
Thrust	−0.651	−0.083	0.069	−0.045	−0.093	−0.148
Lift	0.022	0.108	0.138	2.14	0.075	−0.033

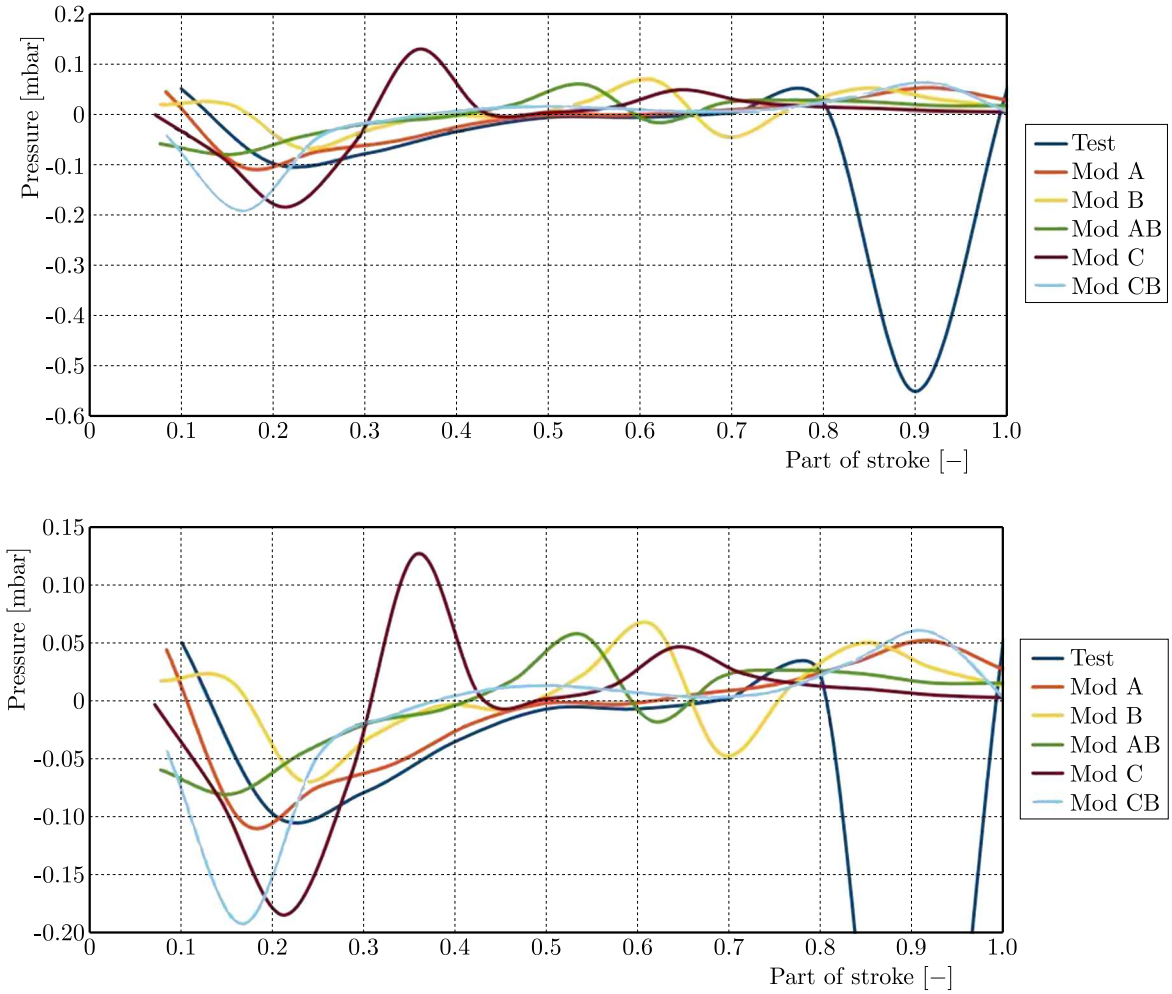


Fig. 6. Pressure difference in front of and behind the wing during one stroke for individual modifications (thrust) [mbar]

5. Conclusions

According to the previous research (Kunicka-Kowalska *et al.*, 2022b), the thrust is generated by the shape of this wing itself, but it is small. The research has shown that only increasing the wing angle of attack produces thrust. Moreover, increasing the stroke amplitude does not have a positive effect on the increasing thrust. However, the initially large angle of attack increases the thrust in the wing descent phase. This has a direct impact on the increasing of pumping effect (creating an air cushion under the wing), but also increases the frontal area when raising the wing. Thus, the original position of the wing at rest is of great importance.

The most important conclusion regarding the measured pressures that influence the lift is that increasing the angle of attack while raising the wing increases the LEV energy and thus increases the lift. In the second phase of movement, the mentioned angle of attack influences the use of resulting air cushion to generate the lift force. Based on the analyses, it was found that AB modification contributes to a significant increase in the lift while reducing the thrust – it can be used as a temporary, very energy-consuming procedure. Modification C, on the other hand, is the closest to hovering. CB modification does not bring the expected physical effects.

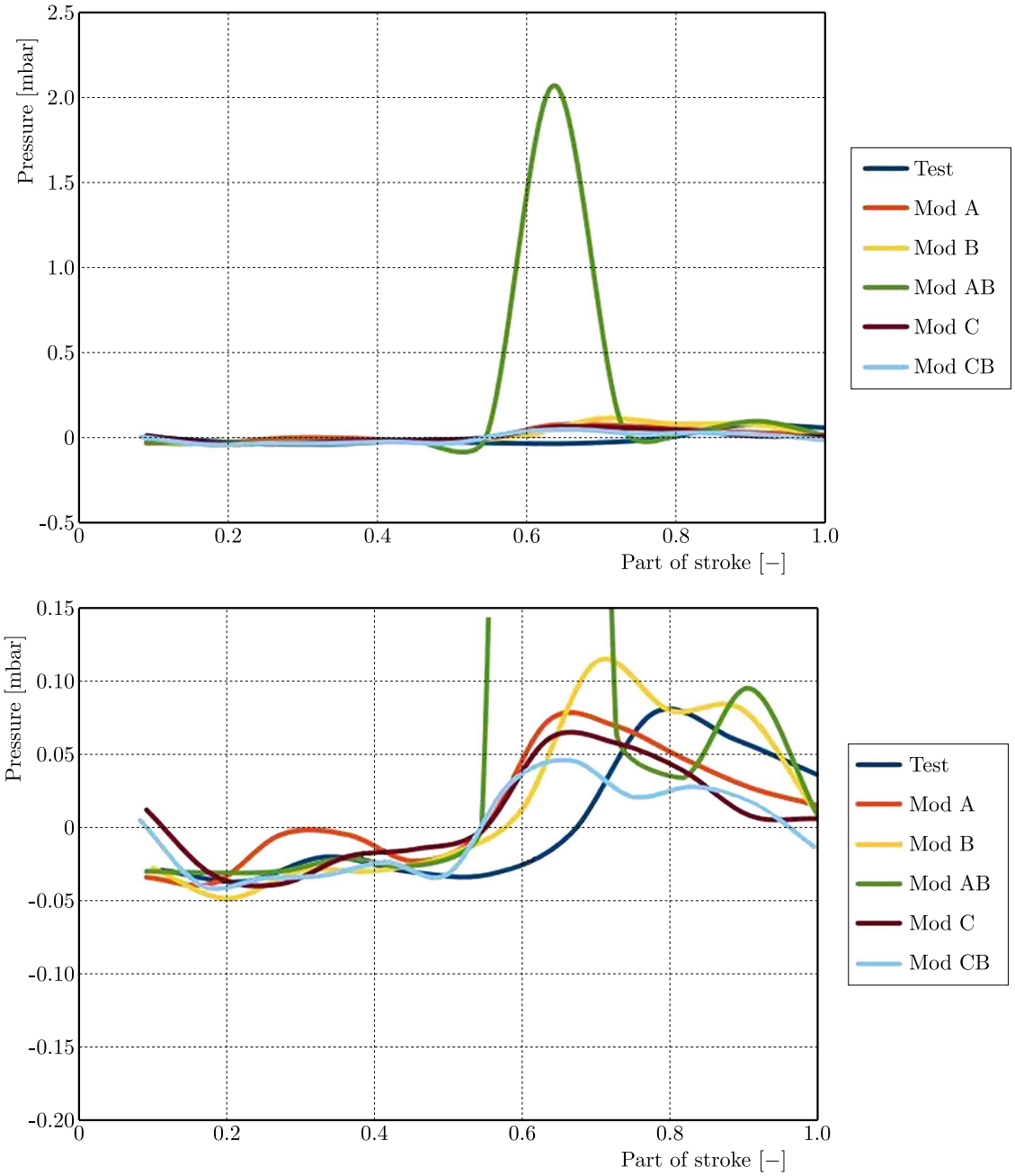


Fig. 7. Pressure difference above and below the wing during one stroke for individual modifications (lift) [mbar]

Acknowledgments

The research was financed by statutory funds of the Faculty of Power and Aeronautical Engineering of the Warsaw University of Technology granted for 2020/2021.

References

1. Aditya K., Malolan V., 2012, *Investigation of Strouhal Number Effect on Flapping Wing Micro Air Vehicle*, https://doi.org/10.2514/6.2007-486 [18.06.2012]
2. Biej-Bijenko G.J., 1976, *Introduction to Entomology* (in Polish), Powszechne Wydawnictwo Rolnicze i Leśne, Warsaw, Poland

3. Bluman J., Kang C.-K., 2017, Achieving hover equilibrium in free flight with a flexible flapping wing, *Journal of Fluids and Structures*, **75**, 117-139

4. Ellington C.P., 1984, The aerodynamics of hovering insect flight I-VI, *Philosophical Transactions of the Royal Society B*, **305**, 1122, 1-181

5. Fry S.N., Sayaman R., Dickinson M.H., 2005, The aerodynamics of hovering flight in *Drosophila*, *Journal of Experimental Biology*, **208**, 12, 2303-2318

6. Ha N.S., Jin T.L., Goo N.S., Park H.C., 2011, Anisotropy and non-homogeneity of an *Allomyrina Dichotoma* beetle hind wing membrane, *Bioinspiration and Biomimetics*, **6**, 4, 046003

7. Hu Y., Wang J., 2010, Experimental investigation on aerodynamic performance of gliding butterflies, *AIAA Journal*, **48**, 10, 2454-2457

8. Hu Y., Wang J., Zhang P.F., Zhang C., 2009, Experimental investigation on the flow structure over a simplified Papilio Ulysses model, *Chinese Science Bulletin*, **54**, 6, 1026-1031

9. Kunicka-Kowalska Z., 2020, Modeling the flow around the insect flapping wing on the example of the butterfly Attacus atlas, PhD Thesis, Warsaw University of Technology, Warsaw, Poland

10. Kunicka-Kowalska Z., Landowski M., Sibilski K., 2022a, Deformable model of a butterfly in motion on the example of Attacus atlas, *Journal of the Mechanical Behavior of Biomedical Materials*, **133**, 105351

11. Kunicka-Kowalska Z., Landowski M., Sibilski K., 2022b, The motion analysis of *Attacus atlas* rigid wing, *International Review of Aerospace Engineering*, **15**, 4, 205-214

12. Kunicka-Kowalska Z., Sibilski K., 2018, Study on Attacus Atlas wing motion during flapping flight for needs of FEM simulation (in Polish), [In:] *Mechanika w Lotnictwie, ML-XVIII 2018*, **II**, K. Sibilski (Edit.), Polish Society of Theoretical and Appplied Mechanics, Warsaw, 35-43

13. Landowski M., Kunicka-Kowalska Z., Sibilski K., 2020, Mechanical and structural investigations of wings of selected insect species, *Acta of Bioengineering and Biomechanics*, **22**, 2, 199-209

14. Okamoto M., Sunada S., Tokutake H., 2009, Stability analysis of gliding flight of a swallowtail butterfly *Papilio xuthus*, *Journal of Theoretical Biology*, **257**, 2, 191-202

15. Piechna J., 1997, How do animals fly? (in Polish), *Wiedza i Życie*, **8**, 24-29

16. Reade J., Jankauski M.A., 2020, Deformable blade element and unsteady vortex lattice fluid-structure interaction modeling of a 2D flapping wing, *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, **83969**, American Society of Mechanical Engineers

17. Senda K., Obara T., Kitamura M., Nishikata T., Hirai N., *et al.*, 2012, Modeling and emergence of flapping flight of butterfly based on experimental measurements, *Robotics and Autonomous Systems*, **60**, 5, 670-678

18. Shyy W., Lian Y., Tang J., Viieru D., Liu H., 2007, *Aerodynamics of Low Reynolds Number Flyers*, Cambridge University Press, Cambridge, UK

19. Steppan S.J., 2000, Flexural stiffness patterns of butterfly wings (Papilionoidea), *The Journal of Research on the Lepidoptera*, **35**, 61-77

20. Sun J., Bhushan B., 2012, The structure and mechanical properties of dragonfly wings and their role on flyability, *Comptes Rendus. Mécanique*, **340**, 1-2, 3-17

Manuscript received January 19, 2024; accepted for publication September 13, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 667-681, Warsaw 2024
https://doi.org/10.15632/jtam-pl/193256

DYNAMIC RESPONSE OF SATURATED FROZEN SOIL FOUNDATION UNDER UNDERLYING BEDROCK TO HARMONIC LOAD

Huaiyuan Chen

School of Civil Engineering, Qinghai University, Xining, China

Qiang Ma

School of Civil Engineering, Qinghai University, Xining, China, and

Qinghai Provincial Key Laboratory of Energy-saving Building Materials and Engineering Safety, Xining, China

e-mail: maqiang0104@163.com

This study aims to provide a theoretical basis for engineering construction in alpine frozen soil area, simplify the soil medium of the site in alpine permafrost region to saturated permafrost, and expand the elastic foundation and saturated ground foundation to better reflect the dynamic response problem of permafrost site in alpine region. Based on the theory of solid porous media with pores, a physical model of dynamic response of saturated frozen soil foundation with underlying bedrock under vertical harmonic load is established, and the effects of temperature, porosity, cementation parameters, load frequency and contact parameters on the dynamic response are analyzed.

Keywords: soil medium, saturated frozen soil foundation, dynamic response

1. Introduction

China is the third largest frozen soil country in the world. Frozen soil is mainly distributed in China's western high-altitude and cold regions. With the implementation of the national western development strategy and the "One Belt, One Road" strategy, there are many infrastructure projects under construction which are proposed to be implemented in the western high-altitude and high-cold areas. Because the soil layer in the western high altitude and cold area is frozen chiefly soil, the mechanical properties of frozen soil are very different from those of unfrozen soil. Therefore, in order to ensure stability as well as safety of engineering buildings in the permafrost region, the research on the dynamic response of permafrost needs to be carried out urgently.

Since the first study of single-phase elastic half-space dynamics by Lamb (1904) in 1904, more and more scholars (Ai *et al.*, 2018; Liang *et al.*, 2020) have gradually improved the theoretical study of the dynamic response of single-phase elastic foundations. However, it is not practical to analyze the soil by simply assuming it to be single-phase elastic, and when the single-phase soil is filled with water, it becomes two-phase saturated soil. Biot (1956; 1962) was the first to study the dynamic response of saturated soil foundation, and put forward the fluctuation equation of saturated soil. Based on Biot's theory, Zhou *et al.* (2013) obtained numerical solutions for the displacement and stress integrals of soil particles and pore water in saturated soils when simple harmonic loads were applied to the surface of the saturated soil foundation. Xu *et al.* (2009) further considered the layered nature of the soil, and used the transfer matrix method and Hankel transform to study the dynamic response of a layered saturated soil foundation when a horizontal simple harmonic load was applied to the surface of the foundation, and found that the layered nature of the soil had a significant effect on the dynamic response of the foundation. Liu *et al.* (2022) investigated the dynamic response of a layered transverse

isotropic saturated land base under two-dimensional moving loads under permeable and impermeable conditions on the surface using Fourier series expansion and dual variable and position methods. However, soils in nature are usually unsaturated soils, which consist of three phases: gas phase, liquid phase and soil particle phase, and the influence of gas phase in unsaturated soils on the dynamic response should not be neglected. Considering the coupling effect of the three phases of unsaturated soils, Xu (2010) and Xu *et al.* (2011) obtained the fluctuation equation of unsaturated soils by using Bishop's effective stress equation combined with the V-G model of capillary pressure function, and investigated the dynamic response of unsaturated semi-space surfaces under arbitrary loading. Shi *et al.* (2021; 2022) investigated the dynamic response of semi-space unsaturated soil foundations under simple harmonic loads under different boundary conditions by utilizing the Fourier transform. Ma *et al.* (2023a) developed a layered foundation model and analyzed and discussed the dynamic response of three types of layered unsaturated foundations, homogeneous soil, hard interbedded and soft interbedded, under moving loads using coordinate transformations and the transfer and reflection matrix method. Ma *et al.* (2023b) investigated the dynamic response of different types of loads on transversely isotropic multilayered unsaturated and ground foundations in a three-dimensional column coordinate system using the Laplace-Hankel transform.

The above studies on the foundation dynamic response mostly regard foundation soil as single-phase elastic soil, two-phase saturated soil and three-phase unsaturated soil, but in the high-latitude and high-altitude region of western China, the site soil medium is mostly permafrost, and many projects in the western region involve permafrost, and the dynamic response of the permafrost is very different from that of the unfrozen soil due to the existence of the ice phase. Therefore, on the basis of the above research, this paper simplifies the frozen soil into saturated frozen soil, establishes the two-dimensional saturated frozen soil foundation power control equation under a simple harmonic load based on the theory of porous solid porous medium, and obtains the answer to the question of the power response of saturated frozen soil foundation in the frequency domain by using the Fourier integral transform, the principle of Helmholtz vector decomposition and combining with the boundary conditions. The effects of temperature, porosity, cementation parameters, and loading frequency on the dynamic response of a two-dimensional saturated frozen soil foundation are analyzed and discussed through numerical examples.

2. Mechanical model and governing equations

2.1. Mechanical model

Considering a saturated frozen soil foundation with finite thickness, the surface action has a vertical strip harmonic load, as shown in Fig. 1, where H is the layer thickness, q_0 is the linear load acting on the surface of the foundation, l is the half of the linear load length, q is the load amplitude, $q = q_0/(2\pi)$, and ω is the circular frequency of the load vibration. Assuming that the surface is permeable and the bottom bedrock is impermeable, the coordinate axis shown in the diagram is established, and the coordinate origin is placed on the surface of the foundation, that is, the surface of the foundation $z = 0$.

2.2. Governing equations

According to the definition of Zhou and Lai (2011), ice is formed in pores and coexists with liquid water in thepores. Therefore, saturated frozen soil can be regarded as a three-phase porous medium composed of soil particle phase, pore liquid water phase, and pore ice phase, that is, solid phase, liquid phase, and ice phase. The volume fraction of each phase is expressed as follows

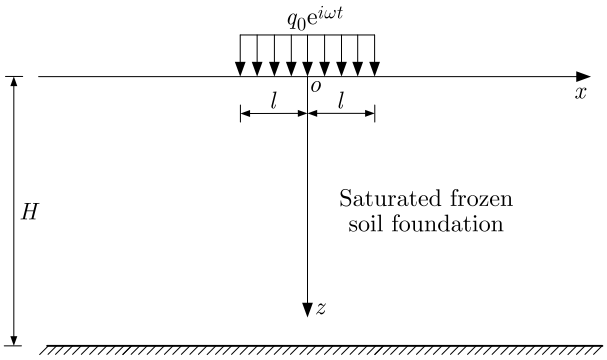


Fig. 1. A schematic diagram of the harmonic load acting on a two-dimensional saturated frozen soil foundation

$$\phi_S = 1 - \phi \qquad \phi_F = \phi S_r \qquad \phi_I = \phi(1 - S_r) \tag{2.1}$$

where S_r denotes the saturation level of pore liquid water, ϕ denotes the porosity of porous media; the subscripts S, F and I denote the volume fraction of soil particles, pore liquid water, and pore ice in saturated frozen soil, respectively.

Under a harmonic load, the stress and displacement of each phase in saturated frozen soil can be expressed as follows

$$G(x, z, t) = G(x, z)^* \mathrm{e}^{\mathrm{i}\omega t} \tag{2.2}$$

where i is an imaginary unit, ω is the frequency of load circle; for the convenience of representation, the superscript $*$ in the following text is omitted.

The governing equation for saturated frozen soil can be derived based on the theory presented by Leclaire *et al.* (1994, 1995) and further developed by Carcione *et al.* (2000, 2003), Carcione and Seriani (2001)

$$\begin{aligned}
&\rho_{11}\ddot{\mathbf{u}}^S + \rho_{12}\ddot{\mathbf{u}}^F + \rho_{13}\ddot{\mathbf{u}}^I = R_{11}\nabla(\nabla\cdot\mathbf{u}^S) + R_{12}\nabla(\nabla\cdot\mathbf{u}^F) + R_{13}\nabla(\nabla\cdot\mathbf{u}^I)\\
&\quad - \mu_{11}\nabla\times\nabla\times\mathbf{u}^S - \mu_{13}\nabla\times\nabla\times\mathbf{u}^I - (b_{12}+b_{13})\dot{\mathbf{u}}^S + b_{12}\dot{\mathbf{u}}^F + b_{13}\dot{\mathbf{u}}^I\\
&\rho_{12}\ddot{\mathbf{u}}^S + \rho_{22}\ddot{\mathbf{u}}^F + \rho_{23}\ddot{\mathbf{u}}^I = R_{12}\nabla(\nabla\cdot\mathbf{u}^S) + R_{22}\nabla(\nabla\cdot\mathbf{u}^F) + R_{23}\nabla(\nabla\cdot\mathbf{u}^I)\\
&\quad + b_{12}\dot{\mathbf{u}}^S - (b_{12}+b_{23})\dot{\mathbf{u}}^F + b_{23}\dot{\mathbf{u}}^I\\
&\rho_{13}\ddot{\mathbf{u}}^S + \rho_{23}\ddot{\mathbf{u}}^F + \rho_{33}\ddot{\mathbf{u}}^I = R_{13}\nabla(\nabla\cdot\mathbf{u}^S) + R_{23}\nabla(\nabla\cdot\mathbf{u}^F) + R_{33}\nabla(\nabla\cdot\mathbf{u}^I)\\
&\quad - \mu_{13}\nabla\times\nabla\times\mathbf{u}^S - \mu_{33}\nabla\times\nabla\times\mathbf{u}^I + b_{13}\dot{\mathbf{u}}^S + b_{23}\dot{\mathbf{u}}^F - (b_{13}+b_{23})\dot{\mathbf{u}}^I
\end{aligned} \tag{2.3}$$

where ρ_{ij} $(i = 1, 2, 3;\ j = 1, 2, 3)$ denote the coupling inertia coefficients between each phase; $\mathbf{u}^S$, $\mathbf{u}^F$ and $\mathbf{u}^I$ denote the displacement vectors of soil particles, pore liquid water, and pore ice in the saturated frozen soil medium, respectively. b_{12}, b_{23} and b_{13} are viscosity parameters; R_{ij} $(i = 1, 2, 3;\ j = 1, 2, 3)$ and μ_{11}, μ_{13}, μ_{33} are stiffness parameters. The coefficients of ρ_{ij}, R_{ij}, μ_{11}, μ_{13}, μ_{33}, b_{12}, b_{23}, b_{13}, etc. are detailed in Qiu *et al.* (2018) literature. The symbols $(\cdot)$, $(\cdot\cdot)$ denote the first and second derivatives with respect to time, respectively; ∇^2 is the Laplacian operator in a Cartesian coordinate system.

The stress-strain relationship in the saturated permafrost medium is as follows

$$\begin{aligned}
\sigma_{ij}^S &= (K_1\theta_S + C_{12}\theta_F + C_{13}\theta_I)\delta_{ij} + 2\mu_{11}d_{ij}^S + \mu_{13}d_{ij}^I\\
\sigma^F &= C_{12}\theta_S + K_2\theta_F + C_{23}\theta_I\\
\sigma_{ij}^I &= (C_{13}\theta_S + C_{23}\theta_F + K_3\theta_I)\delta_{ij} + 2\mu_{33}d_{ij}^I + \mu_{13}d_{ij}^S
\end{aligned} \tag{2.4}$$

and

$$
\begin{aligned}
&C_{12} = (1-c_1)\phi_S\phi_F K_{av} \qquad C_{13} = (1-c_1)(1-c_3)\phi_S\phi_I K_{av}\\
&C_{23} = (1-c_3)\phi_I\phi_F K_{av}\\
&K_1 = [(1-c_1)\phi_S]^2 K_{av} + K_{sm} \qquad K_2 = \phi_F^2 K_{av} \qquad K_3 = [(1-c_3)\phi_I]^2 K_{av} + K_{im}\\
&\theta_S = u^S_{i,i} \qquad \theta_F = u^F_{i,i} \qquad \theta_I = u^I_{i,i}\\
&d^S_{ij} = \varepsilon^S_{ij} - \frac{1}{3}\theta_S\delta_{ij} \qquad d^I_{ij} = \varepsilon^I_{ij} - \frac{1}{3}\theta_I\delta_{ij} \qquad \varepsilon^S_{ij} = \frac{1}{2}(u^S_{i,j} + u^S_{j,i})\\
&\varepsilon^I_{ij} = \frac{1}{2}(u^I_{i,j} + u^I_{j,i})
\end{aligned}
\tag{2.5}
$$

where σ^S_{ij}, σ^F, σ^I_{ij} denote the stress components acting on the solid phase, liquid phase, and ice phase, respectively; K_1, K_2 and K_3 are bulk moduli of elasticity; C_{12}, C_{13} and C_{23} are the bulk modulus between the three phases. δ_{ij} is the Kronecker symbol; ε^a_{ij}, d^a_{ij} and θ_a denote the body strain, bias strain, and strain of the a ($a = S, F, I$) phase, respectively. The coefficients of c_1, c_3, K_{av}, K_{im}, K_{sm}, etc. are detailed in Qiu *et al.* (2018).

According to Zhou (2020), among the parameters involved in the governing equations of saturated frozen soil, porosity, and ice content are considered to be significant factors that influence the characteristics of the soil. Leclaire *et al.* (1994) used the normal distribution of pores to describe the ice content and considered that at a certain temperature, when the pore is less than a certain scale, the water in the pore will not freeze, and the water in the pore larger than this pore diameter will freeze completely. Therefore, the relationship between temperature and pore ice content is calculated as follows

$$
\phi_F = (1-\phi_S)A \int_0^{r_0/\ln(T_0/T_{k0})} \exp\Big[-\frac{(r-r_{av})^2}{2\Delta r^2}\Big]\, dr \tag{2.6}
$$

By normalizing the Gaussian probability function from $r = 0 \to \infty$, the expression of unfrozen water content used in this paper can be expressed as follows

$$
\phi_F = (1-\phi_S)\frac{\operatorname{erf}(\zeta) + \operatorname{erf}(\eta)}{1 + \operatorname{erf}(\eta)} \qquad \zeta = \frac{r_0/\ln(T_0/T_{k0})}{\sqrt{2}\Delta r} - \eta \qquad \eta = \frac{r_{av}}{\sqrt{2}\Delta r} \tag{2.7}
$$

where r_{av} is the average pore radius; Δr is the standard deviation of pore radius; $r_0 = 0.228\,\mathrm{nm}$, $T_{k0} = T + T_0$, T_{k0}, T_0, T all indicate temperature, T_{k0}, T_0 in Kelvin, $T_0 = 273.15\,\mathrm{K}$, T in Celsius. The parameters η, ζ are explained in detail in Carcione and Seriani (2001).

Substituting Eq. (2.2) into Eqs. (2.3), we can get the following

$$
\begin{aligned}
&a_{11}\mathbf{u}^S + a_{12}\mathbf{u}^F + a_{13}\mathbf{u}^I = R_{11}\nabla(\nabla\cdot\mathbf{u}^S) + R_{12}\nabla(\nabla\cdot\mathbf{u}^F) + R_{13}\nabla(\nabla\cdot\mathbf{u}^I)\\
&\quad - \mu_{11}\nabla\times\nabla\times\mathbf{u}^S - \mu_{13}\nabla\times\nabla\times\mathbf{u}^I\\
&a_{12}\mathbf{u}^S + a_{22}\mathbf{u}^F + a_{23}\mathbf{u}^I = R_{12}\nabla(\nabla\cdot\mathbf{u}^S) + R_{22}\nabla(\nabla\cdot\mathbf{u}^F) + R_{23}\nabla(\nabla\cdot\mathbf{u}^I)\\
&a_{13}\mathbf{u}^S + a_{23}\mathbf{u}^F + a_{33}\mathbf{u}^I = R_{13}\nabla(\nabla\cdot\mathbf{u}^S) + R_{23}\nabla(\nabla\cdot\mathbf{u}^F) + R_{33}\nabla(\nabla\cdot\mathbf{u}^I)\\
&\quad - \mu_{13}\nabla\times\nabla\times\mathbf{u}^S - \mu_{33}\nabla\times\nabla\times\mathbf{u}^I
\end{aligned}
\tag{2.8}
$$

where $a_{11} = -\rho_{11}\omega^2 + (b_{12} + b_{13})\mathrm{i}\omega$, $a_{12} = -\rho_{12}\omega^2 - b_{12}\mathrm{i}\omega$, $a_{13} = -\rho_{13}\omega^2 - b_{13}\mathrm{i}\omega$, $a_{22} = -\rho_{22}\omega^2 + (b_{12} + b_{23})\mathrm{i}\omega$, $a_{23} = -\rho_{23}\omega^2 - b_{23}\mathrm{i}\omega$, $a_{33} = -\rho_{33}\omega^2 + (b_{13} + b_{23})\mathrm{i}\omega$.

The displacement vector of each phase can be expressed by the Helmholtz vector decomposition theorem as

$$
u^a = \nabla\varphi_a + \nabla\times\psi_a \qquad \nabla\cdot\psi_a = 0 \qquad a = S, F, I \tag{2.9}
$$

where φ_a and ψ_a are the three-phase medium scalar and vector potential functions, respectively.

Substituting Eq. (2.9) into Eqs. (2.8), we can get

$$\begin{aligned} &a_{11}\varphi_S + a_{12}\varphi_F + a_{13}\varphi_I = R_{11}\nabla^2\varphi_S + R_{12}\nabla^2\varphi_F + R_{13}\nabla^2\varphi_I \\ &a_{12}\varphi_S + a_{22}\varphi_F + a_{23}\varphi_I = R_{12}\nabla^2\varphi_S + R_{22}\nabla^2\varphi_F + R_{23}\nabla^2\varphi_I \\ &a_{13}\varphi_S + a_{23}\varphi_F + a_{33}\varphi_I = R_{13}\nabla^2\varphi_S + R_{23}\nabla^2\varphi_F + R_{33}\nabla^2\varphi_I \end{aligned} \tag{2.10}$$

$$\begin{aligned} &a_{11}\psi_S + a_{12}\psi_F + a_{13}\psi_I = \mu_{11}\nabla^2\psi_S + \mu_{13}\nabla^2\psi_I \\ &a_{12}\psi_S + a_{22}\psi_F + a_{23}\psi_I = 0 \qquad a_{13}\psi_S + a_{23}\psi_F + a_{33}\psi_I = \mu_{13}\nabla^2\psi_S + \mu_{33}\nabla^2\psi_I \end{aligned} \tag{2.11}$$

3. The solution of the potential function

The above governing equations can be transformed into ordinary differential equations, which are easy to solve by the Fourier transform of the horizontal coordinate x. The Fourier transform of the function $f(x,z)$ is as follows

$$\widetilde{f}(\xi,z) = \int_{-\infty}^{\infty} f(x,z)\mathrm{e}^{-\mathrm{i}\xi x}\,dx \tag{3.1}$$

where ξ is the Fourier transform parameter, the superscript $\widetilde{(\cdot)}$ denotes the Fourier transform of the space coordinate x.

The Fourier transform of Eqs. (2.10) can be obtained

$$\begin{aligned} &R_{11}\frac{d^2\widetilde{\varphi}_S}{dz^2} + A_{11}\widetilde{\varphi}_S + R_{12}\frac{d^2\widetilde{\varphi}_F}{dz^2} + A_{12}\widetilde{\varphi}_F + R_{13}\frac{d^2\widetilde{\varphi}_I}{dz^2} + A_{13}\widetilde{\varphi}_I = 0 \\ &R_{12}\frac{d^2\widetilde{\varphi}_S}{dz^2} + A_{12}\widetilde{\varphi}_S + R_{22}\frac{d^2}{\widetilde{\varphi}_F}dz^2 + A_{22}\widetilde{\varphi}_F + R_{23}\frac{d^2\widetilde{\varphi}_I}{dz^2} + A_{23}\widetilde{\varphi}_I = 0 \\ &R_{13}\frac{d^2\widetilde{\varphi}_S}{dz^2} + A_{13}\widetilde{\varphi}_S + R_{23}\frac{d^2\widetilde{\varphi}_F}{dz^2} + A_{23}\widetilde{\varphi}_F + R_{33}\frac{d^2\widetilde{\varphi}_I}{dz^2} + A_{33}\widetilde{\varphi}_I = 0 \end{aligned} \tag{3.2}$$

where $A_{11} = -a_{11} - \xi^2 R_{11}$, $A_{12} = -a_{12} - \xi^2 R_{12}$, $A_{13} = -a_{13} - \xi^2 R_{13}$, $A_{22} = -a_{22} - \xi^2 R_{22}$, $A_{23} = -a_{23} - \xi^2 R_{23}$, $A_{33} = -a_{33} - \xi^2 R_{33}$.

The solution of Eqs. (3.2) is assumed as follows

$$[\widetilde{\varphi}_S, \widetilde{\varphi}_F, \widetilde{\varphi}_I]^{\mathrm{T}} = [c^S, c^F, c^I]^{\mathrm{T}} \exp(\lambda z) \tag{3.3}$$

By substituting Eq. (3.3) into Eqs. (3.2), we obtain the following system of linear equations

$$\begin{bmatrix} \lambda^2 R_{13} + A_{13} & \lambda^2 R_{12} + A_{12} & \lambda^2 R_{11} + A_{11} \\ \lambda^2 R_{23} + A_{23} & \lambda^2 R_{22} + A_{22} & \lambda^2 R_{12} + A_{12} \\ \lambda^2 R_{33} + A_{33} & \lambda^2 R_{23} + A_{23} & \lambda^2 R_{13} + A_{13} \end{bmatrix} \begin{bmatrix} c^I \\ c^F \\ c^S \end{bmatrix} = 0 \tag{3.4}$$

When the coefficient matrix determinant is zero, Eq. (3.4) has a non-zero solution, that is

$$\beta_1\lambda^6 + \beta_2\lambda^4 + \beta_3\lambda^2 + \beta_4 = 0 \tag{3.5}$$

where

$$\begin{aligned} \beta_1 &= -R_{11}R_{22}R_{33} + R_{11}R_{23}^2 + R_{12}^2R_{33} - 2R_{12}R_{13}R_{23} + R_{13}^2R_{22} \\ \beta_2 &= -A_{11}R_{22}R_{33} + A_{11}R_{23}^2 + 2A_{12}R_{12}R_{33} - 2A_{12}R_{13}R_{23} - 2A_{13}R_{12}R_{23} + 2A_{13}R_{13}R_{22} \\ &\quad - A_{22}R_{11}R_{33} + A_{22}R_{13}^2 + 2A_{23}R_{11}R_{23} - 2A_{23}R_{12}R_{13} - A_{33}R_{11}R_{22} + A_{33}R_{12}^2 \\ \beta_3 &= -A_{11}A_{22}R_{33} + 2A_{11}A_{23}R_{22} - A_{11}A_{33}R_{22} + A_{12}^2R_{33} - 2A_{12}A_{13}R_{23} - 2A_{12}A_{23}R_{13} \\ &\quad + 2A_{12}A_{33}R_{12} + A_{13}^2R_{22} + 2A_{13}A_{22}R_{13} - 2A_{13}A_{23}R_{12} - A_{22}A_{33}R_{11} + A_{23}^2R_{11} \\ \beta_4 &= -A_{11}A_{22}A_{33} + A_{11}A_{23}^2 + A_{12}^2A_{33} - 2A_{12}A_{13}A_{23} + A_{13}^2A_{22} \end{aligned}$$

The roots of Eq. (2.5), denoted as $\pm\lambda_n$ $(n = 1, 2, 3)$, can be obtained by solving the following equation

$$\lambda_n = \sqrt{d_n}(\mathrm{Re}[\lambda_n] \geqslant 0 \qquad n = 1, 2, 3 \tag{3.6}$$

where d_n are determined by the equation $\beta_1 d_n^3 + \beta_2 d_n^2 + \beta_3 d_n + \beta_4 = 0$.

The general solution of the system of Eqs. (3.2) can be obtained as

$$\begin{aligned} &\widetilde{\varphi}_S = \sum_{n=1}^{3}(D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) \qquad \widetilde{\varphi}_F = \sum_{n=1}^{3}\delta_{pn}^F(D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) \\ &\widetilde{\varphi}_I = \sum_{n=1}^{3}\delta_{pn}^I(D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) \end{aligned} \tag{3.7}$$

where D_i and E_i $(i = 1, 2, 3)$ are undetermined coefficients

$$\begin{aligned} \delta_{pn}^F &= \frac{(R_{11}R_{23} - R_{12}R_{13})d_n^2 + (A_{11}R_{23} - A_{12}R_{13} - A_{13}R_{12} + A_{23}R_{11})d_n + A_{11}A_{23} - A_{12}A_{13}}{(-R_{12}R_{23} + R_{13}R_{22})d_n^2 + (-A_{12}R_{23} + A_{13}R_{22} + A_{22}R_{13} - A_{23}R_{12})d_n - A_{12}A_{23} + A_{13}A_{22}} \\ \delta_{pn}^I &= \frac{(-R_{11}R_{22} + R_{12}^2)d_n^2 + (-A_{11}R_{22} + 2A_{12}R_{12} - A_{22}R_{11})d_n - A_{11}A_{22} + A_{12}^2}{(-R_{12}R_{23} + R_{13}R_{22})d_n^2 + (-A_{12}R_{23} + A_{13}R_{22} + A_{22}R_{13} - A_{23}R_{12})d_n - A_{12}A_{23} + A_{13}A_{22}} \end{aligned}$$

The Fourier transform of Eqs. (2.11) can be obtained

$$\begin{aligned} &\mu_{11}\frac{d^2\widetilde{\psi}_S}{dz^2} + B_{11}\widetilde{\psi}_S + B_{12}\widetilde{\psi}_F + \mu_{13}\frac{d^2\widetilde{\psi}_I}{dz^2} + B_{13}\widetilde{\psi}_I = 0 \\ &B_{21}\widetilde{\psi}_S + B_{22}\widetilde{\psi}_F + B_{23}\widetilde{\psi}_I = 0 \\ &\mu_{13}\frac{d^2\widetilde{\psi}_S}{dz^2} + B_{31}\widetilde{\psi}_S + B_{32}\widetilde{\psi}_F + \mu_{33}\frac{d^2\widetilde{\psi}_I}{dz^2} + B_{33}\widetilde{\psi}_I = 0 \end{aligned} \tag{3.8}$$

where $B_{11} = -a_{11} - \xi^2\mu_{11}$, $B_{12} = -a_{12}$, $B_{13} = -a_{13} - \xi^2\mu_{13}$, $B_{21} = a_{12}$, $B_{22} = a_{22}$, $B_{23} = a_{23}$, $B_{31} = -a_{13} - \xi^2\mu_{13}$, $B_{32} = -a_{23}$, $B_{33} = -a_{33} - \xi^2\mu_{33}$.

The solution of Eqs. (3.8) is assumed as follows

$$[\widetilde{\psi}_S, \widetilde{\psi}_F, \widetilde{\psi}_I]^{\mathrm{T}} = [h^S, h^F, h^I]^{\mathrm{T}} \exp(rz) \tag{3.9}$$

Substituting Eq. (3.9) into Eqs. (3.8), the linear equations are obtained

$$\begin{bmatrix} \mu_{13}r^2 + B_{13} & B_{12} & \mu_{11}r^2 + B_{11} \\ B_{23} & B_{22} & B_{21} \\ \mu_{33}r^2 + B_{33} & B_{32} & \mu_{13}r^2 + B_{31} \end{bmatrix} \begin{bmatrix} h^I \\ h^F \\ h^S \end{bmatrix} = 0 \tag{3.10}$$

When the coefficient matrix determinant is zero, Eq. (9.10) has a non-zero solution, that is

$$\beta_5 r^4 + \beta_6 r^2 + \beta_7 = 0 \tag{3.11}$$

where

$$\begin{aligned} \beta_5 &= -\mu_{11}\mu_{33}B_{22} + \mu_{13}^2 B_{22} \\ \beta_6 &= -\mu_{11}B_{22}B_{33} + \mu_{11}B_{23}B_{32} - \mu_{13}B_{12}B_{23} + \mu_{13}B_{13}B_{22} - \mu_{13}B_{21}B_{32} + \mu_{13}B_{22}B_{31} \\ &\quad - \mu_{33}B_{11}B_{22} + \mu_{33}B_{12}B_{21} \\ \beta_7 &= -B_{11}B_{22}B_{33} + B_{11}B_{23}B_{32} + B_{12}B_{21}B_{33} - B_{12}B_{23}B_{31} - B_{13}B_{21}B_{32} + B_{13}B_{22}B_{31} \end{aligned}$$

The roots of Eq. (3.11), denoted as $\pm r_n$ $(n = 1, 2)$, can be obtained by solving the following equation

$$r_n = \sqrt{t_n}(\mathrm{Re}[r_n] \geqslant 0 \qquad n = 1, 2 \tag{3.12}$$

where t_n are determined from the equation $\beta_5 t_n^2 + \beta_6 t_n + \beta_7 = 0$.

The general solution of the system of Eqs. (3.8) can be obtained as

$$\begin{aligned} &\widetilde{\psi}_S = \sum_{n=1}^{2}(M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z}) \qquad \widetilde{\psi}_F = \sum_{n=1}^{2}\delta_{sn}^F(M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z}) \\ &\widetilde{\psi}_I = \sum_{n=1}^{2}\delta_{sn}^I(M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z}) \end{aligned} \tag{3.13}$$

where M_1, M_2, N_1 and N_2 are undetermined coefficients

$$\delta_{sn}^F = \frac{(\mu_{11}t_n + B_{11})B_{23} - B_{21}(\mu_{13}t_n + B_{13})}{-B_{23}B_{12} + B_{22}\mu_{13}t_n + B_{13})} \qquad \delta_{sn}^I = \frac{(-\mu_{11}t_n + B_{11})B_{22} - B_{12}B_{21}}{-B_{23}B_{12} + B_{22}(\mu_{13}t_n + B_{13})}$$

4. Solution of the dynamic response of the saturated frozen soil foundation

4.1. Dynamic response of the foundation

The displacement component in the right angle coordinate system can be expressed by the potential functions φ and ψ

$$u_x = \frac{\partial \varphi}{\partial x} - \frac{\partial \psi}{\partial z} \qquad u_z = \frac{\partial \varphi}{\partial z} + \frac{\partial \psi}{\partial x} \tag{4.1}$$

Combining Eqs. (2.5) with Eq. (4.1), then substituting into Eqs. (2.4), the constitutive equation for the saturated frozen soil medium are be obtained as follows

$$\begin{aligned} \sigma_{zz}^S &= \Big(K_1 - \frac{2\mu_{11}}{3}\Big)\nabla^2\widetilde{\varphi}_S + C_{12}\nabla^2\widetilde{\varphi}_F + \Big(C_{13} - \frac{\mu_{13}}{3}\Big)\nabla^2\widetilde{\varphi}_I + 2\mu_{11}\Big(\frac{\partial^2\widetilde{\varphi}_S}{\partial z^2} + \frac{\partial^2\widetilde{\psi}_S}{\partial x \partial z}\Big) \\ &\quad + \mu_{13}\Big(\frac{\partial^2\widetilde{\varphi}_I}{\partial z^2} + \frac{\partial^2\widetilde{\psi}_I}{\partial x \partial z}\Big) \\ \sigma_{xz}^S &= \mu_{11}\Big(2\frac{\partial^2\widetilde{\varphi}_S}{\partial x \partial z} + \frac{\partial^2\widetilde{\psi}_S}{\partial x^2} - \frac{\partial^2\widetilde{\psi}_S}{\partial z^2}\Big) + \frac{1}{2}\mu_{13}\Big(2\frac{\partial^2\widetilde{\varphi}_I}{\partial x \partial z} + \frac{\partial^2\widetilde{\psi}_I}{\partial x^2} - \frac{\partial^2\widetilde{\psi}_I}{\partial z^2}\Big) \\ \sigma^F &= C_{12}\nabla^2\widetilde{\varphi}_S + K_2\nabla^2\widetilde{\varphi}_F + C_{23}\nabla^2\widetilde{\varphi}_I \\ \sigma_{zz}^I &= \Big(C_{13} - \frac{\mu_{13}}{3}\Big)\nabla^2\widetilde{\varphi}_S + C_{23}\nabla^2\widetilde{\varphi}_F + \Big(K_3 - \frac{2\mu_{33}}{3}\Big)\nabla^2\widetilde{\varphi}_I + \mu_{13}\Big(\frac{\partial^2\widetilde{\varphi}_S}{\partial z^2} + \frac{\partial^2\widetilde{\psi}_S}{\partial x \partial z}\Big) \\ &\quad + 2\mu_{33}\Big(\frac{\partial^2\widetilde{\varphi}_I}{\partial z^2} + \frac{\partial^2\widetilde{\psi}_I}{\partial x \partial z}\Big) \\ \sigma_{xz}^I &= \mu_{33}\Big(2\frac{\partial^2\widetilde{\varphi}_I}{\partial x \partial z} + \frac{\partial^2\widetilde{\psi}_I}{\partial x^2} - \frac{\partial^2\widetilde{\psi}_I}{\partial z^2}\Big) + \frac{1}{2}\mu_{13}\Big(2\frac{\partial^2\widetilde{\varphi}_S}{\partial x \partial z} + \frac{\partial^2\widetilde{\psi}_S}{\partial x^2} - \frac{\partial^2\widetilde{\psi}_S}{\partial z^2}\Big) \end{aligned} \tag{4.2}$$

Substituting Eqs. (3.7) and (3.13) into Eqs. (4.1) and (4.2) and combining the relationship between the parameters by Qiu *et al.* (2018), the expressions of displacement and stress of each phase in the saturated frozen soil medium in the Fourier transform domain can be obtained

$$
\begin{aligned}
\widetilde{u}_x^S &= i\xi \sum_{n=1}^{3}(D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) + \sum_{n=1}^{2} r_n (M_n \mathrm{e}^{-r_n z} - N_n \mathrm{e}^{r_n z}) \\
\widetilde{u}_z^S &= \sum_{n=1}^{3} -\lambda_n (D_n \mathrm{e}^{-\lambda_n z} - E_n \mathrm{e}^{\lambda_n z}) + \mathrm{i}\xi \sum_{n=1}^{2} (M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z}) \\
\widetilde{u}_z^F &= \sum_{n=1}^{3} -\lambda_n \delta_{pn}^F (D_n \mathrm{e}^{-\lambda_n z} - E_n \mathrm{e}^{\lambda_n z}) + \mathrm{i}\xi \sum_{n=1}^{2} \delta_{sn}^F (M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z}) \\
\widetilde{u}_x^I &= i\xi \sum_{n=1}^{3} \delta_{pn}^I (D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) + \sum_{n=1}^{2} r_n \delta_{sn}^I (M_n \mathrm{e}^{-r_n z} - N_n \mathrm{e}^{r_n z}) \\
\widetilde{u}_z^I &= \sum_{n=1}^{3} -\lambda_n \delta_{pn}^I (D_n \mathrm{e}^{-\lambda_n z} - E_n \mathrm{e}^{\lambda_n z}) + \mathrm{i}\xi \sum_{n=1}^{2} \delta_{sn}^I (M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z})
\end{aligned}
\tag{4.3}
$$

and

$$
\begin{aligned}
\widetilde{\sigma}_{xz}^S &= \mathrm{i}\xi \sum_{n=1}^{3} -\lambda_n (2\mu_{11} + \mu_{13}\delta_{pn}^I)(D_n \mathrm{e}^{-\lambda_n z} - E_n \mathrm{e}^{\lambda_n z}) \\
&\quad - \sum_{n=1}^{2} (r_n^2 + \xi^2)\Big(\mu_{11} + \frac{1}{2}\mu_{13}\delta_{sn}^I]\Big)(M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z}) \\
\widetilde{\sigma}_{zz}^S &= \sum_{n=1}^{3} \chi_n (D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) - \mathrm{i}\xi \sum_{n=1}^{2} r_n (2\mu_{11} + \mu_{13}\delta_{sn}^I)(M_n \mathrm{e}^{-r_n z} - N_n \mathrm{e}^{r_n z}) \\
\widetilde{\sigma}_{zz}^F &= \sum_{n=1}^{3} (R_{12} + R_{22}\delta_{pn}^F + R_{23}\delta_{pn}^I)(\lambda_n^2 - \xi^2)(D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) \\
\widetilde{\sigma}_{xz}^I &= \mathrm{i}\xi \sum_{n=1}^{3} -\lambda_n (2\mu_{33}\delta_{pn}^I + \mu_{13})(D_n \mathrm{e}^{-\lambda_n z} - E_n \mathrm{e}^{\lambda_n z}) \\
&\quad - \sum_{n=1}^{2} (r_n^2 + \xi^2)\Big(\mu_{33}\delta_{sn}^I + \frac{1}{2}\mu_{13}\Big)(M_n \mathrm{e}^{-r_n z} + N_n \mathrm{e}^{r_n z}) \\
\widetilde{\sigma}_{zz}^I &= \sum_{n=1}^{3} \alpha_n (D_n \mathrm{e}^{-\lambda_n z} + E_n \mathrm{e}^{\lambda_n z}) - \mathrm{i}\xi \sum_{n=1}^{2} r_n (\mu_{13} + 2\mu_{33}\delta_{sn}^I)(M_n \mathrm{e}^{-r_n z} - N_n \mathrm{e}^{r_n z})
\end{aligned}
\tag{4.4}
$$

where for $n = 1, 2, 3$

$$
\begin{aligned}
\chi_n &= [(R_{11} - 2\mu_{11}) + R_{12}\delta_{pn}^F + (R_{13} - \mu_{13})\delta_{pn}^I](\lambda_n^2 - \xi^2) + (2\mu_{11} + \mu_{13}\delta_{pn}^I)\lambda_n^2 \\
\alpha_n &= [(R_{13} - \mu_{13}) + R_{23}\delta_{pn}^F + (R_{33} - 2\mu_{33})\delta_{pn}^I](\lambda_n^2 - \xi^2) + (\mu_{13} + 2\mu_{33}\delta_{pn}^I)\lambda_n^2
\end{aligned}
$$

4.2. Boundary conditions and solution

Considering the boundary condition that the surface of foundation ($z = 0$) is permeable and subjected to a vertical harmonic load at $-l < x < l$, and the bottom surface of bedrock ($z = H$) is fixed and impermeable, we get:

— at $z = 0$

$$
\begin{aligned}
&\widetilde{\sigma}_{zz}^S + \widetilde{\sigma}_{zz}^F + \widetilde{\sigma}_{zz}^I = q_0 \frac{\sin \xi l}{\xi l} \qquad \widetilde{\sigma}_{xz}^S + \widetilde{\sigma}_{xz}^I = 0 \qquad \widetilde{\sigma}_{zz}^F = 0 \\
&\widetilde{u}_z^S = \widetilde{u}_z^I \qquad \widetilde{u}_x^S = \widetilde{u}_x^I
\end{aligned}
\tag{4.5}
$$

— at $z = H$

$$
\widetilde{u}_x^S = 0 \qquad \widetilde{u}_z^S = 0 \qquad \widetilde{u}_z^F = 0 \qquad \widetilde{u}_x^I = 0 \qquad \widetilde{u}_z^I = 0 \tag{4.6}
$$

By substituting Eqs. (4.3) and (4.4) into Eqs. (4.5) and (4.6) boundary conditions, we obtain the following

$$\mathbf{Jx} = \mathbf{f} \tag{4.7}$$

where $\mathbf{J}$ is a 10×10 matrix, and

$$\mathbf{x} = [D_1, D_2, D_3, M_1, M_2, E_1, E_2, E_3, N_1, N_2]^{\mathrm{T}} \qquad \mathbf{f} = \left[q_0 \frac{\sin(\xi l)}{\xi l}, 0, 0, 0, 0, 0, 0, 0, 0, 0\right]^{\mathrm{T}}$$

By solving Eq. (4.7), $\mathbf{x}$ can be obtained. Combined with Eqs. (4.3) and (4.5), the stress and displacement of the saturated frozen soil foundation in the wavenumber domain can be found.

5. Numerical examples

In this Section, the effect of temperature T, porosity ϕ, cementation parameter ϖ, load frequency f, and contact parameter ε on the dynamic response is studied by numerical calculation when the boundary condition is permeable at the surface of the foundation and subjected to the vertical harmonic load, and the bottom is fixed and impermeable.

This paper chooses a collection of physical and mechanical parameters pertaining to the saturated frozen soil foundation, as illustrated in Table 1 (Qiu *et al.*, 2018). Take the load amplitude $q = 1\,\mathrm{kPa}$, saturated frozen soil thickness $H = 20\,\mathrm{m}$, distribution length $l = 1\,\mathrm{m}$. The closed-form solution of the inverse Fourier transform is challenging to obtain due to the intricate nature of the integrand function. Therefore, this paper employs the Fast Fourier Transform (FFT) method to perform the inverse Fourier transform. The wave number is discretized into 1024 points. The vertical and horizontal displacement examples in this paper take the calculation results at the surface of the foundation ($z = 0$)

Table 1. Physical and mechanical parameters of the saturated frozen soil foundation

Material parameters						
Hydrodynamic Viscosity coefficient	Density of soil particle	Density of liquid	Density of particles	Bulk modulus of soil particles	Bulk modulus of liquid	Bulk modulus of ice particle
η_F [Kg/(m s)]	ρ_s [Kg/m^3]	ρ_F [Kg/m^3]	ρ_I [Kg/m^3]	K_S [GPa]	K_F [GPa]	K_I [GPa]
$1.8 \cdot 10^{-3}$	2580	1000	920	20.9	2.25	8.58

Material parameters			
Shear modulus of soil particle	Shear modulus of ice particle	Reference value of soil skeleton permeability coeff.	Reference value of ice permeability coefficient
μ_S [GPa]	μ_I [GPa]	κ_{s0} [m^2]	κ_{i0} [m^2]
6.85	3.32	$1.0 \cdot 10^{-11}$	$5.0 \cdot 10^{-5}$

5.1. Verification

To verify the accuracy of the numerical calculations in this paper, the model is validated by degrading the Lamb problem in which the seismic source is located on the surface of a semi-infinite, single-phase homogeneous medium, and the results are compared with those of Yuan (1999) under the same conditions. A saturated permafrost foundation is degraded to an elastic

foundation when saturation $S_r \to 0$ and porosity $\phi \to 0$. Taking $\omega = 0.01\,\mathrm{rad/s}$, Poisson's ratio $\nu = 0.3$, soil particle density $S_r = 1884\,\mathrm{kg/m^3}$, and modulus of elasticity $E = 1 \cdot 10^7\,\mathrm{Pa}$, Fig. 2 shows the variation of positive stress with depth under the load edge at the surface of the foundation where the bar load is applied. It can be seen that the vertical positive stresses in this paper are in good agreement with the results given by Yuan.

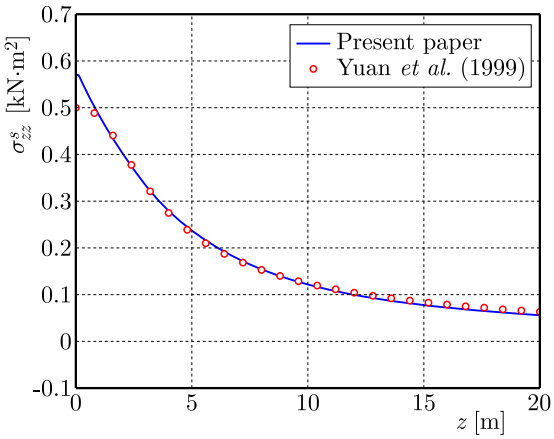


Fig. 2. The normal stress under the edge of the load when the strip load acts on the surface of the foundation

5.2. Effect of temperature

Changes in temperature cause changes in the composition of ice and liquid phases in saturated frozen soil, which in turn cause changes in the bearing capacity of the soil particle phase. Figures 3a and 3b show the effect curves of varying temperatures on the displacement of foundation under the condition of overburdened soil thickness $H = 20\,\mathrm{m}$, porosity $\phi = 0.3$, Poisson's ratio $\nu = 0.3$, frequency $f = 1\,\mathrm{Hz}$, and contact parameter $\varepsilon = 0.5$. As depicted in Figs. 3a

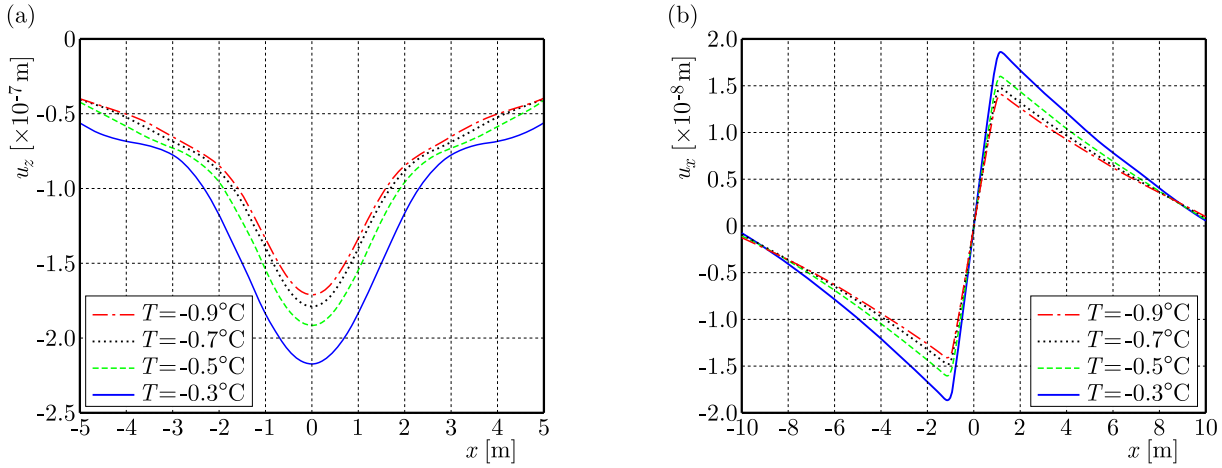


Fig. 3. Variation curves of the vertical and horizontal displacement of the foundation surface with temperature

and 3b, it becomes apparent that as the temperature rises, there is a noticeable escalation in both the magnitude of vertical and horizontal surface displacements. Furthermore, this increase in temperature leads to a gradual augmentation of the displacement amplitudes. This is because as the temperature increases, the ice particle content in the pore space gradually decreases, and the interaction with the soil particle skeleton gradually weakens so that the load borne by the

ice phase of the pore space decreases and the load borne by the soil particle phase increases, which finally makes the vertical displacement and horizontal displacement increase.

5.3. Effect of porosity

The ratio of the volume of connected pores in the unit body to the volume of unit body is the porosity, which is consistent with the actual geotechnical material, and the porosity is taken to be in the range of $\phi = 0.05$-0.50. The other parameters in the numerical calculation are selected as follows: $H = 20\,\mathrm{m}$, $T = -0.5^{\circ}\mathrm{C}$, $\nu = 0.3$, $f = 1\,\mathrm{Hz}$, $\varepsilon = 0.5$, ϕ taken as 0.1, 0.2, 0.3, 0.4, respectively. Figures 4a and 4b give the displacement curves of the saturated frozen soil foundation when the porosity is changed. From Figs. 4a and 4b, it is evident that variations in porosity exert a pronounced influence on both vertical and horizontal displacements of the foundation. Moreover, the displacements demonstrate an escalating trend with increasing porosity. This phenomenon can be attributed to the increase in porosity, which refers to the volume of interconnected pores within the unit body. As porosity increases, the volume of ice particles within the pores remains constant. Consequently, the supporting effect of ice particles on the soil particle framework gradually weakens. This causes an augmentation in the load carried by the soil particle framework, leading to a subsequent increase in the overall soil displacement.

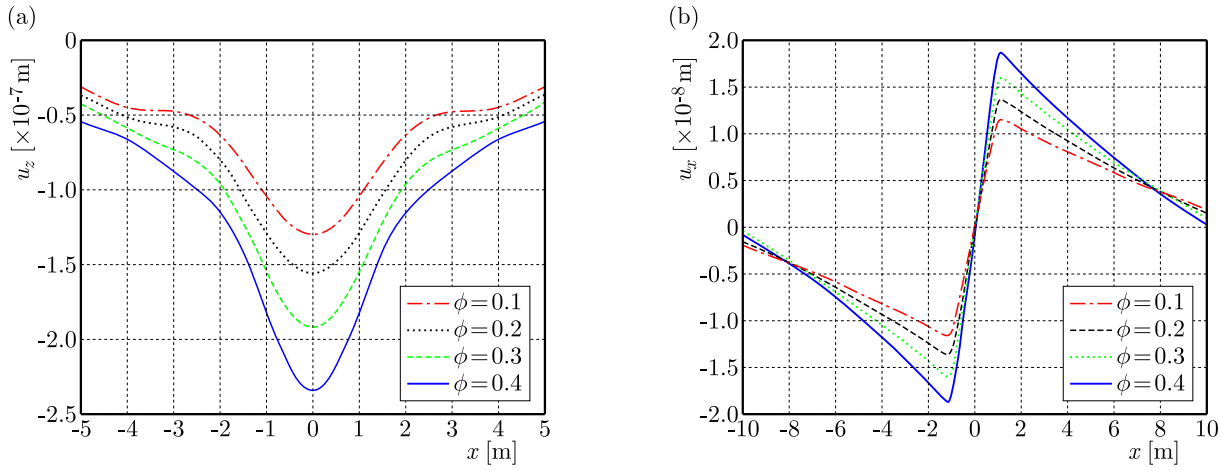


Fig. 4. Variation curves of the vertical and horizontal displacement of the foundation surface with porosity

5.4. Effect of cementation parameters

Imitating the relationship between modulus-porosity-Poisson's ratio of the two-phase medium and ignoring the effect of porosity on Poisson's ratio of the skeleton, Luo (1999) and Liu *et al.* (2015) believed that the cementation parameters can be taken as

$$\varpi = \frac{1+\nu}{2(1-2\nu)} \tag{5.1}$$

where ϖ denotes the level of overall cementation within the skeleton, ranging from 0.5 to infinity. A higher value of ϖ signifies a lower degree of cementation between particles, indicating a decreased bonding strength within the soil particle skeleton. ν denotes Poisson's ratio of the soil particle skeleton.

The range of Poisson's ratio considered in this study is $\nu = 0.100$ to 0.498, within which the cementation parameters exhibit an increasing trend with higher values of Poisson's ratio. Considering the overburdened thickness $H = 20\,\mathrm{m}$, temperature $T = -0.5^{\circ}\mathrm{C}$, porosity $\phi = 0.3$, load frequency $f = 1\,\mathrm{Hz}$, and contact parameter $\varepsilon = 0.5$, Figs. 5a and 5b present the curves of

displacement along the surface when Poisson's ratio ν gradually increases from 0.1 to 0.4, respectively. From these figures, it is evident that changes in Poisson's ratio, which corresponds to variations in the cementation parameter, have a significant impact on the amplitude of vertical and horizontal displacements of the ground surface. As Poisson's ratio increases (indicating an increase in the cementation parameter), both the vertical and horizontal displacements experience amplified amplitudes with the magnitude of displacement amplification gradually intensifying. This law can be explained as follows: an increase in the cementation parameters corresponds to a decrease in the degree of cementation between soil particles in saturated frozen soil and ice particles within the pores. This, in turn, results in a decrease in the skeleton modulus and softening of the soil. As a consequence, there is an increase in displacement observed.

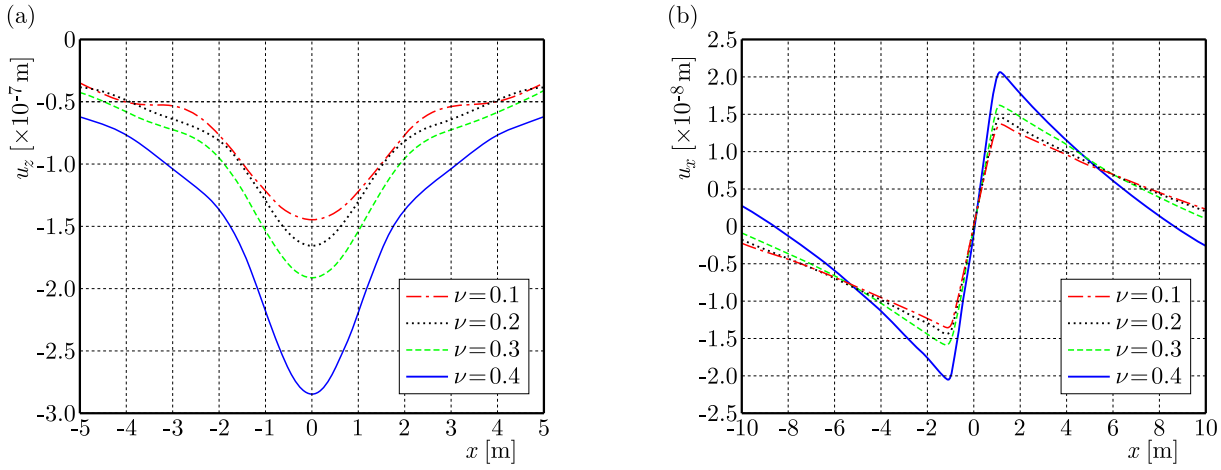


Fig. 5. Variation curves of the vertical and horizontal displacement of the foundation surface with cementation parameters

5.5. Effect of load frequency

Figures 6a and 6b analyze the influence of load frequency on the surface displacement of foundation. The following parameters are employed in the numerical calculations: overburden soil thickness $H = 20\,\mathrm{m}$, temperature $T = -0.5°\mathrm{C}$, Poisson's ratio $\nu = 0.3$, porosity $\phi = 0.3$, contact parameter $\varepsilon = 0.5$. The load frequencies considered are 1 Hz, 10 Hz, 20 Hz, and 30 Hz. From the observations made in Figs. 6a and 6b, it is apparent that as the load frequency increases, both the amplitude of vertical displacement and horizontal displacement of the ground surface also increase. This indicates that with an increase of load frequency, the load borne by the pore water and pore ice phases decreases, and the load borne by the soil particle phase increases.

5.6. Effect of contact parameters

The contact parameter ε characterizes the support of ice in the pore space on the skeleton of soil particles, and its value ranges from 0 to 1. When $\varepsilon = 1$, the ice has the smallest support to the soil particle skeleton, and the ice is suspended in the pores; when $\varepsilon = 0$, the ice has the largest support to the soil particle skeleton. The contact parameter affects the dynamic response of saturated permafrost foundation by influencing the bulk modulus K_{sm} and shear modulus μ_{sm} of the soil particles skeleton, which in turn affects the stiffness parameter. The relationship between the contact parameters and the volume and shear modulus of the skeletal mold of the soil particles is described in detail in Qiu *et al.* (2018). Figures 7a and 7b display the influence of contact parameters on the vertical and horizontal displacements of the surface under specific conditions: overlying soil layer thickness $H = 20\,\mathrm{m}$, Poisson's ratio $\nu = 0.3$, temperature

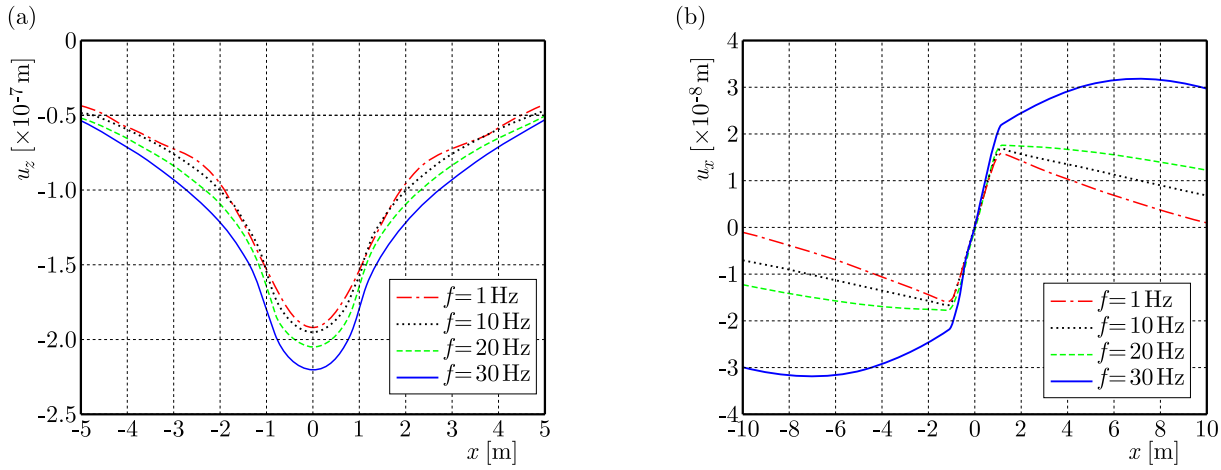


Fig. 6. Variation curves of the vertical and horizontal displacement of the foundation surface with load frequency

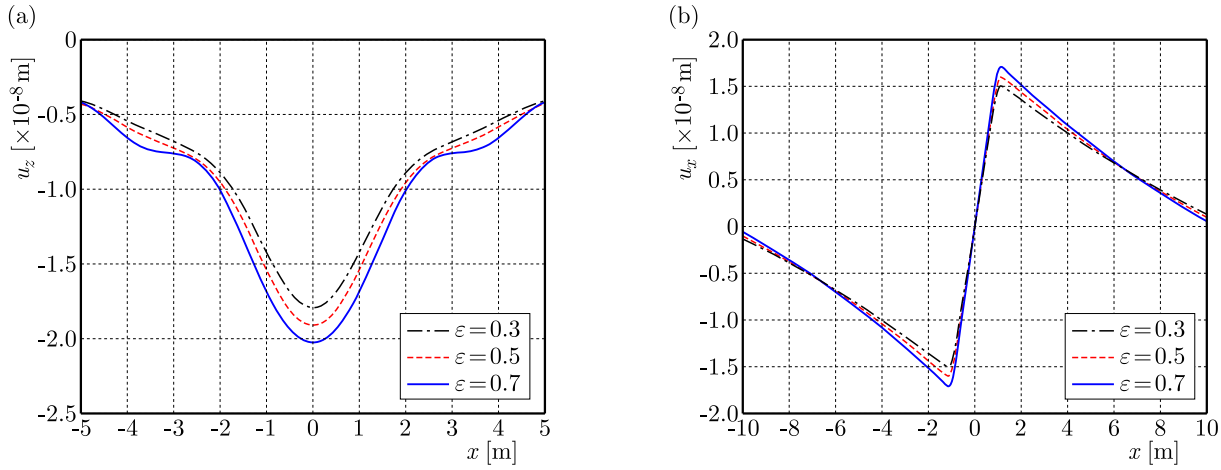


Fig. 7. Variation curves of the vertical and horizontal displacement of the foundation surface with contact parameters

$T = -0.5$°C, porosity $\phi = 0.3$, load frequency $f = 1$ Hz, and contact parameters $\varepsilon = 0.3$, 0.5, and 0.7. From Figs. 7a and 7b, it can be observed that an increase in the contact parameters leads to amplification in both the vertical and horizontal displacement amplitudes of the surface. This phenomenon can be attributed to the enlargement of pores between the ice particles and soil particle skeleton as the contact parameters increase. Consequently, there is a reduction in the friction force and an increase in the relative displacement between the ice and solid phases within the pores.

6. Conclusions

Based on the porous theory of porous solid medium, this paper establishes the governing equation of the two-dimensional dynamic response problem of a saturated frozen soil medium. Using the Fourier transform and boundary conditions, the solutions of the corresponding forces and displacements in the frequency domain of the saturated frozen soil foundation under the underlying bedrock under a harmonic load are obtained. Finally, the FFT method obtains the numerical results, and the influence of various parameters on the dynamic response of saturated frozen soil is analyzed in detail. The following conclusions are obtained:

- The amplitudes of vertical and horizontal displacements on the surface of saturated frozen soil foundations exhibit an increasing trend with elevated temperatures, porosity levels, cementation parameters, load frequencies, and contact parameters.
- The variations in porosity and cementation parameters within a saturated frozen soil foundation have a notable impact on the amplitude of both vertical and horizontal displacements.

Acknowledgments

The authors would like to express their gratitude to the National Natural Science Foundation of China (No. 52168053), the Qinghai Province Science and Technology Department Project (No. 2021-ZJ-922), and the China Postdoctoral Science Foundation (No. 2022MD723837) for their support. Additionally, the authors appreciate the insightful suggestions and comments provided by the reviewers.

References

1. Ai Z.Y., Mu J.J., Ren G.P., 2018, 3D dynamic response of a transversely isotropic multilayered medium subjected to a moving load, *International Journal for Numerical and Analytical Methods in Geomechanics*, **42**, 4, 636-654
2. Biot M.A., 1956, Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range, *The Journal of the Acoustical Society of America*, **28**, 2, 179-191
3. Biot M.A., 1962, Mechanics of Deformation and Acoustic Propagation in Porous Media, *Journal of Applied Physics*, **33**, 4, 1482-1498
4. Carcione J.M., Gurevich B., Cavallini F., 2000, A generalized Biot-Gassmann model for the acoustic properties of shaley sandstones, *Geophysical Prospecting*, **48**, 3, 539-557
5. Carcione J.M., Santos J.E., Ravazzoli C.L., Helle H.B., 2003, Wave simulation in partially frozen porous media with fractal freezing conditions, *Journal of Applied Physics*, **94**, 12, 7839-7847
6. Carcione J.M., Seriani G., 2001, Wave simulation in frozen porous media, *Journal of Computational Physics*, **170**, 2, 676-695
7. Lamb H., 1904, On the Propagation of Tremors over the Surface of an Elastic Solid, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **203**, 359-371, 1-42
8. Leclaire P., Cohen-Ténoudji F., Aguirre-Puente J., 1994, Extension of Biot's theory of wave propagation to frozen porous media, *The Journal of the Acoustical Society of America*, **96**, 6, 3753-3768
9. Leclaire P., Cohen-Ténoudji F., Aguirre-Puente J., 1995, Observation of two longitudinal and two transverse waves in a frozen porous medium, *The Journal of the Acoustical Society of America*, **97**, 4, 2052-2055
10. Liang J., Wu M., Ba Z., Lee V.W., 2020, Transfer matrix solution to free-field response of a multi-layered transversely isotropic poroelastic half-plane, *Soil Dynamics and Earthquake Engineering*, **134**, 106168
11. Liu K., Zhang Z., Pan E., 2022, Dynamic response of a transversely isotropic and multilayered poroelastic medium subjected to a moving load, *Soil Dynamics and Earthquake Engineering*, **155**, 107154
12. Liu Z.J., 2015, Research on wave propagation characteristics and related problems in two-phase porous media, Ph.D. Thesis (in Chinese), Zhejiang University, China
13. Luo J., Stevens R., 1999, Porosity-dependence of elastic moduli and hardness of 3Y-TZP ceramics, *Ceramics International*, **25**, 3, 281-286

14. Ma Q., Huang Y.Y., Zhou F.X., 2023, Dynamic response study of layered unsaturated foundation under moving load based on TRM method (in Chinese), *Engineering Mechanics*, 1-11

15. Ma Q., Shi L.W., 2022, Dynamic response of two-dimensional double-layered unsaturated soil foundations under a strip load, *Journal of Vibration Engineering and Technologies*, **10**, 4, 1221-1233

16. Ma W., Wang X., Wang B., Zhou S., Leong E.-C., Wang C., 2023, Three-dimensional axisymmetric transient response of an unsaturated poroelastic transversely isotropic half-space, *Computers and Geotechnics*, **159**, 105482

17. Qiu H.M., Xia T.D., Zheng Q.Q., Zhou F., 2018, Parametric studies of body wave propagation in saturated frozen soil (in Chinese), *Rock and Soil Mechanics*, **39**, 4053-4062

18. Shi L.W., Ma Q., Ma Y., 2021, Dynamic responses of unsaturated half-space soils to a strip load at different boundary conditions, *Arabian Journal of Geosciences*, **14**, 947

19. Xu M.J., 2010, *Investigation on Dynamic Response of Unsaturated Soils and Foundation* (in Chinese), South China University of Technology, China

20. Xu M.J., Wei D.M., 2011, 3D non-axisymmetrical dynamic response of unsaturated soils (in Chinese), *Engineering Mechanics*, **28**, 3, 78-85

21. Xu M.Q., Jin L.H., Li J.H., 2009, Dynamic responses of a layered saturated soil subjected to harmonic horizontal loads (in Chinese), *Rock and Soil Mechanics*, **30**, 9, 536-540

22. Yuan J.Y., Zhao X.H., 1999, Formulas for calculating stresses in soil subjecting to vertical line load and strip distributed load beneath the surface of ground (in Chinese), *Chinese Quarterly of Mechanics*, **2**, 156-165

23. Zhou B., 2020, Study on ground motion and Rayleigh wave propagation characteristics in frozen soil area, Master's Thesis (in Chinese), Zhejiang Ocean University, China

24. Zhou F.X., Lai Y.M., 2011, Propagation characteristics of elastic waves in saturated frozen soil (in Chinese), *Rock and Soil Mechanics*, **32**, 2669-2674

25. Zhou F.X., Lai Y.M., Ren Y.Y., 2013, An analysis on saturated soil foundation under harmonic loads (in Chinese), *Chinese Journal of Solid Mechanics*, **34**, 5, 536-540

Manuscript received July 20, 2023; accepted for publication September 12, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 683-694, Warsaw 2024
https://doi.org/10.15632/jtam-pl/193746

METHOD FOR DETERMINING THE S-N CURVE FOR A LOW PROBABILITY OF FAILURE

Przemysław Strzelecki, Dariusz Skibicki, Łukasz Pejkowski
Bydgoszcz University of Science and Technology, Faculty of Mechanical Engineering, Bydgoszcz, Poland
corresponding author P. Strzelecki, e-mail: p.strzelecki@pbs.edu.pl

This paper presents a new method for determining the S-N curve for a low probability of failure, e.g., 5%. To apply this method, only eight fatigue tests are needed, which is fewer than standard methods require. This could be achieved because the standard deviation, which is necessary for estimating the normal distribution of fatigue life, was derived from the distribution of logarithm of the yield strength. The tensile tests necessary to get the yield strength are relatively simple and cost-effective. Verification of the method was performed for fatigue tests on S355J2+C structural steel, 1.4301 and 1.4404 stainless steels, medium carbon steel C45 and AW 6063 & AW 2017A aluminium alloys. The results showed that the proposed method gave fatigue strength for 5% failure probability with a more reliable fatigue life than the S-N curve estimated according to ASTM E-739-10, 2015. Considering that the proposed method is conservative and low-cost, it can be used in engineering practice.

Keywords: S-N curve, high-cycle fatigue, probability of failure, normal distribution, scatter of fatigue tests

Nomenclature

F_p – F-distribution value with desired confidence interval p for $n_1 = 2$ and $n_2 = k - 2$
N – number of cycles
K_{ASTM} – confidence band acc. ASTM standard
S_a, S_u, S_y – stress amplitude, ultimate tensile stress, yield stress, respectively
b, b_{inv} – intercept coefficient of S-N curve for relationship $\log(S_a) \sim \log(N)$ and $\log(N) \sim \log(S_a)$, respectively
f_X, f_Y – normal distribution of base and inverted function
$g(x), h(x)$ – base and inverted function
k – number of observed data
m, m_{inv} – slope coefficient of S-N curve for relationship $\log(S_a) \sim \log(N)$ and $\log(N) \sim \log(S_a)$, respectively
σ – standard deviation of fatigue life for small number of specimens
$\sigma_N, \sigma_R, \sigma_S$ – standard deviation of fatigue life, yield stress and fatigue strength, respectively

Indexes: i – i-th number of sample, $\overline{(\cdot)}$ – mean value, $\widehat{(\cdot)}$ – estimated value.

1. Introduction

In the design of a new machine the fatigue strength of the entire service life must be known. For this purpose, the stress-number of cycles curve (S-N curve), referring to a maximum 5% probability of failure, can be used (PN-EN 13749, 2011). Fatigue tests necessary for getting S-N curves are time-consuming, need specialised equipment and skilled personnel. For example,

to obtain 10^5 cycles at a load frequency of 30 Hz, a test takes 56 minutes. However, to obtain 10^6 cycles, the test will take over 9 hours. The time of sample preparation, installation in the machine holder, setting of research parameters, etc. is not included. The costs of performing fatigue testing were calculated by Shen, 1994 to be from 500 up to 1000 dollars per specimen. Therefore, developing fatigue characteristics is very costly. For this reason, normative documents specify a minimum number of specimens necessary to determine reliable fatigue characteristics ASTM E-739-10, 2015. Unfortunately, still a significant number of samples is required. According to the standard ASTM E-739-10, 2015, 12-24 samples for reliability and design purposes and 6-12 for preliminary and exploratory tests are required.

Several analytical methods for determining S-N curves, based on tensile testing, have been developed to reduce the number of specimens needed while maintaining the reliability of the characteristics. These methods assume that S-N curves can be achieved by correlation with tensile test parameters, such as yield strength, tensile strength, or hardness (Lee *et al.*, 2005). Verification of these methods can be found in papers by Strzelecki and Sempruch (2016). According to the verification results, the fatigue life prediction error can be as high as 263%. This is because the estimation of the "fatigue limit" is done by multiplying the tensile strength or hardness by a factor appropriate for the material and assuming a constant slope factor of the regression line. "Fatigue limit" derived from tensile tests can exhibit a large scatter, as shown in Pang *et al.* (2014). Also, the slope coefficient has a large variation according to Goedel *et al.* (2018). Analytical-experimental methods have been developed, e.g., Goedel *et al.* (2018), Strzelecki and Sempruch (2016) to improve the accuracy of S-N curves. In these methods, some parameters must be additionally determined experimentally, e.g., "fatigue limit" can be determined by the Locati method. However, these methods still give a large error of up to 50% (Strzelecki *et al.*, 2015).

Another approach to the formulation of S-N curves is represented by methods based on statistical analysis of fatigue life test results. These methods utilise the relationship between the number of samples used to create an S-N curve and the level of failure probability predicted by the curve. Gope (1999) considering the number of specimens at each stress level, determined the failure probability level and confidence level of the S-N curve. He estimated that for a 10% probability of failure and a 90% confidence level, 10 samples are required. Lewis and Sadhasivini (2004) proposed 7 samples for a two-parameter Weibull distribution for 5% probability and a 95% confidence level at each stress level. Later, Gope (2012) estimated the minimum number of samples for S-N curve estimation as 13 samples. Soh Fotsing *et al.* (2010) suggested 7 samples for a 50% failure probability. As seen, there is no consensus on a guideline specifying the number of samples for a given level of damage probability. For this reason, many methods have been proposed to improve the estimation of the S-N curve from the reliability point of view.

One of these propositions is the backward statistical inference method presented in Xie *et al.* (2014). The method assumes that the standard deviation is constant for each stress amplitude and that the curves of the probability density function of fatigue life at each i-th stress level are the same. An improvement of the backward statistical inference method has been proposed by Li *et al.* (2020) using the modified distribution coefficients. Both methods, original and improved, require as many as 15 specimens to establish a fatigue curve. Bai *et al.* (2019) presented a new method with the similar assumption that the curves of the fatigue life probability density function at each i-th stress level were the same. An additional assumption was that the coefficient of variation had to be also the same for each stress level. This method gives similar results to the backward statistical inference method, and the improved backward statistical inference method. Liu and Sun (2020) assumed that a linear regression distribution could be obtained using non--invasive polynomial chaos. Their study plan required 16 specimens. Zu *et al.* (2020) proposed a α-S-N method based on uncertainty theory. This method gives better results than the standard ISO-12107 (2012) but still requires at least 15 specimens. In many research studies, at least three

S-N curves must be estimated for the tested material, like in multiaxial fatigue, which gives over 45 specimens.

All P-S-N methods presented above use statistical methods to improve probability calculations. In this respect, the work by Strzelecki (2021) stands out, which describes a method of increasing the accuracy of fatigue characteristics by using tensile test data. In that study, the S-N curve for the relationship $\log(S_a) \sim \log(N)$ was estimated. However, in normative documents (e.g. ASTM E-739-10, 2015) the inverse relation $\log(N) \sim \log(S_a)$ is required, as the number of cycles N varies depending on the stress amplitude S_a. Therefore, the use of the relationship $\log(S_a) \sim \log(N)$ should be considered formally incorrect.

The goal of this work is to create a new method for determining P-S-N fatigue characteristics with a low failure probability (e.g. 5%). This was done using a small number of fatigue tests and an additional series of monotonic tensile tests. The deficiency of the earlier proposal by Strzelecki (2021) was corrected in such a way that the S-N median curve was determined based on fatigue tests of a limited number of samples, and the standard deviation of fatigue life was replaced by the value from the tensile test σ_R. A value of fatigue life standard deviation σ_N was obtained by the inverse of the yield stress distribution.

2. Proposed method

The S-N curve can be described using the Basquin equation, as proposed by ASTM E-739-10 (2015)

$$\log(S_{ai}) = m\log(N_i) + b \tag{2.1}$$

The least squares method is recommended for estimating the S-N curve parameters ASTM E-739-10 (2015). The resulting estimated curve follows a normal distribution and is expressed as

$$f(\log(S_{ai})) = \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\Big(\frac{[\log(S_{ai}) - (m\log(N_i) + b)]^2}{\sigma_S^2}\Big) \tag{2.2}$$

The experimental results of the yield stress determination can also be expressed by the normal distribution function (Fig. 1)

$$f(\log(S_y)) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\Big(\frac{[\log(S_{yi}) - \overline{\log(S_y)}]^2}{\sigma_R^2}\Big) \tag{2.3}$$

The proposed method assumes that the S-N curve for 50% probability of failure is estimated for a small number of specimens, e.g. 8 (Eq. (2.2)), and the standard deviation is taken from the distribution of yield strength (Eq. (2.3)). Such a distribution was written below and was named P-S-N (Fig. 1)

$$f(\log(S_{ai})) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\Big(\frac{[\log(S_{ai}) - (m\log(N_i) + b)]^2}{\sigma_R^2}\Big) \tag{2.4}$$

Because, the number of cycles N is a dependent variable, so Eq. (2.1) must be given by an inverse function. The inverse of function (2.1) is as follows

$$\log(N_i) = \frac{1}{m}[\log(S_{ai}) - b] \qquad \log(N_i) = m_{inv}\log(S_{ai}) + b_{inv} \tag{2.5}$$

Let us assume that the base function is $g(x)$ and the inverted function can be denoted as $h(y)$. If $g(x)$ is monotonic and differentiable, one can get a distribution for the inverted function. It must be calculated distribution acc. to the following equation, Walpole *et al.* (2012)

$$f_Y(y) = f_X(h(y))\left|\frac{dx(y)}{dy}\right| \tag{2.6}$$

Calculating the inverse distribution according to Eq. (2.6) and assuming linear regression according to Eq. (2.1), the following equation is obtained

$$f_Y(y) = -\frac{1}{m} f_X\Big(\frac{y-b}{m}\Big) \tag{2.7}$$

where: f_Y – normal distribution of the inverted function, f_X – normal distribution of the base function.

After substituting Eq. (2.7) into Eq. (2.4), the resulting equation is as follows

$$f(\log(N_i)) = \frac{1}{m_{inv}\sqrt{2\pi\sigma_R^2}} \exp\Big(\frac{[\log(N_i) - (m_{inv}\log(S_{ai}) + b_{inv})]^2}{m_{inv}^2\sigma_R^2}\Big) \tag{2.8}$$

The following substitution can be made in Eq. (2.8)

$$\sigma_N = m_{inv}\sigma_R \tag{2.9}$$

Then, Eq. (2.8) has the following form (Fig. 1)

$$f(\log(N_i)) = \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\Big(\frac{[\log(N_i) - (m_{inv}\log(S_{ai}) + b_{inv})]^2}{\sigma_N^2}\Big) \tag{2.10}$$

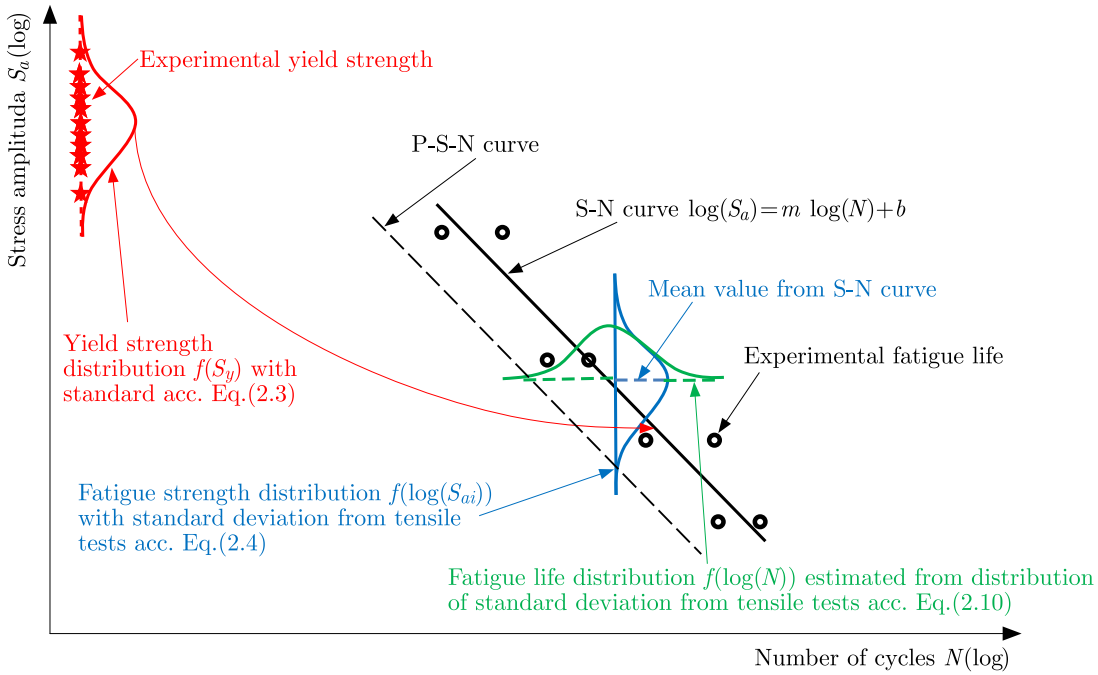


Fig. 1. Scheme of an S-N curve obtained by the proposed method

3. Method proposed in ASTM standard

According to standard ASTM E-739-10 (2015), the confidence band of median S-N curve can be estimated for desired probability. It is assumed that the logarithm of cycles N is a dependent variable and the logarithm of stress amplitude S_a is independent. For this statement, Eq. (2.5) is valid. The following equation is proposed for estimating the P-S-N curve

$$\log(N) = \widehat{m_{inv}} \log(S_a) + \widehat{b_{inv}} \pm K_{ASTM} \tag{3.1}$$

where

$$K_{ASTM} = \sqrt{2F_p}\sigma \sqrt{\frac{1}{k} + \frac{[\log(S_a) - \overline{\log(S_a)}]^2}{\sum_{i=1}^{k}[\log(S_{ai}) - \overline{\log(S_a)}]^2}} \tag{3.2}$$

where: F_p – F-distribution value with the desired confidence interval p for $n_1 = 2$ and $n_2 = k-2$, k – number of observed data, σ – standard deviation for a small number of specimens.

The positive value of K_{ASTM} in Eq. (3.1) is for the upper confidence band, and the negative value is for the lower confidence band. Because the value of stress amplitude S_a in Eq. (3.2) is substituted for each level of the load, K_{ASTM} must be calculated for each stress level separately, Lee *et al.* (2005). A line regression is provided through the estimated values of KASTM. Scheme of the ASTM method was presented in Fig. 2.

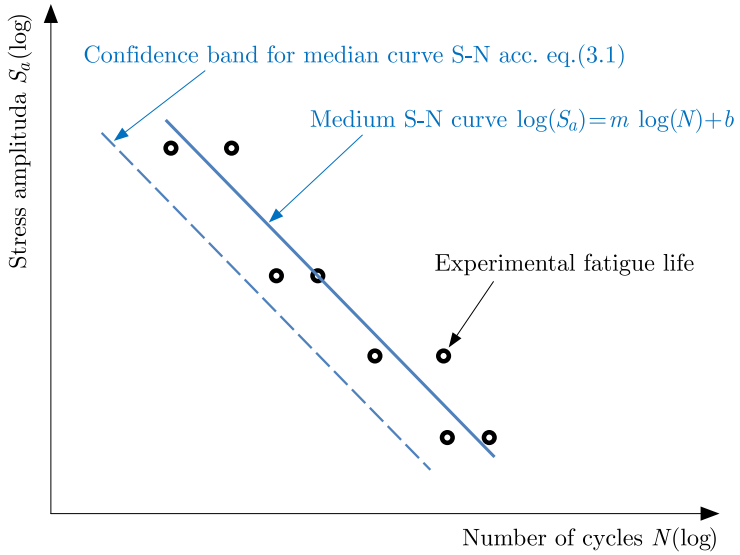


Fig. 2. Scheme of an S-N curve obtained by the ASTM method

4. Verification of the proposed method

The use of the proposed method requires knowledge of histograms and distributions of the yield strength. The tensile tests were performed by Instron 8874 testing machine for S355J2+C (Strzelecki, 2018), 1.4301 (Strzelecki, 2021), 1.4404 (Skibicki *et al.*, 2014) amd AW 6063 T6 (Strzelecki and Wachowski, 2022). The histograms and distributions of yield strength were presented in Figs. 3c-6c and Table 1. The Shapiro-Wilk test of normality was performed for those data. The hypothesis of normality of the data can be rejected if p-value is higher than 0.1 acc. to R Core Team (2023). Whereas the tensile test for C45 and AW 2017A T4 was performed by Instron 8501. Unfortunately, the histograms of yield strength for C45 and AW 2017A T4 are

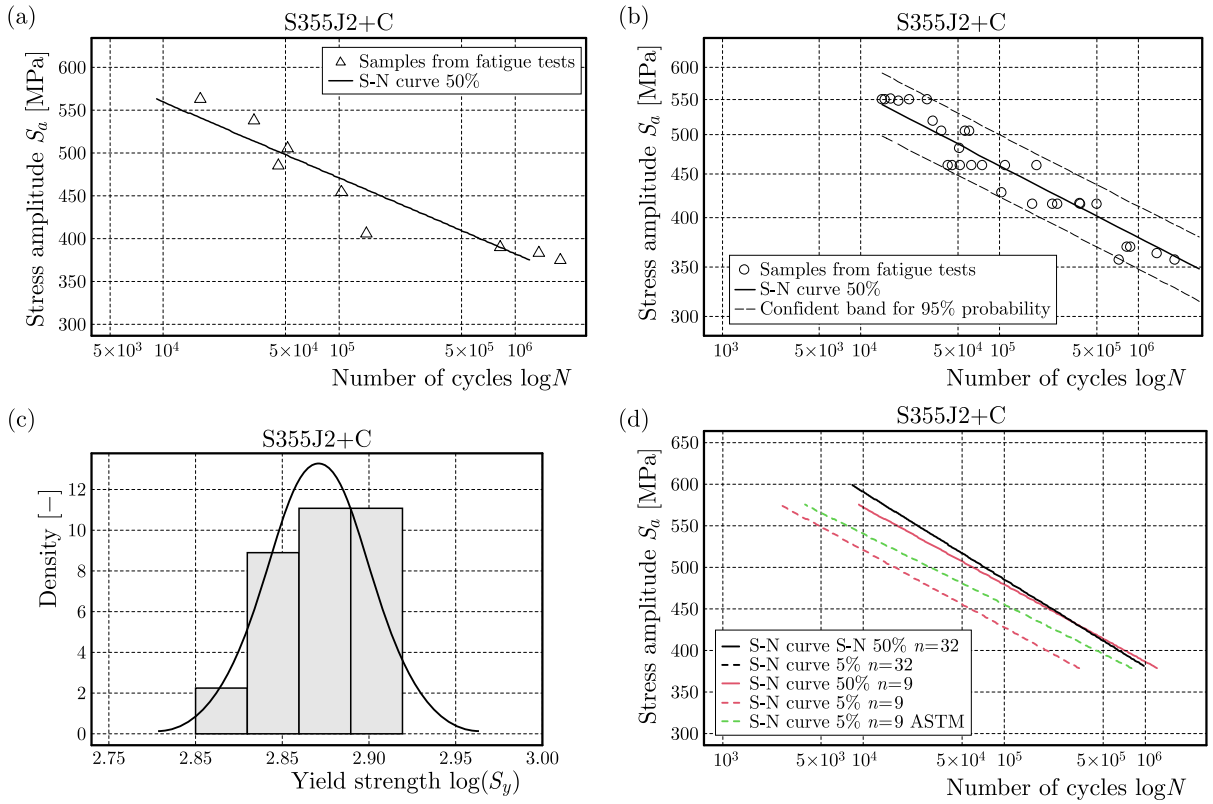


Fig. 3. S-N curves for S355J2+C steel for (a) 9 specimens and (b) 32 specimens, (c) histogram of the yield strength with normal distribution and (d) S-N curves on one diagram for 32 specimens and 9 specimens with standard deviation from the tensile test

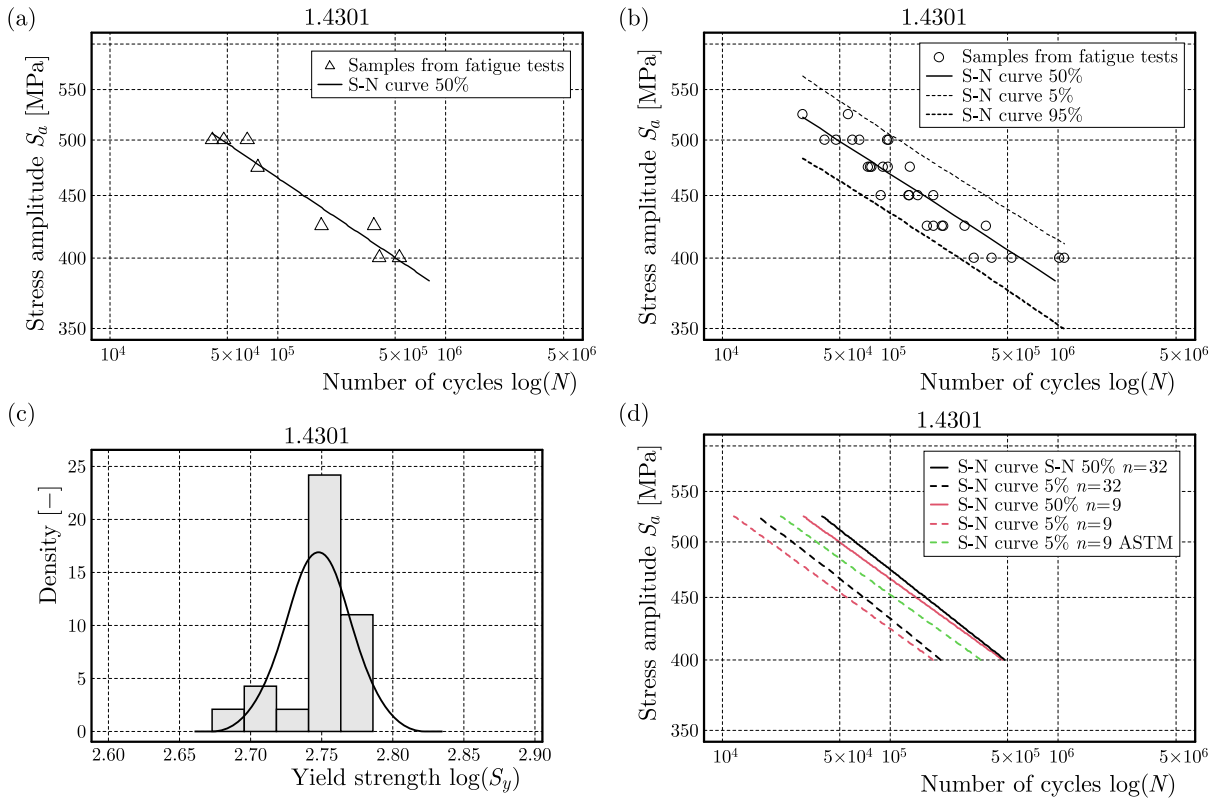


Fig. 4. S-N curves for 1.4301 steel for (a) 9 specimens and (b) 32 specimens, (c) histogram of the yield strength with normal distribution and (d) S-N curves on one diagram for 32 specimens and 9 specimens with standard deviation from the tensile test

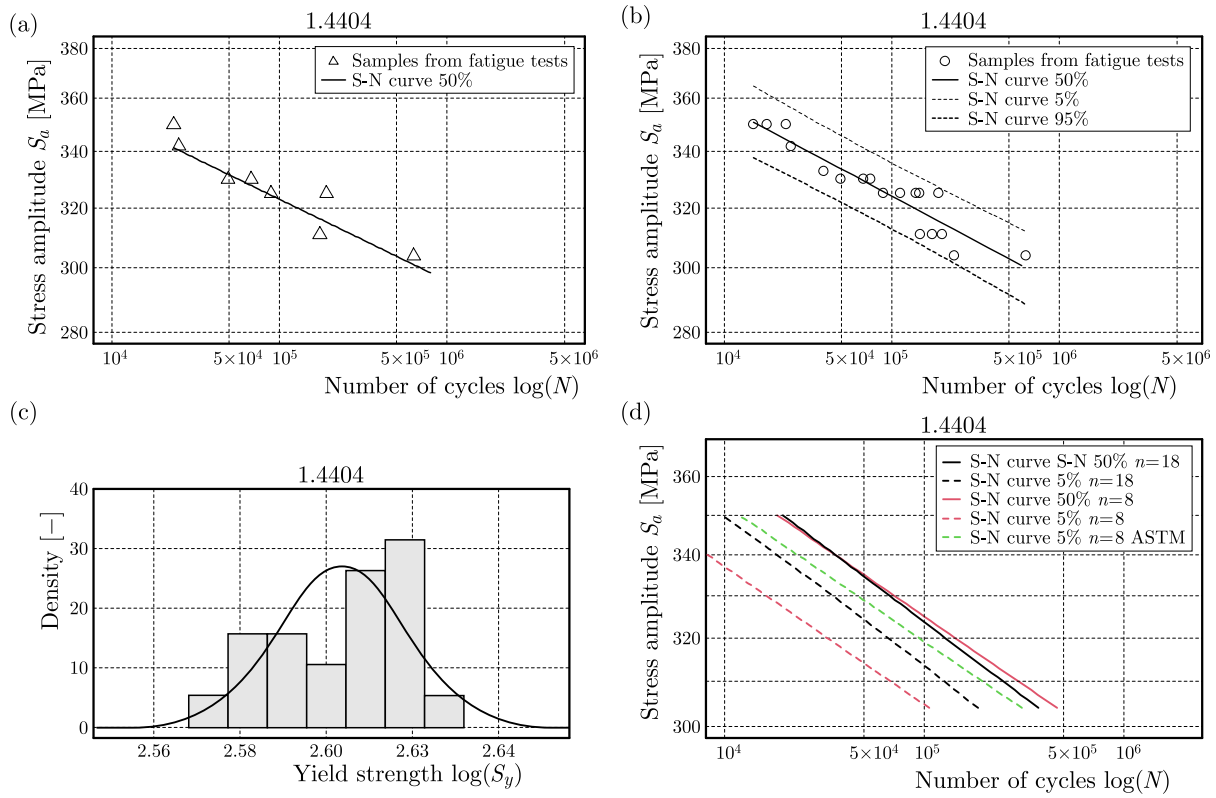


Fig. 5. S-N curves for 1.4404 steel for (a) 8 specimens and (b) 18 specimens, (c) histogram of the yield strength with normal distribution and (d) S-N curves on one diagram for 18 specimens and 8 specimens with standard deviation from the tensile test

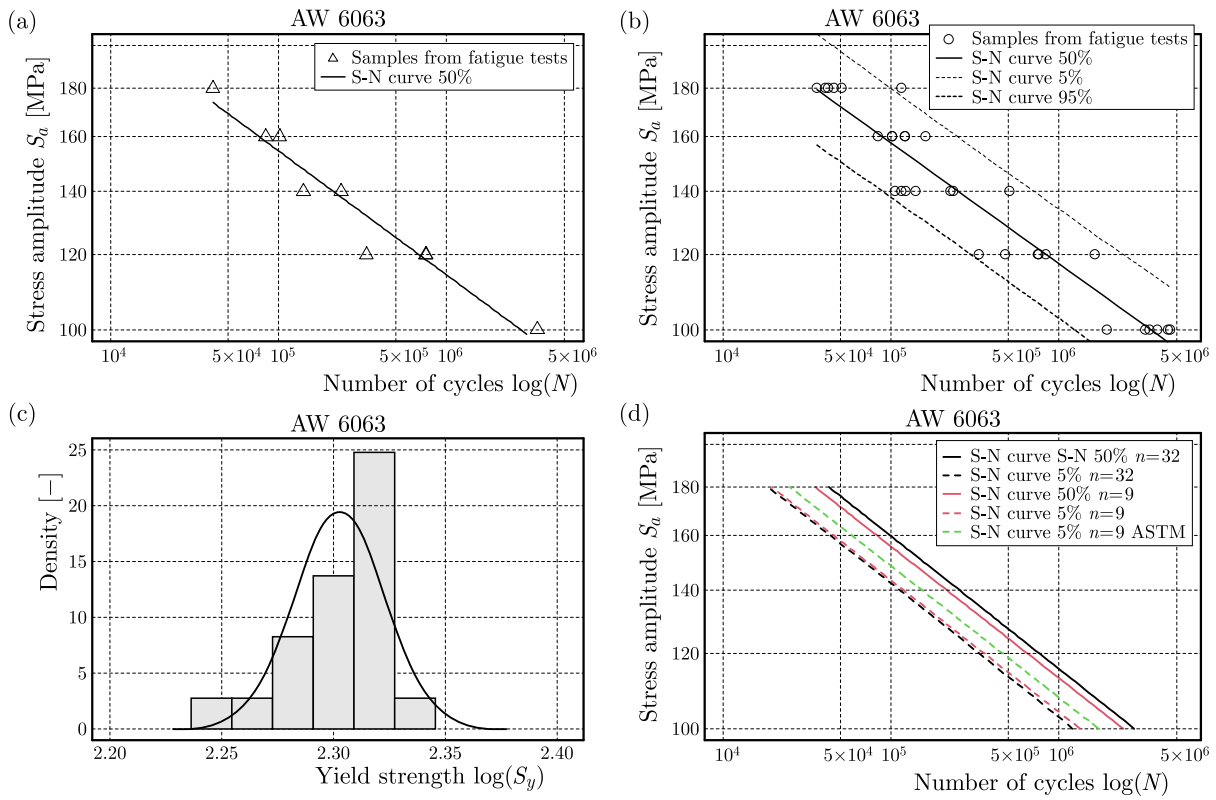


Fig. 6. S-S-N curves for AW 6063 T6 steel for (a) 9 specimens and (b) 32 specimens, (c) histogram of the yield strength with normal distribution and (d) S-N curves on one diagram for 32 specimens and 9 specimens with standard deviation from the tensile test

Table 1. Value of parameters for the distribution acc. to Eq. (2.3)

Material	Specimens No.	$\overline{\log(S_y)}$	σ_R	Shapiro-Wilk statistic W	Shapiro-Wilk p-value
1	2	3	4	5	6
S355J2+C (Strzelecki, 2018)	15	2.871	0.030	0.977	0.08
1.4301 (Strzelecki, 2021)	20	2.748	0.023	0.886	0.02
1.4404 (Skibicki *et al.*, 2014)	21	2.603	0.014	0.915	0.07
C45 (Ligaj and Szala, 2013)	–	2.633	0.045	–	–
AW 6063 (Strzelecki and Wachowski, 2022)	20	2.302	0.021	0.884	0.02
AW 2017A (Ligaj and Szala, 2013)	–	2.499	0.028	–	–

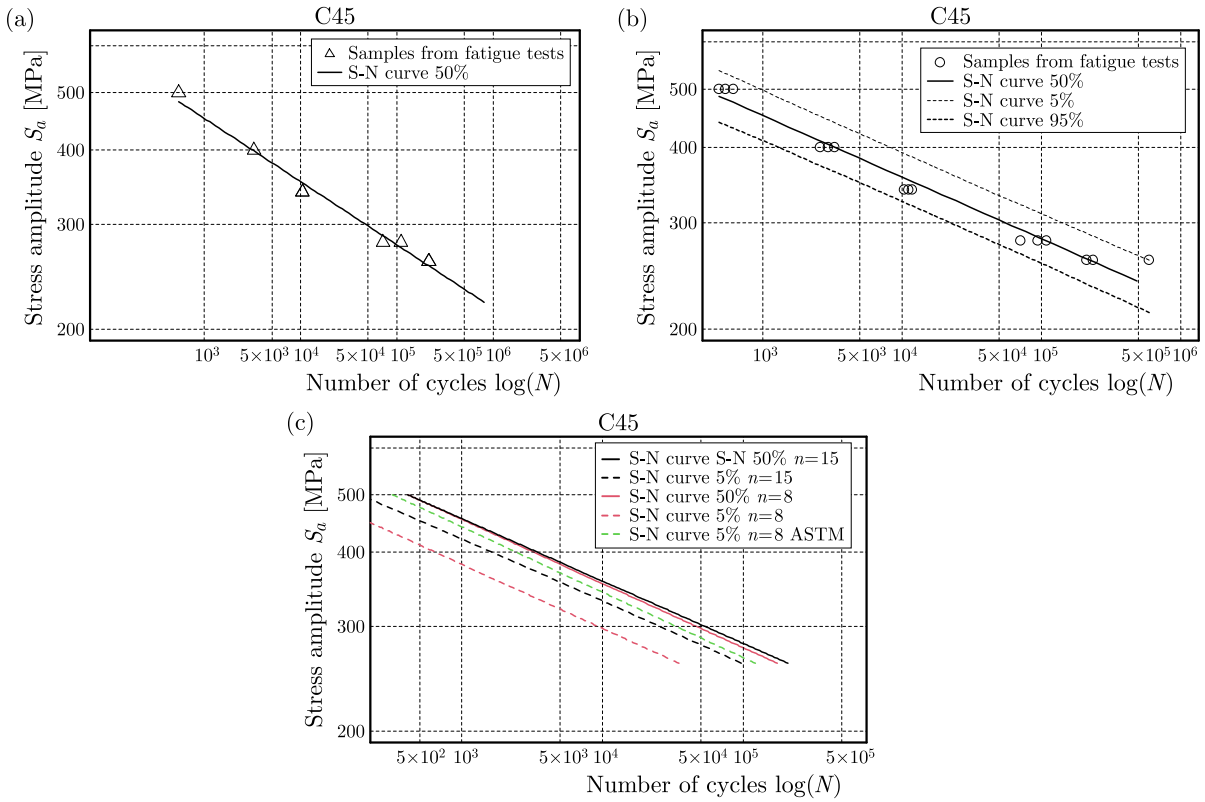


Fig. 7. S-N curves for C45 steel for (a) 8 specimens and (b) 15 specimens, (c) and (d) S-N curves on one diagram for 15 specimens and 8 specimens with standard deviation from the tensile test

not presented, because only the value of standard deviation is known from the literature. All calculations were made in R software ver. 4.3.1 (R Core Team, 2023).

The proposed method was verified for six structural materials, four steels: S355J2+C (Strzelecki, 2018), 1.4301 (Strzelecki, 2021), 1.4404 (Skibicki *et al.*, 2014), C45 (Ligaj and Szala, 2013), and two aluminium alloys: AW 6063 T6 (Strzelecki and Wachowski, 2022) and AW 2017A T4 (Ligaj and Szala, 2013). Fatigue tests were performed by rotating the bending machine for S355J2+C, 1.4301 and AW 6063 T6. Loads were applied with 28.5 Hz frequencies for S355J2+C and 50 Hz frequency for 1.4301 and AW 6063. The fatigue test of stainless steel 1.4404 was performed by fatigue machine Instron 8874. Instron 8501 was used to test the rest two materials: C45, AW 2017A T4. Axial loading was applied with a stress ratio $R = 1$. Test results are presented in Figs. 3b-8b.

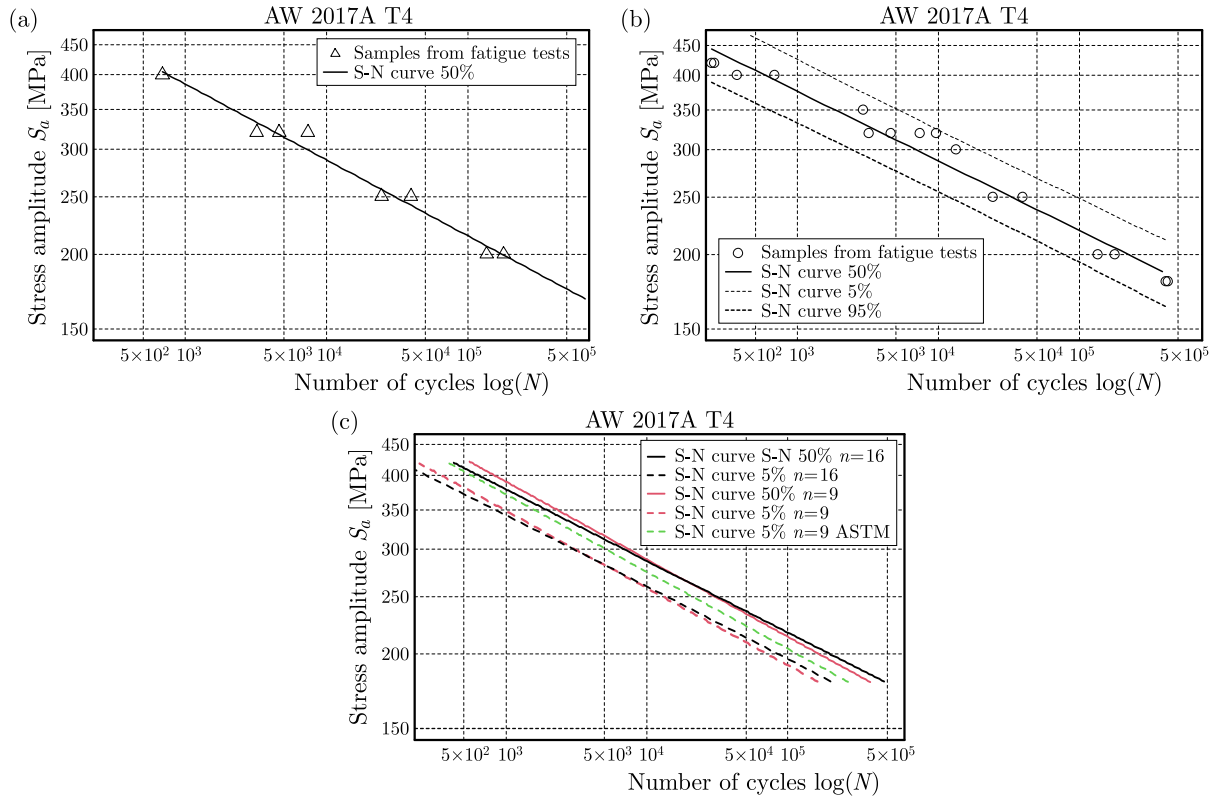


Fig. 8. S-N curves for AW 2017A T4 for (a) 9 specimens and (b) 16 specimens, (c) and (d) S-N curves on one diagram for 16 specimens and 9 specimens with standard deviation from the tensile test

Table 2. Value of parameters for the estimated S-N curves

Material	Type of load	Parameter	Standard method	Method with small no. of specimens	Proposed method	$\mathcal{A}$	$\mathcal{B}$
1	2	3	4	5	6	7	8
S355J2+C (Strzelecki, 2018)	Rotating bending	m_{inv}	−9.917	−9.922	−9.922	36.32%	48.18%
		b_{inv}	31.325	31.300	31.300		
		σ_N	0.1748	0.2745	0.3373		
1.4301 (Strzelecki, 2021)	Rotating bending	m_{inv}	−9.190	−10.041	−10.041	−83.42%	15.04%
		b_{inv}	29.595	31.795	31.795		
		σ_N	0.2135	0.1164	0.2513		
1.4404 (Skibicki *et al.*, 2014)	Axial	m_{inv}	−20.952	−22.890	−22.890	8.79%	59.19%
		b_{inv}	57.594	62.501	62.501		
		σ_N	0.1587	0.1740	0.3889		
C45 (Ligaj and Szala, 2013)	Axial	m_{inv}	−9.526	−9.256	−9.256	−41.88%	58.58%
		b_{inv}	28.326	27.595	27.595		
		σ_N	0.1738	0.1225	0.4196		
AW 6063*	Rotating bending	m_{inv}	−7.108	−7.176	−7.176	−33.70%	−30.89%
		b_{inv}	20.664	20.735	20.735		
		σ_N	0.2051	0.1534	0.1567		
AW 2017A (Ligaj and Szala, 2013)	Axial	m_{inv}	−8.288	−7.691	−7.691	−50.72%	10.63%
		b_{inv}	24.375	22.918	22.918		
		σ_N	0.1985	0.1317	0.2221		

$\mathcal{A}$ – Difference standard method – small no. of specimens $E = (\sigma_{N5} - \sigma_{N4})/\sigma_{N5}$

$\mathcal{B}$ – Difference standard method – proposed method $E = (\sigma_{N6} - \sigma_{N4})/\sigma_{N6}$

* (Strzelecki and Wachowski, 2022)

The estimated parameters for Eq. (2.5) are shown in Table 2. The least squares method was used to determine the parameters for median S-N curves. The S-N curves obtained acc. to ASTM E-739-10 (2015) were named "standard method". In column 4, the parameters for "standard method" for reliability and design purposes were placed. That case involved testing at least 15 specimens for each material. In column 5, the parameters for preliminary tests were placed. In that case a small number, namely 8-9 specimens, were tested. Such values were chosen, because they are in the middle range of preliminary tests according to ASTM E-739-10 (2015). Additionally, standard ISO-12107 (2003) requires a minimum of eight specimens for exploratory testing.

As seen, there are little differences between the "standard method" for reliability and design purposes versus preliminary tests. The biggest differences are for standard deviation σ_N. The parameters for the proposed method, where the standard deviation was calculated according to Eq. (2.9), are shown in the 6th column of Table 2. In the 7th and 8th columns, differences between the standard method, method with a small number of specimens and the proposed method are presented. In all cases, the proposed method gets higher values of standard deviation than the small number of specimens method. This means that the method is conservative, except for material AW6063. However, the proposed method got higher values of standard deviation than the method with a small number of specimens.

5. Discussion

It can be seen in Figs. 3d-6d and Figs. 7c-8c that the S-N found by the proposed method is shifted and has a different angle compared to the S-N curve acc. to the standard method for all materials. This is expressed in different values of parameters m_{inv} and b_{inv}, as shown in Table 2. However, these differences are little and are expected according to widely scattered fatigue tests results. Dispersion of the fatigue test results is caused by microstructural inhomogeneity in material properties, geometry of specimen, differences in the surface roughness, test conditions, environment, and personal aspects (skill of laboratory technicians).

The influence of surface roughness was investigated by Nanninga and White (2009). They found a little difference in the fatigue life for roughness line in the transverse and longitudinal direction. However, they stated that "while it is probably not statistically significant, some polished transverse specimens exhibited fatigue lives that were higher than those of the longitudinal specimens". Additionally, Kurek *et al.* (2017) found that the type of load causes a different value of standard deviation. They stated that the lower scatter of the fatigue test is for tension-compression than for repeated bending. Another factor of the scatter of test results is an error of the applied load, which was analysed in the paper by Strzelecki (2018). It was found that the applied load could cause up to 46% of standard deviation. Gope (2012) found that the error of estimation fatigue life for salt solution was around half smaller than that for air condition. Standard deviation for different types of materials was tested by Wormsen *et al.* (2015). Standard deviation for the same type of material AISI 8630 M, but from different suppliers, have values 0.133, 0.188, 0.126, 0.186 and 0.106 for the axial load with asymmetry of cycles $R = -1$.

Taking these factors into account, it is impossible to estimate the exact value of standard deviation of the fatigue life. It can only be stated that it is from 0.1 to 0.25, see Wormsen *et al.* (2015). Values in this range, namely 0.159 to 0.214, were estimated while analysing materials, Table 2. So, the scatter of fatigue tests can be estimated only from experimental results.

It is worth the notice, that the relation of hardness, tensile strength or yield strength has been used to develop analytical methods for many years (e.g. presented in Pang *et al.* (2014)). However, the authors have not found such correlations for the scatter of fatigue life. The test results presented above show good correlations of standard deviation for the yield strength

with fatigue life. Despite of different damage mechanisms in these tests, in both cases, they are dominated by the elastic strain. So, it can be assumed that the standard deviation can be the same, which has been proven for the presented materials.

6. Conclusion

Upon analysing the obtained results for standard deviation using the standard method, it becomes clear that the standard deviation for rotary bending was higher than that for axial loading. For stainless steels (1.4301 and 1.4404) differences of the standard deviations were significant and equaled 26%. However, those differences for aluminium alloy (AW 6063 T6 and AW 2017A T4) were little and equaled 3%. For most cases, the standard deviation for a small number of specimens was smaller than the standard deviation σ_N from the standard method. An exception was S355J2+C steel.

The proposed method of determination of the P-S-N curve has a lower value fatigue life than the ASTM method. An overestimated fatigue life was obtained by the ASTM method (except for S355J2+C). The proposed method gives underestimated values (except for aluminium alloys). Thus, the proposed method can be used by engineers.

References

1. Bai X., Zhang P., Zhang Z., Liu R., Zhang Z., 2019, New method for determining *P-S-N* curves in terms of equivalent fatigue lives, *Fatigue and Fracture of Engineering Materials and Structures*, **42**, 10, 2340-2353
2. Goedel F., Mezzomo G.P., Pravia Z.M.C., 2018, Fatigue lifespan of a fillet welded joint – Hybrid approach to obtain the S-N curve with a reduced number of tests, *Latin American Journal of Solids and Structures*, **15**, 10
3. Gope P.C., 1999, Determination of sample size for estimation of fatigue life by using Weibull or log-normal distribution, *International Journal of Fatigue*, **21**, 8, 745-752
4. Gope P.C., 2012, Scatter analysis of fatigue life and prediction of S-N curve, *Journal of Failure Analysis and Prevention*, **12**, 5, 507-517
5. Kurek A., Koziarska J., Kluger K., Łagoda T., 2017, Fatigue life of 2017A-T4 aluminium alloy under different types of stress, *Journal of Machine Construction and Maintenance*, **4**, 53-61
6. Lee Y.-L., Pan J., Hathaway R.B., Barkey M.E., 2005, *Fatigue Testing and Analysis – Theory and Practice*, Elsevier Butterworth-Heinemann
7. Lewis G., Sadhasivini A., 2004, Estimation of the minimum number of test specimens for fatigue testing of acrylic bone cement, *Biomaterials*, **25**, 18, 4425-4432
8. Li C., Wu S., Zhang J., Xie L., Zhang Y., 2020, Determination of the fatigue *P-S-N* curves – A critical review and improved backward statistical inference method, *International Journal of Fatigue*, **139**, 105789
9. Ligaj B., Szala G., 2013, *Hybrid Method for Fatigue Life Calculations* (in Polish), J. Szala (Edit.), Wydawnictwo Naukowe Instytutu Technologii Eksploatacji Państwowego Instytutu Badawczego, Radom
10. Liu X., Sun Q., 2020, Small sample-based fatigue reliability analysis using non-intrusive polynomial chaos, *IEEE Access*, **8**, 59678-59683
11. Nanninga N., White C., 2009, The relationship between extrusion die line roughness and high cycle fatigue life of an AA6082 alloy, *International Journal of Fatigue*, **31**, 7, 1215-1224

12. Pang J.C., Li S.X., Wang Z.G., Zhang Z.F., 2014, Relations between fatigue strength and other mechanical properties of metallic materials, *Fatigue and Fracture of Engineering Materials and Structures*, **37**, 9, 958-976

13. R Core Team, 2023, R: A Language and Environment for Statistical Computing, The R Foundation for Statistical Computing, Vienna, http://www.r-project.org/

14. Shen, C., 1994, The statistical analysis of fatigue data, Ph.D. Thesis, The University of Arizona, https://repository.arizona.edu/handle/10150/186810

15. Skibicki D., Sempruch J., Pejkowski Ł., 2014, Model of non-proportional fatigue load in the form of block load spectrum, *Materialwissenschaft und Werkstofftechnik*, **45**, 2, 68-78

16. Soh Fotsing B. D., Anago G.F., Fogue M., 2010, Statistical techniques of sample size estimating in fatigue tests, *International Journal of Engineering and Technology*, **2**, 6, 477-481

17. Strzelecki P., 2018, Fatigue scatter of the tests results related to error of applied stress, [In:] *Advances in Mechanics: Failure, Deformation, Fatigue, Waves and Monitoring. Proceedings of the 11th International Conference on Structural Integrity and Failure*, Perth, Australia, 3-6 December 2018, Dyskin A.V., Pasternak E. (Edits.), University of Western Australia, Perth, WA, Australia, 28-32

18. Strzelecki P., 2021, Determination of standard deviation for fatigue strength based on the tensile test, *ICSID 2021 5th International Conference on Structural Integrity and Durability*, Željko Božić (Edit.), University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture, Dubrovnik, 13

19. Strzelecki P., Sempruch J., 2016, Verification of analytical models of the S-N curve within limited fatigue life, *Journal of Theoretical and Applied Mechanics*, **54**, 1, 63-73

20. Strzelecki P., Sempruch J., Nowicki K., 2015, Accuracy of analytical-experimental method for determining the fatigue characteristics in a limited life region, *Solid State Phenomena*, **224**, 63-68

21. Strzelecki P., Wachowski M., 2022, Effect of the stress concentration factor on the final fracture zone of aluminium AW 6063 T6 for rotating bending specimens, *Materials Today Communications*, **31**, 103766

22. Walpole R.E., Myers R.H., Myers S.L., Ye K.E., 2012, *Probability and Statistics for Engineers and Scientists*, 9-th ed., Pearson Education, Inc.

23. Wormsen A., Avice M., Fjeldstad A., Reinås L., Macdonald K.A., Muff A.D., 2015, Base material fatigue data for low alloy forged steels used in the subsea industry. Part 1: In air S-N data, *International Journal of Fatigue*, **80**, 477-495

24. Xie L., Liu J., Wu N., Qian W., 2014, Backwards statistical inference method for P-S-N curve fitting with small-sample experiment data, *International Journal of Fatigue*, **63**, 62-67

25. Zu T., Kang R., Wen M., Chen Y., 2020, α-S-N curve: A novel S-N curve modeling method under small-sample test data using uncertainty theory, *International Journal of Fatigue*, **139**, 105725

Manuscript received December 8, 2023; accepted for publication September 25, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 695-711, Warsaw 2024
https://doi.org/10.15632/jtam-pl/193523

DESIGN OF A SEMI-ACTIVE SUSPENSION CONTROL METHOD BASED ON AN ENHANCED INVERSE MODEL FOR NONLINEAR MAGNETORHEOLOGICAL DAMPERS

Lei Jia, Chun Wang
School of Mechanical Engineering, Shenyang Ligong University, Shenyang, China
corresponding author C. Wang, e-mail: wangchun016050@163.com

Yang Tian
School of Mechanical Engineering, Liaoning Engineering Vocational College, Tieling, China

Jiankang Yang, Ziliang Liu, Xin Zhang
School of Mechanical Engineering, Shenyang Ligong University, Shenyang, China

The nonlinear hysteresis characteristics of magnetorheological dampers lead to low fitting accuracy and poor practicality of their inverse models. Hence, to improve the accuracy of an inverse model generated with BP neural network, this research presents a novel optimization approach called Beluga Whale Optimization. The prediction accuracy of current is enhanced by the optimized inverse model. Under the enhanced inverse model, a variable universe fuzzy PID control is created. Based on the research outcomes, it has been shown that the introduction of control contributes to noteworthy improvements in the suspension performance metrics, both in terms of time and frequency domains.

Keywords: semi-active suspension, BP neural network, beluga whale optimization, variable universe fuzzy PID, magnetorheological damper

1. Introduction

The suspension of a vehicle plays a crucial role in minimizing the effects of road vibrations and ensuring stability of the vehicle body (Tseng and Hrovat, 2015). The passive suspension cannot satisfy the needs for complex operating conditions due to its simple structure and fixed parameters, so the intelligent suspension (active suspension, semi-active suspension) has emerged as a hot topic of research (Na *et al.*, 2022; Krzyzynski and Maciejewski, 2019). The active suspension mostly uses the target forces generated by the actuator to cut down vibration. Semi-active suspension has gained considerable attention as an alternative to address the challenges of complexity, exorbitant costs, and excessive energy consumption associated with active suspensions (Liu *et al.*, 2022).

Magnetorheological dampers (MRD) are widely employed intelligent devices for damping vibrations in various fields. They were first introduced into the suspension system by Lou *et al.* (1994), which has aroused the attention of researchers. MRD is a device that uses the magnetorheological effect to control the damping force. MRD offers fast responsiveness, wide control range and low energy consumption, but it also has a powerful hysteresis and nonlinearity. Designing a semi-active suspension control has made it crucial to develop a highly precise MRD model. Consequently, establishing this model has emerged as one of the utmost essential objectives (Mai *et al.*, 2020; Dong *et al.*, 2010). The range of mechanical models has been presented by the researchers (Chen, 2022), including the Bingham model, the Bouc-wen model, and others. The frequent utilization of the Bouc-Wen model (Tang *et al.*, 2020; Gong and Hen, 2020) has been observed to describe the properties of MRD. Boada *et al.* (2018) and Maciejewski *et al.* (2019)

developed an inverse model to output the control current. Zhang and Zhao (2017) has built the inverse model with the BP neural network (BPNN), which output a fairly accurate predicted currents but still had some defects. Ma *et al.* (2021) obtained a higher accurate system with the Elman neural network compared to BPNN. BWO, a bio-inspired optimization technique created by Zhong *et al.* (2022), is applied to tackle challenging engineering and optimization issues. In this paper, the BPNN is optimized to provide a more accurate inverse model thanks to the benefit that the BWO can ensure the algorithm converge globally.

Effective and reasonable control algorithm is another crucial role of the intelligent suspension system for vibration reduction. Successful implementations have been carried out for vibration control of vehicles using control methods such as skyhook control and PID control (Papaioannou *et al.*, 2021; Shin *et al.*, 2016), but semi-active control still faces a huge challenge for the nonlinear characteristics of MRD. The issue at hand has garnered heightened interest in the implementation of nonlinear control techniques, namely fuzzy control, sliding model control, H_∞ control, and neural network control in MRD semi-active suspension systems. This approach aims to address and mitigate the aforementioned challenge (Al Aela *et al.*, 2022). Morales *et al.* (2018) proposed a variable damping control strategy for a semi-active suspension system to improve smoothness of the vehicle ride. As well as compared to other control methods, fuzzy control as an intelligent control can handle many problems that cannot be accurately mathematically modeled in complex systems. Trikande *et al.* (2018) set up a quarter suspension system with fuzzy acceleration and velocity of the suspension. According to the experimental results, it has been shown that the incorporation of fuzzy control significantly improves the stability and ride comfort of the suspension system. In this study, in the realm of active suspension systems, Na *et al.* (2020) introduced an innovative and dynamic approach to regulate control strategies. This scheme effectively ensures vehicle stability in the presence of suspension parameter uncertainties and time delays. Despite the fuzzy control has small overshoot, good robustness and strong adaptability, there is inevitable steady-state control error. Henceforth, merging the concept of fuzzy control with other control approaches becomes imperative in order to obtain optimal control with utmost stability and precision. A fuzzy control strategy for the active suspension in vehicles was introduced by Wang *et al.* (2015). The proposal garnered significant attention as it addresses the issue of balancing control precision and speed, while also confronting the impact of uncertain control parameters. According to the work of Muthalif *et al.* (2017), a semi-active suspension system model was created with integrating fuzzy PID control, where the objective was to enhance control performance by applying fuzzy logic to manipulate the input variables, including body speed and acceleration. Fuzzy PID control not only tackles the issue of steady-state error in control systems, but also provides adaptive adjustment of PID control parameters. Although these aforementioned techniques greatly enhance the precision and adaptability of suspension control, they primarily focus on addressing the uncertainty inherent within their own parameters. As such, they fail to incorporate external excitations and are unable to proactively pre-adjust control parameters to promptly react to variations in external stimuli. In an effort to improve control precision of the fuzzy controller, Pang *et al.* (2018) put forward a variable universe fuzzy T-S with a semi-active suspension system, whereas Zhang *et al.* (2024) designed variable universe fuzzy PID (VUF-PID) for a continuous damping semi-active suspension. The conventional fuzzy PID technique is unable to scale owing to variation of the initial theoretical domain error, causing the system control accuracy decline. The VUF-PID strategy has the advantages of the conventional fuzzy PID, and meanwhile, the utilization rate of fuzzy rules is largely enhanced by adjusting theoretical domain factor, which enhances the system accuracy. The major contribution of this work is a new inverse model of MRD, which estimates the current required to make the desired force exerted by MRD. Prior studies have suggested that the use of fuzzy PID in the suspension system offers considerable robustness. In this study, the notion of variable domains is incorporated into the realm of fuzzy control. By modifying the control

domain, both the controller input and output can be flexibly tailored, thereby enhancing the system adaptability and the performance of the suspension.

Structural organization of the paper is a follows: Section 2 describes the semi-active suspension and MRD dynamic model, Section 3 focuses on design of the MRD inverse model and system controller, Section 4 performs an experimental simulation of the developed controller, and Section 5 provides a summary of the entire work accomplished.

2. Dynamic modeling

2.1. Semi-active suspension system

The vertical force and vibration are the main factors affecting the vehicle suspension system, and the single suspension system mainly responds to vertical motion of the vehicle, so it can fulfill the requirement of the research. This paper focuses on the two-degrees of freedom semi-active suspension system of a quarter vehicle model, which is shown as a simplified model in Fig. 1.

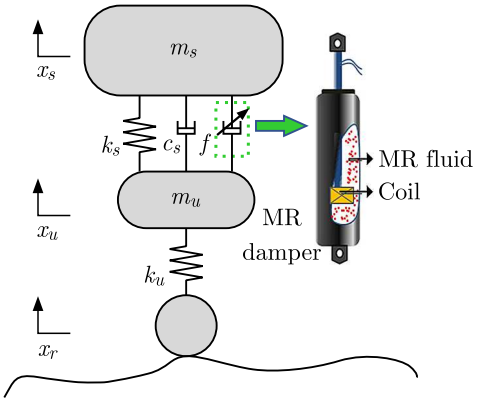


Fig. 1. Modeling of a semi-active suspension with MRD

The differential equations for semi-active suspension motion can be derived based on Newton's law of motion and Lagrange's law

$$\begin{aligned} &m_s\ddot{x}_s + c_s(\dot{x}_s - \dot{x}_u) + k_s(x_s - x_u) + f = 0 \\ &m_u\ddot{x}_u - c_s(\dot{x}_s - \dot{x}_u) - k_s(x_s - x_u) + k_u(x_u - x_r) - f = 0 \end{aligned} \tag{2.1}$$

In Eqs. (2.1), m_s and m_u are body mass and wheels mass, k_s is the suspension stiffness, x_s refers to the absolute displacement of vehicle body, c_s is the damping coefficient, x_r is the road surface excitation, k_t is the tire stiffness, x_u is the absolute displacement of the wheel, f stands for the damping force. The parameter values for the quarter semi-active suspension are shown Table 1.

Table 1. Parameter values of the quarter semi-active suspension

Parameter	Value	Unit
m_s	345	kg
m_u	40.5	kg
k_s	17	kN/m
k_u	192	kN/m
c_s	1500	Ns/m

To facilitate the design of the controller, the state is definined as

$$x_1 = x_s - x_u \qquad x_2 = \dot{x}_s \qquad x_3 = x_u - x_r \qquad x_4 = \dot{x}_u \tag{2.2}$$

The state-space equation is acquired form Eqs. (2.1) and (2.2) as follows

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}f + \mathbf{C}\dot{x}_u \tag{2.3}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & -1 & 0 & -1 \\ -\dfrac{k_s}{m_s} & -\dfrac{c_s}{m_s} & 0 & \dfrac{c_s}{m_s} \\ 0 & 0 & 0 & 1 \\ \dfrac{k_s}{m_u} & \dfrac{c_s}{m_u} & -\dfrac{k_u}{m_u} & -\dfrac{c_s}{m_u} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 \\ \dfrac{1}{m_s} \\ 0 \\ -\dfrac{1}{m_u} \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

2.2. Dynamic model of a magnetorheological damper

According to the modified Bouc-Wen model, this paper investigates MRD. Its structure includes a damping element, elastic element and hysteresis operator, as show in Fig. 2.

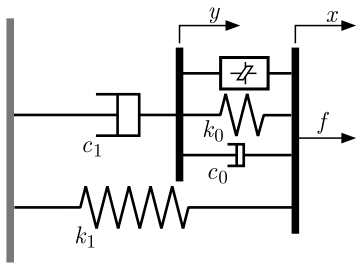


Fig. 2. Modified Bouc-wen model

The function expression is

$$\begin{aligned} &f = c_1\dot{y} + k_1(x - x_0) \qquad y = \frac{1}{c_0} + c_1[\delta z + c_0\dot{x} + k_0(x - y)] \\ &z = -\gamma|\dot{x} - \dot{y}|z|z|^{n-1} - \psi(\dot{x} - \dot{y})|z|^n + G(\dot{x} - \dot{y}) \end{aligned} \tag{2.4}$$

where c_1 and δ can be expressed as

$$c_1 = c_{1a} + c_{1b}\mathrm{i} \qquad \delta = \delta_a + \delta_b\mathrm{i} \tag{2.5}$$

In equations (2.4) and (2.5), x denotes the displacement of the spring, x_0 stands for the initial value, y is the internal displacement, c_0 represents viscous damping factor at a high speed, z is an evolutionary variable, k_0 stands for the stiffness coefficient, k_1 expresses the stiffness coefficient of the accumulator, c_1 represents the viscous damping factor at a low speed, δ indicates the scale factor of hysteresis operator, γ, ψ, n, G represent the adjustment coefficient.

3. Control system design

A novel control technique is developed utilizing the semi-active suspension of MRD. This technique incorporates the VUF-PID control model and the inverse model of MRD, which has been created using BPNN optimized with BWO, as depicted in Fig. 3.

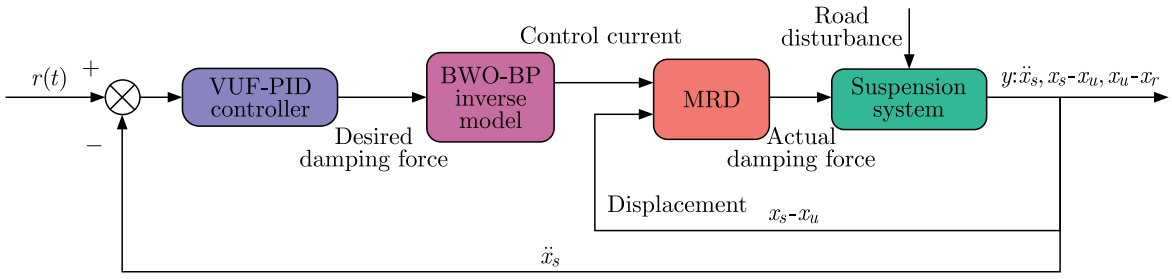


Fig. 3. Control schematic diagram

3.1. An inverse model of the magnetorheological damper

3.1.1. An inverse model based on BPNN

The BPNN is employed in constructing the inverse model of MRD, as depicted in Fig. 4. This figure also illustrates the correlation between the current and the damping force, as represented by the mapping relationship. Damping forces are obtained from the Modified Bouc-Wen model of magnetorheological dampers, the control current from random white noise, displacement of MRD data from suspension dynamic travel. 2000 sets of data were collected in total, with 1500 sets designated for network training and the remaining 500 sets reserved for testing purposes.

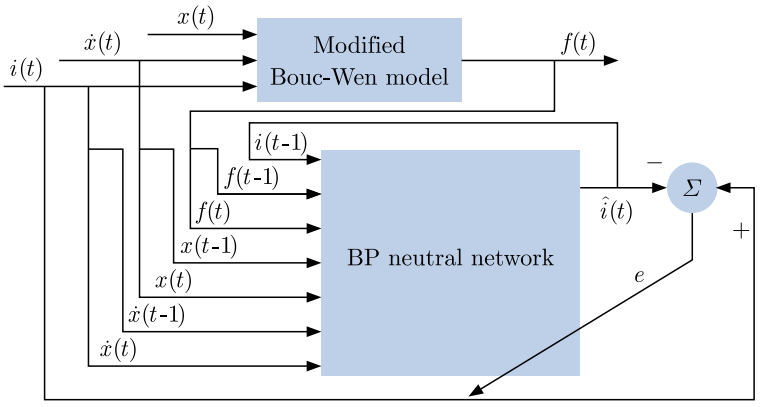


Fig. 4. Inverse model identification scheme of the MRD with BPNN

The neural network used in this study consists of three layers, namely the input layer, the hidden layer, and the output layer. It is composed of a total of seven input nodes, while the hidden layer comprises twelve nodes, and finally, a single output node. The inputs chosen for the neural network include the damping force, velocity, and displacement of MRD at time t and $t-1$, along with the control current at time $t-1$. The neural network output represents the current at time t.

3.1.2. Optimization of inverse models

During training, BPNN is likely to generate local optimum. In order to obtain more accurate current signals, this paper focuses on optimizing BPNN with the Beluga Whale Optimization (GWO). BWO, a bio-inspired optimization technique is applied to tackle challenging engineering and optimization issues. BPNN is most sensitive to weights and thresholds in initialization parameters. In the process of optimizing BPNN, BWO is equivalent to constantly updating the weights and thresholds and calculating the global optimum through multiple iterations. Figure 5 expresses the flow chart of the algorithm.

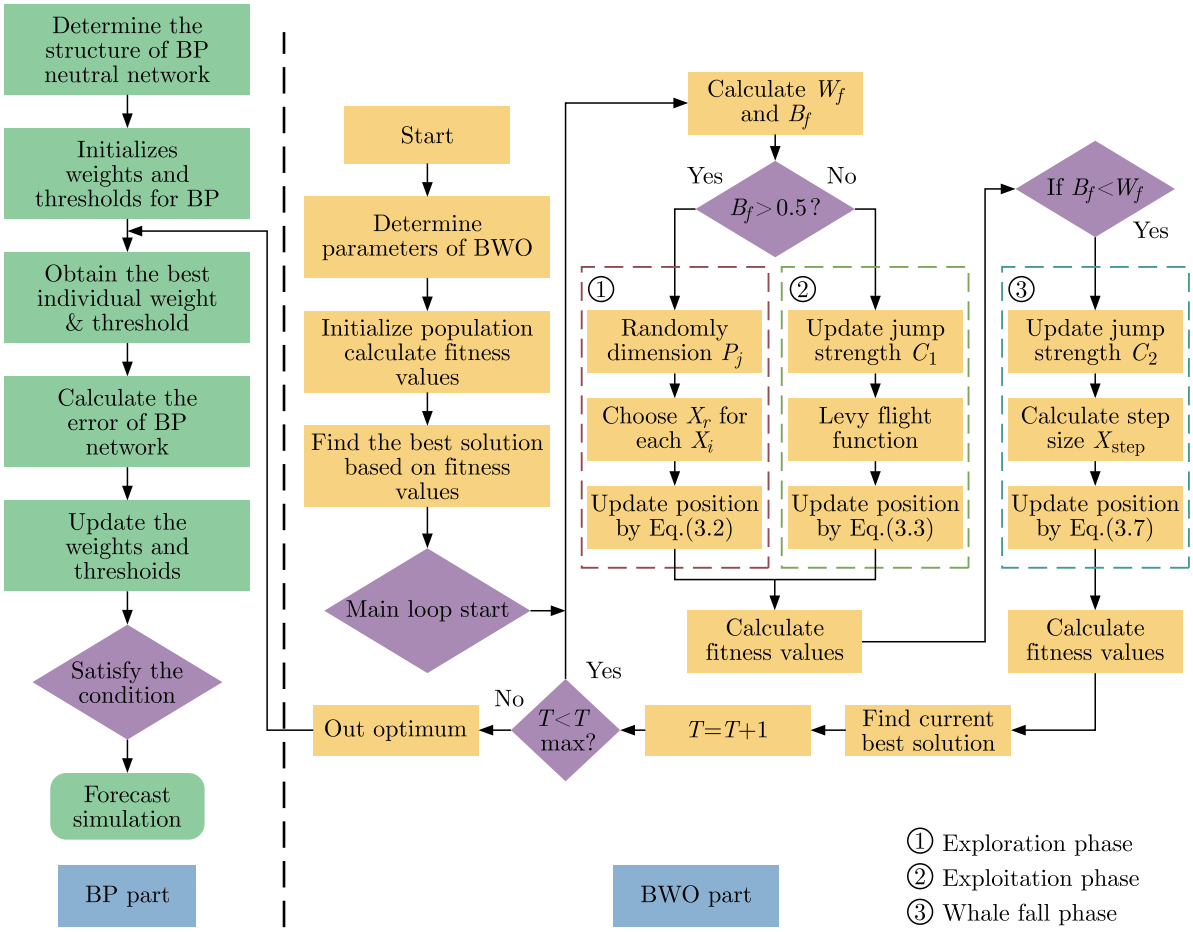


Fig. 5. The flow chart of BP neural network optimized by BWO

The main focus of this program lies in acquiring knowledge from the exploration and exploitation phases, as well as adopting behavioral traits observed in whales. In the BWO method, with the help of a balancing factor B_f, B_0 is randomly varied between $(0, 1)$ at each iteration. It allows conversion from exploration to exploitation. T and T_{max} denote the current and maximum number of iterations, respectively. The balance factor B_f can be expressed as

$$B_f = B_0\Big(1 - \frac{T}{2T_{max}}\Big) \tag{3.1}$$

When the B_f exceeds 0.5, the exploration phase commences, whereas when B_f falls below 0.5, it is the start of exploitation phase.

I) Exploration phase

Now, it is about to mimic beluga whales' swimming style. Depending on the odd and even numbers of positions, different positions are updated. Here is the mathematical model

$$x_{i,j}^{T+1} = \begin{cases} x_{r,P_j}^T + (x_{r,P_1}^T - x_{i,P_j}^T)(1 + r_1)\sin(2\pi r_2) & \text{for} \quad j = \text{even} \\ x_{r,P_j}^T + (x_{r,P_1}^T - x_{i,P_j}^T)(1 + r_1)\cos(2\pi r_2) & \text{for} \quad j = \text{odd} \end{cases} \tag{3.2}$$

where $x_{i,j}^{T+1}$ represents the i-th beluga whale in the j-th dimension, T denotes current iteration, r denotes randomly selected beluga whale. r_1 and r_2 indicate random numbers that range from 0 to 1, $\sin(2\pi r_2)$ and $\cos(2\pi r_2)$ denote mirrored beluga whales with their fins facing water.

II) Exploitation phase

The idea of this stage comes from the feeding behavior of beluga whales, which share information between themselves and their neighboring partners, so that the position of the best individual and other individuals can be updated. Introducing the Lévy flight strategy of predation, which is mathematically represented as

$$x_i^{T+1} = r_3 x_{best}^T - r_4 x_i^T + C_1 L_F (x_r^T - x_i^T) \tag{3.3}$$

where x_i^T stands for the i-th beluga whale's current position, x_r^T stands for random beluga whale's current positions, the optimal position for beluga whales is x_{best}^T, r_3 and r_4 indicate random numbers that range from 0 to 1, C_1 stands for the random jump strength used to gauge the Lévy flight intensity and represented by equation $(3.4)_1$, L_F represents the Lévy flight function, calculated from equation $(3.4)_2$

$$C_1 = 2r_4\Big(\frac{1-T}{T_{max}}\Big) \qquad L_F = 0.05\frac{u\sigma}{|v|^{1/\zeta}} \tag{3.4}$$

where u and v represent normally distributed random variables, while ζ represents a constant with a default value of 1.5σ, which is represented as

$$\sigma = \Big(\frac{\Gamma(1+\beta)\sin(\pi\zeta/2)}{\Gamma[(1+\beta)/2]\zeta 2^{(\zeta-1)/2}}\Big)^{1/\zeta} \tag{3.5}$$

III) Whale fall

Belugas face threats from orcas, polar bears and humans as they migrate and forage. The majority of belugas demonstrate their intelligence by evading these threats through effective communication. Nonetheless, some belugas unfortunately perish, sinking to depths of the ocean floor. This occurrence is referred to as a "whale fall".

In each iteration, to imitate the behavior of a whale fall, we simulate slight variations in the groups by selecting a likelihood of the whale fall from the individual. The beluga whales in question have either repositioned themselves or been subjected to expulsion and have subsequently descended into depths of the ocean. To establish the revised position, the utilization of beluga whale locations and the magnitude of their stride during falls is imperative. It guarantees the population size remain unaltered. The mathematical representation of the model is as follows

$$x_i^{T+1} = r_5 x_i^T - r_6 x + r_7 x_{step} \tag{3.6}$$

where r_5, r_6 and r_7 indicate a random numbers that range from 0 to 1, x_{step} indicates the step size for the whale fall, and x_{step} can be expressed as

$$x_{step} = (u_b - l_b)\exp\Big(-C_2\frac{T}{T_{max}}\Big) \tag{3.7}$$

where u_b represents the maximum limit of the variable, while l_b represents the minimum limit of the variable lower boundary. C_2 is expressed as

$$C_2 = 2W_f n \tag{3.8}$$

where C_2 denotes the step factor, W_f can be expressed as

$$W_f = 0.1 - 0.05\frac{T}{T_{max}} \tag{3.9}$$

As beluga whales progressively approach their source of sustenance, the level of risk associated with them diminishes throughout the optimization process. The evident proof arises from a drop in the likelihood of the whale fall, plunging from 0.1 in the initial iteration to 0.05 in the final iteration.

3.2. Control design of the suspension system

3.2.1. VUF-PID theory

The fuzzy PID controller integrates a variable universe controller known as VUF-PID. To enhance its performance, the fuzzy PID controller applies the expansion factor to incorporate the variable domain concept. The number of fuzzy rules at the local level is adjusted by this effect while overcoming the restricted accuracy caused by the limited count of fuzzy control rules found in the parameterization of the PID control problem. As shown in Fig. 6, the concept of fuzzy variable universe is aimed at changing the size of basic fuzzy variable set as the system error changes during the control process, thus ensuring dynamic adaptation for the optimized performance. In Fig. 6, x stands for input variables of the control system, μ represents the membership degree function, $\alpha(x)$ indicates the expansion factor of the input variable domain, e denotes the input initial domain.

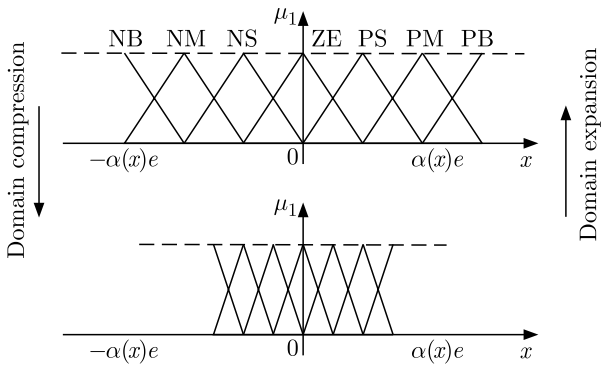


Fig. 6. Schematic diagram of universe transformation

3.2.2. Fuzzy PID controller

Figure 7 depicts the flow chart of fuzzy PID control, which builds upon the foundational PID control system through incorporation of fuzzy theory. This integration allows for the enhancement and improvement of the basic PID control approach. This paper adopts the two-input, three-output form of fuzzy control, controller inputs are error e and change rate of error ec. The error e is calculated as a difference between vertical acceleration of the vehicle $\ddot{x}_s$, and the set value $r(t) = 0$, ΔK_p, ΔK_i and ΔK_d as the controller output. The input and output variables are grouped into seven levels, denoted as NB, NM, NS, ZE, PS, PM, PB. All domains have a range of $(-6, 6)$ configured. The triangle function is employed for the membership functions of input and output. Fuzzy inference rules are established based on the suspension system characteristics, displayed in Table 2.

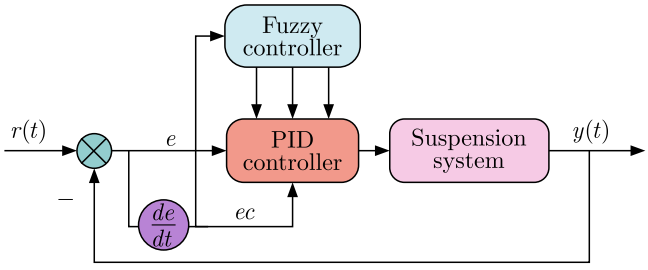


Fig. 7. Fuzzy-PID controller scheme

Table 2. Fuzzy rule table

	NB	NM	NS	ZE	PS	PM	PB
NB	PB/NB/PS	PB/NB/NS	PM/NM/NB	PM/NM/NB	PS/NS/NB	ZE/ZE/NM	ZE/ZE/PS
NM	PB/NB/PS	PB/NB/NS	PM/NM/NB	PM/NS/NM	PS/NS/NM	ZE/ZE/NS	NS/ZE/ZE
NS	PM/NM/ZE	PM/NM/NS	PM/NS/NM	PS/NS/NM	ZE/ZE/NS	NS/PS/NS	NS/PS/ZE
ZE	PM/NM/ZE	PM/NM/NS	PS/NS/NS	ZE/ZE/NS	NS/PS/NS	NM/PM/NS	NM/PM/ZE
PS	PS/NM/ZE	PS/NS/ZE	ZE/ZE/ZE	NS/PS/ZE	NS/PS/ZE	NM/PM/ZE	PM/PB/ZE
PM	PS/ZE/PB	ZE/ZE/NS	NS/PS/PS	NM/PS/PS	NM/PM/PS	NM/PB/PS	NB/PB/PB
PB	ZE/ZE/PB	ZE/ZE/PM	NM/PS/PM	NM/PM/PM	NM/PM/PS	NB/PB/PS	NB/PB/PB

Based on the control principle of fuzzy PID, the formula for adjusting the parameters of PID is presented below

$$K_p = K_{p0} + \Delta K_p \qquad K_i = K_{i0} + \Delta K_i \qquad K_d = K_{d0} + \Delta K_d \tag{3.10}$$

3.2.3. VUF-PID controller

In variable universe fuzzy control, the expansion factor directly determines the precise control effect and overall control performance. The universe formula for the scaling factor is as follows

$$X_e = [-\alpha e, \alpha e] \qquad Y = [-\beta u, \beta u] \tag{3.11}$$

where α, β can be regarded as the scaling factors for input and output, e, u represent the original input and output universe, X_e, Y denote the resultant input and output universe.

The present investigation utilizes the VUF-PID function (Zeng *et al.*, 2020), as depicted in Fig. 8. The scaling coefficients primarily derived from functional models are produced via functions parameterized by the error and its rate of change. The scaling factors change in the form of exponent

$$\alpha(x) = 1 - \lambda \mathrm{e}^{-kx^2} \qquad \beta(t) = K_I \sum_{i=1}^{n} P_i \int_0^t e_i(\tau)\, d\tau + \beta(0) \tag{3.12}$$

In Eq. (3.12), $9 < \lambda < 1$, $k > 0$, the value of k reflects changing speed of universe of the controller, λ reflects accuracy of the controller, K_I, P_i are constants, $\beta(0)$ is the initial value, $e_i(\tau)$ is the error vector, τ is the adjustment parameter.

The domain of the input variable is

$$X_e(e) = [-\alpha(x)e, \alpha(x)e] \qquad X_e(ec) = [-\alpha(x)ec, \alpha(x)ec] \tag{3.13}$$

Based on the characteristics of MRD semi-active suspension, taking into account the influence of parameters λ and k in the scaling factor, based on experience and after multiple experiments, the value of λ is 0.6 and k is 0.5, then

$$\alpha(e) = 1 - 0.6\mathrm{e}^{-0.5e^2} \qquad \alpha(ec) = 1 - 0.6\mathrm{e}^{-0.5ec^2} \tag{3.14}$$

The variation of integral coefficient K_i tends to opposite direction to the variation of the system deviation, whereas the variation of the output variables K_p and K_d are consistent with variation of the system deviation. The following are the scaling factors for K_p, K_i and K_d

$$\beta(K_p) = 3|e| \qquad \beta(K_i) = \frac{1}{|e| + 0.9} \qquad \beta(K_d) = 3|e| \tag{3.15}$$

The changed output variables $Y(K_p)$, $Y(K_p)$ and $Y(K_d)$ domains can be expressed as

$$\begin{aligned} Y(K_p) &= [-\beta(K_p)K_p, \beta(K_p)K_p] \qquad Y(K_i) = [-\beta(K_i)K_i, \beta(K_i)K_i] \\ Y(K_d) &= [-\beta(K_d)K_d, \beta(K_d)K_d] \end{aligned} \tag{3.16}$$

The fuzzy tuning relation is defined by Θ, the output of the fuzzy control is

$$\begin{aligned} \Delta K_p &= \Theta[\beta(K_p)\Theta(\alpha(x)e, \alpha(x)ec)] \qquad \Delta K_i = \Theta[\beta(K_i)\Theta(\alpha(x)e, \alpha(x)ec)] \\ \Delta K_d &= \Theta[\beta(K_d)\Theta(\alpha(x)e, \alpha(x)ec)] \end{aligned} \tag{3.17}$$

In summary, the output of the controller is

$$f(k) = (\Delta K_p + K_{p0})e(k) + (\Delta K_i + K_{i0}) \int_0^k e(k)\, dk + (\Delta K_d + K_{d0})\frac{de(k)}{dk} \tag{3.18}$$

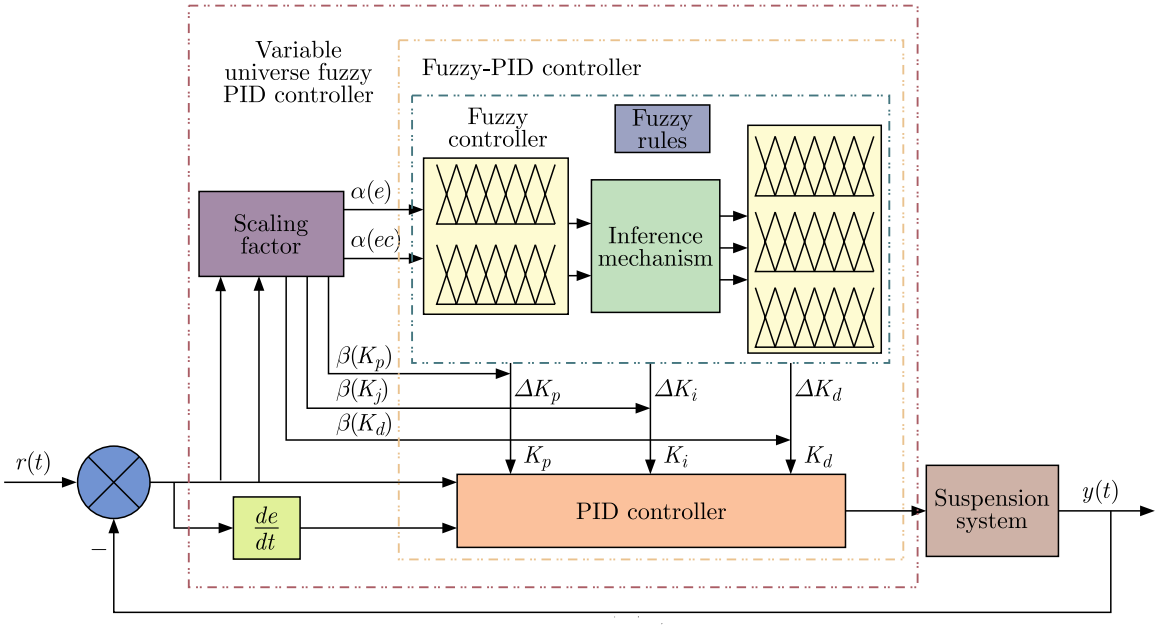


Fig. 8. Variable universe fuzzy PID control system

4. Simulation analysis

4.1. Performance of MRD

According to Eqs. (2.4) and (2.5), the model of MRD is established. Table 3 displays the research parameters for the Bouc-Wen phenomenon model (Sosthene *et al.*, 2018). The displacement follows a sinusoidal pattern with 10mm amplitude and 2Hz frequency, while the current ranges from 0A, 1A, 2A to 3A. Figure 9 illustrates the property of MRD divided into three portions for description, which included the force-time curve, force-displacement curve and force-velocity curve.

To assess the efficacy of the BWO, comparison and analysis of the prediction results of the inverse model built by BPNN before and after optimization are made. Figure 10a illustrates the original current and predictive control current with BPNN and BWO-BPNN, in Fig. 10b, a comparison is displayed between the control current of the BPNN and the BWO-BPNN, highlighting the errors. Based on the comparison of prediction results and errors provided by

Table 3. Bouc-Wen model parameters (Sosthene *et al.*, 2018)

Parameter	Value	Parameter	Value
C_0 [Ns/m]	997	k_1 [–]	134
C_{1a} [Ns/(mA)]	8186	x_0 [–]	0.115
C_{1b} [Ns/(mA)]	2725	ψ [m^2]	70000
δ_a [N/m]	0	G [–]	300
δ_b [N/m]	1723	γ [1/m^2]	70000
k_0 [–]	1072	n [–]	2

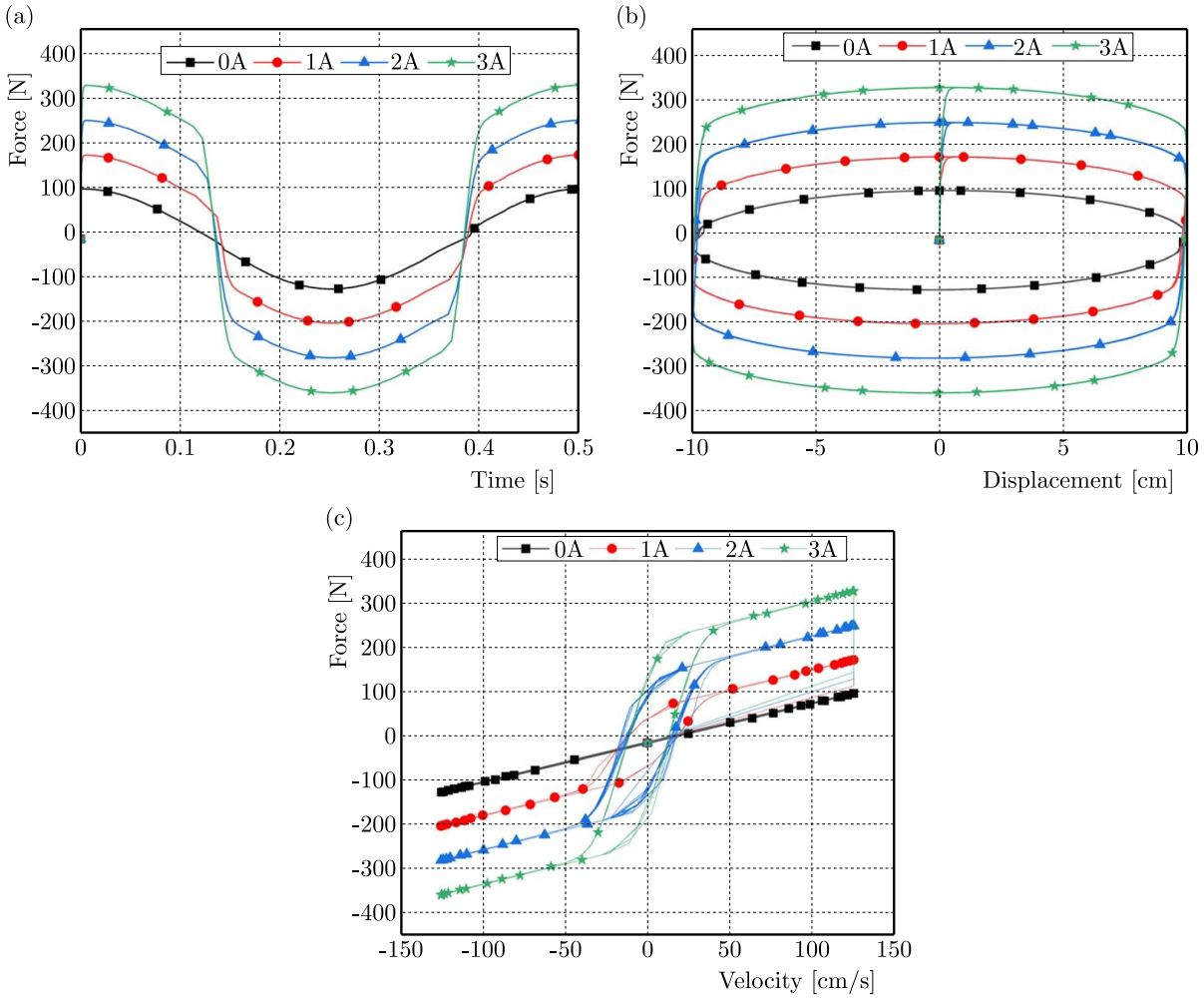


Fig. 9. Performance of MRD with different applied currents: (a) force versus time curve, (b) force versus displacement curve, (c) force versus velocity curve

Fig. 10, it is seen that BWO-BPNN has a superior fitting results of the control current predicted to original current sample than that of BPNN. Moreover, the root mean square (RMS) value of BWO-BPNN prediction error is smaller than that of BPNN, with RMS value of 0.0025621 and BPNN of 0.006093. In addition, the BWO-BPNN fitting error is reduced by 57.97%, which demonstrates the superior performance of the inverse model by BWO-BPNN.

4.2. Performance of the system controller

With the aim at thoroughly assessing the effectiveness of this controller, two types of road excitation random and sinusoidal are introduced.

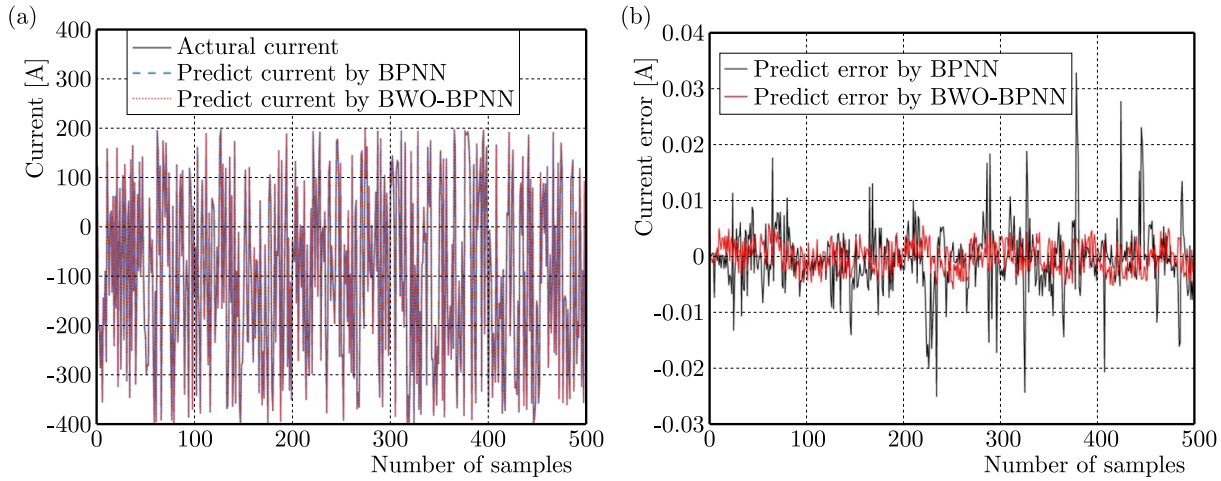


Fig. 10. The predicted values of the current using the BWO-BP and BPNN: (a) current, (b) current error

4.2.1. Random road

The random road is created by white noise. The model expression is

$$\dot{q}_t = -2\pi f_{min} q_t + 2\pi n_0 \sqrt{G q(n_0)} W_t \tag{4.1}$$

where $\dot{q}_t$ is the impact strength, f_{min} is the time frequency, n_0 denotes the spatial frequency, $G_q(n_0)$ stands for the road unevenness coefficient, v is the travel speed, W_t expresses the Gaussian white noise. In this simulation: $v = 40\,\mathrm{m/s}$, $G_q(n_0) = 64 \cdot 10^{-6}\,\mathrm{m/s}$, $n_0 = 0.1\,\mathrm{m}^{-1}$.

Figure 11 interprets variation of the suspension performance under a random surface. In this investigation, a comparison is made between four distinct suspension systems, namely passive suspension, fuzzy control, fuzzy PID control, and VUF-PID control. Performance in the time and frequency domain is analyzed to evaluate their respective effectiveness. Figures 11a,b represent body vertical acceleration, by making use of VUF-PID. The body vertical acceleration is effectively reduced, improving riding comfort. Figures 11c,d indicate the suspension dynamic displacement, VUF-PID control reduces its value to avoid the breakdown phenomenon. The stability of the vehicle is ensured by VUF-PID control, as demonstrated in Figs. 11e,f, which depicts the dynamic load on the wheel. According to the frequency domain results, the VUF-PID controller has the best results in the low frequency range of 0-20 Hz.

Table 4 displays RMS values of the suspension performance index for various control methods employed on the random road surface. Compared to the passive suspension, the vehicle body experiences a decrease in vertical acceleration by 15.85%, 36.21%, and 41.29%, while the suspension deflects dynamically by 5.57%, 33.44%, and 37.70%, respectively. Additionally, the dynamic load on the wheel witnesses a reduction of 4.01%, 35.02%, and 40.46% correspondingly.

Table 4. Simulation results under the random road

Control method	Performance indicators		
	body vertical acceleration [m/s^2]	suspension dynamic travel [m]	wheel dynamic load [N]
Passive	0.322836	0.008760	183.2071
Fuzzy	0.271669	0.008272	175.8765
Fuzzy-PID	0.205912	0.005831	119.0387
VUF-PID	0.189534	0.005458	109.0793

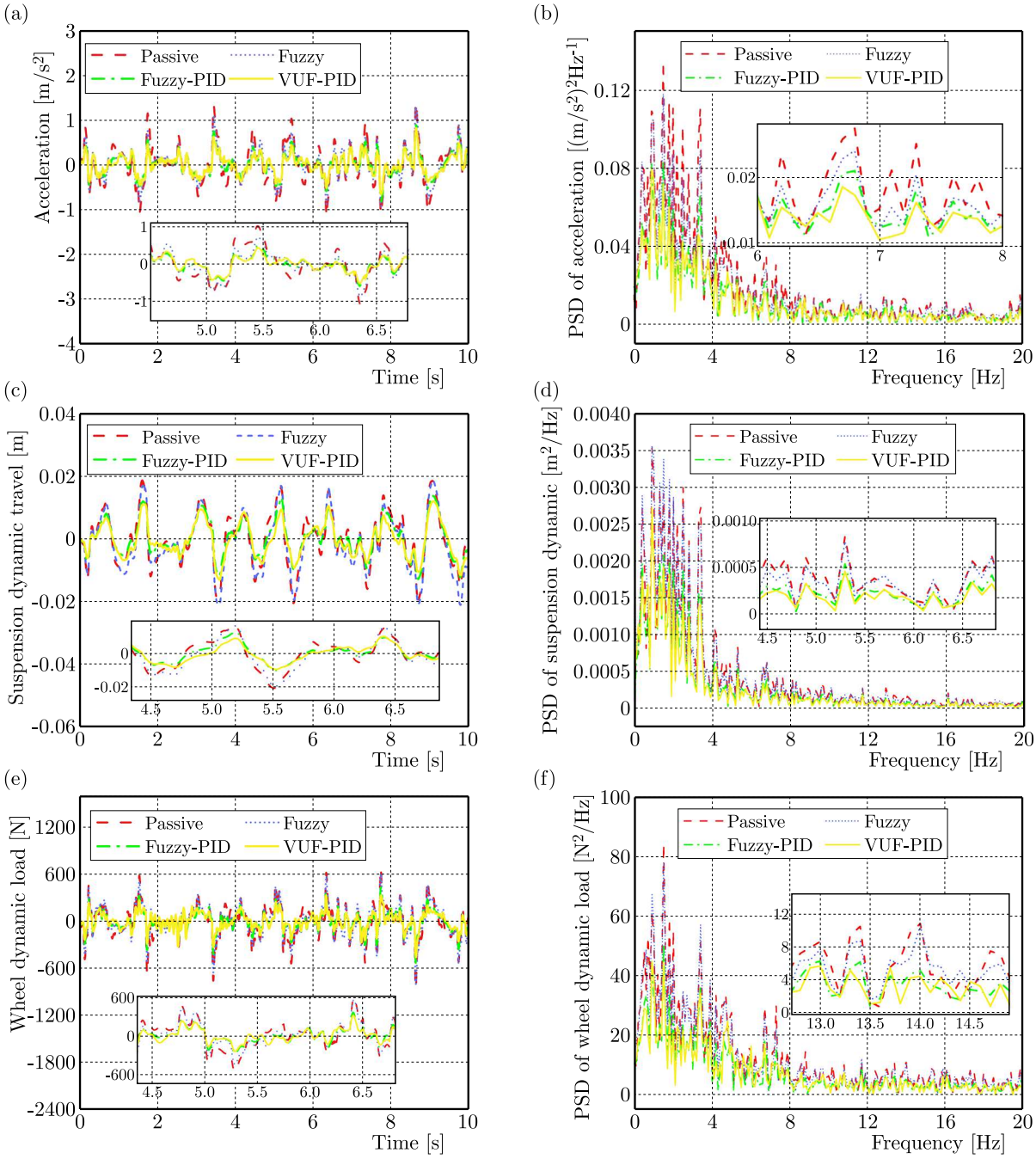


Fig. 11. Performance indexes of different control under the random road suspension system: (a) body vertical acceleration (time domain), (b) body vertical acceleration (frequency domain), (c) suspension dynamic travel (time domain), (d) suspension dynamic travel (frequency domain), (e) wheel dynamic load (time domain), (f) wheel dynamic load (frequency domain)

4.2.2. Sinusoidal road

The second road model is adopted: a sinusoidal road surface. A sinusoidal road signal is a common road excitation for suspension vibration analysis. Its is model expression is

$$x_r = \xi \sin(\omega x) \tag{4.2}$$

where ξ [m] represents the road displacement, ω [rad/s] represents the angular frequency. In this simulation, the road displacement is 0.02 m and the angular frequency is 4π rad/s.

Figure 12 shows the variation of suspension performance index under the sinusoidal road signal. The performance measures of the suspension system are enhanced when comparing it to the passive suspension through implementations of fuzzy control, fuzzy PID control, and VUF-PID control.

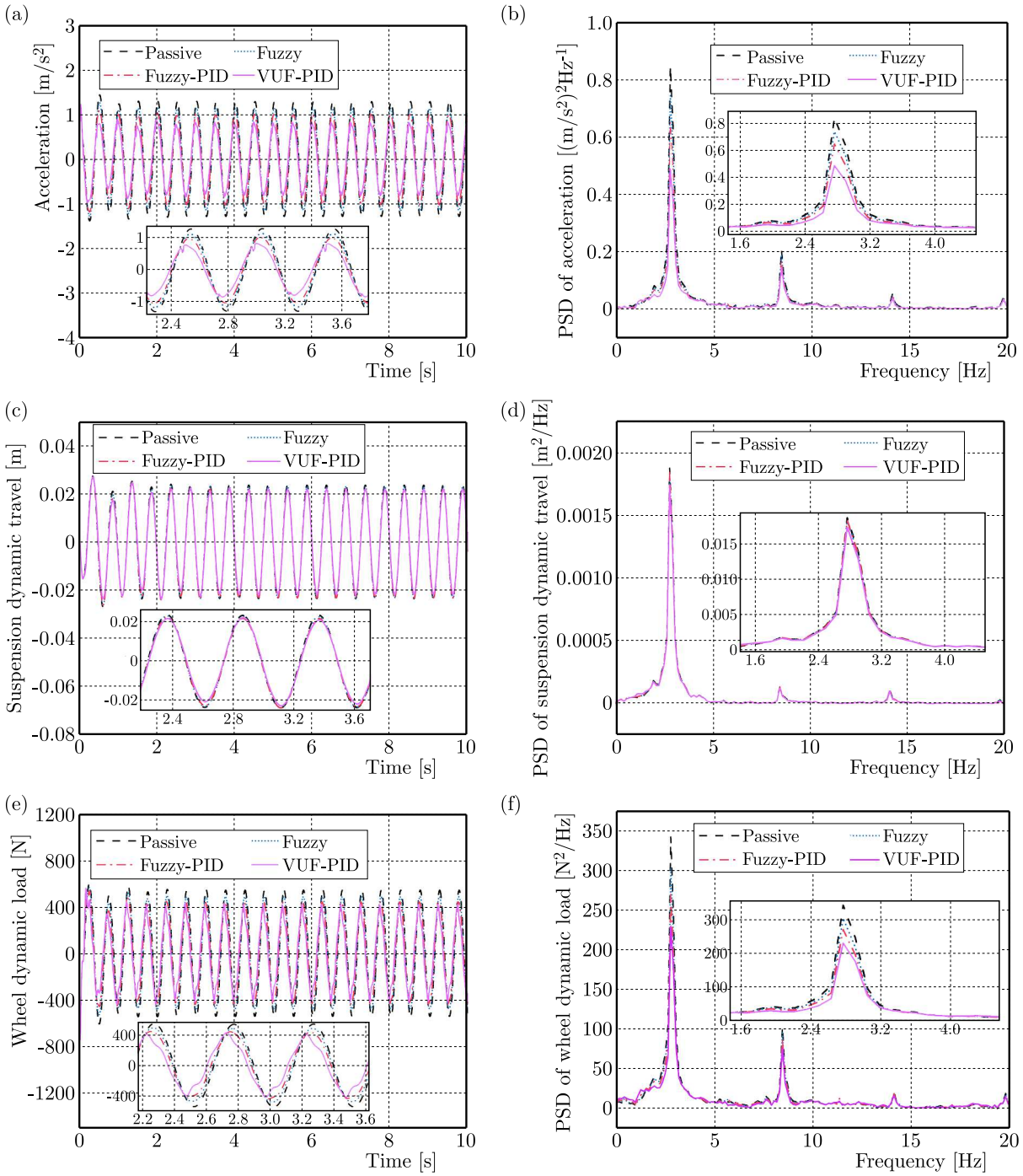


Fig. 12. Performance indexes of different control under the sinusoidal road suspension system: (a) body vertical acceleration (time domain), (b) body vertical acceleration (frequency domain), (c) suspension dynamic travel (time domain), (d) suspension dynamic travel (frequency domain), (e) wheel dynamic load (time domain), (f) wheel dynamic load (frequency domain)

Table 5 shows the RMS values of the suspension performance index across different control strategies when subjected to a sinusoidal road surface. In comparison to the passive suspension,

the vehicle body experiences a reduction in vertical acceleration by 11.46%, 22.12%, and 39.90%, while the suspension dynamic deflection decreases by 1.88%, 3.67%, and 7.04% correspondingly. Additionally, there is a consecutive decrease in the dynamic load observed for the wheel by 10.47%, 19.39%, and 29.28%.

Table 5. Simulation results under the sinusoidal road

Control method	Performance indicators		
	body vertical acceleration [m/s^2]	suspension dynamic travel [m]	wheel dynamic load [N]
Passive	0.845642	0.018207	344.8268
Fuzzy	0.748745	0.017864	308.722
Fuzzy-PID	0.658616	0.017539	277.9482
VUF-PID	0.508267	0.016926	243.8728

5. Conclusions

In this study, a functional VUF-PID control is developed with a quarter magnetorheological semi-active suspension system. In the semi-active suspension control system with MRD, the damping effect depends not only on the control algorithm of the system, but also on the precise mechanical model of MRD.

- The modified Bouc-Wen model of MRD is created, and the relationship between the damping force and displacement as well as velocity under different control currents is analyzed, which shows accurate characteristics of MRD. By using the BWO-BPNN to create an inverse model of the MRD, the fitting error has decreased by 57.97%, and the output prediction current has been more accurate compared with the BPNN.
- In this study, a functional VUF-PID control is developed with a quarter magnetorheological semi-active suspension system. With the additional control methods stated in this research, the performance indices of the suspension improved under different conditions (random road, sinusoidal road). The controller is better suited and more efficient.

Acknowledgments

The authors research is supported by Liaoning Provincial Department of Education Basic Research Project (Project No. LJKMZ20220602) and 2021 Scientific research support for high-level talent from Shenyang Ligong University (1010147001001).

References

1. Al Aela A.M., Kenne J.-P., Mintsa H.A., 2022, Adaptive neural network and nonlinear electrohydraulic active suspension control system, *Journal of Vibration and Control*, **28**, 3-4, 243-259

2. Boada M.J.L., Boada B.L., Diaz V., 2018, A novel inverse dynamic model for a magnetorheological damper based on network inversion, *Journal of Vibration and Control*, **24**, 15, 3434-3453

3. Chen K.Y., 2022, A new state observer-based vibration control for a suspension system with magnetorheological damper, *Vehicle System Dynamics*, **60**, 9, 3127-3150

4. Dong X., Yu M., Liao C., Chen W., 2010, Comparative research on semi-active control strategies for magneto-rheological suspension, *Nonlinear Dynamics*, **59**, 3, 433-453

5. Gong M., Chen H., 2020, Variable damping control strategy of a semi-active suspension based on the actuator motion state, *Journal of Low Frequency Noise, Vibration and Active Control*, **39**, 3, 787-802
6. Krzyzynski T., Maciejewski I., 2019, Computational method for shaping the vibro-isolation properties of semi-active and active systems, *Archives of Mechanics*, **71**, 4-5, 291-313
7. Liu P., Zheng M., Ning D., Zhang N., Du H., 2022, Decoupling vibration control of a semi-active electrically interconnected suspension based on mechanical hardware-in-the-loop, *Mechanical Systems and Signal Processing*, **166**, 108455
8. Lou Z., Ervin R.D., Filisko F.E., 1994, A preliminary parametric study of electrorheological dampers, *Journal of Fluids Engineering*, **116**, 3, 570-576
9. Ma T., Bi F., Wang X., Tian C., Lin J., *et al.*, 2021, Optimized fuzzy skyhook control for semi-active vehicle suspension with new inverse model of magnetorheological fluid damper, *Energies*, **14**, 6, 1674
10. Maciejewski I., Krzyzynski T., Pecolt S., Chamera S., 2019, Semi-active vibration control of horizontal seat suspension by using magneto-rheological damper, *Journal of Theoretical and Applied Mechanics*, **57**, 2, 411-420
11. Mai V.N., Yoon D.-S., Choi S.-B., Kim G.-W., 2020, Explicit model predictive control of semi-active suspension systems with magneto-rheological dampers subject to input constraints, *Journal of Intelligent Material Systems and Structures*, **31**, 9, 1157-1170
12. Morales A.L., Nieto A.J., Chicharro J.M., Pintado P., 2018, A semi-active vehicle suspension based on pneumatic springs and magnetorheological dampers, *Journal of Vibration and Control*, **24**, 4, 808-821
13. Muthalif A.G.A., Kasemi H.B., Nordin N.H.D., Rashid M.M., Razali M.K.M., 2017, Semi-active vibration control using experimental model of magnetorheological damper with adaptive F-PID controller, *Smart Structures and Systems*, **20**, 1, 85-97
14. Na J., Huang Y., Wu X., Liu Y.-J., Li Y., Li G., 2022, Active suspension control of quarter-car system with experimental validation, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **52**, 8, 4714-4726
15. Na J., Huang Y., Wu X., Su S.-F., Li G., 2020, Adaptive finite-time fuzzy control of nonlinear active suspension systems with input delay, *IEEE Transactions on Cybernetics*, **50**, 6, 2639-2650
16. Pang H., Liu F., Xu Z., 2018, Variable universe fuzzy control for vehicle semi-active suspension system with MR damper combining fuzzy neural network and particle swarm optimization, *Neurocomputing*, **306**, 130-140
17. Papaioannou G., Koulocheris D., Velenis E., 2021, Skyhook control strategy for vehicle suspensions based on the distribution of the operational conditions, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **235**, 10-11, 2776-2790
18. Shin D.K., Phu D.X., Choi S.M., Choi S.B., 2016, An adaptive fuzzy sliding mode control of magneto-rheological seat suspension with human body model, *Journal of Intelligent Material Systems and Structures*, **27**, 7, 925-934
19. Sosthene K., Josee M., Xiong H., 2018, Fuzzy logic controller for semi active suspension based on magneto-rheological damper, *International Journal of Automotive Engineering and Technologies*, **7**, 1, 38-47
20. Tang X., Ning D., Du H., Li W., Wen W., 2020, Takagi-Sugeno fuzzy model-based semi-active control for the seat suspension with an electrorheological damper, *IEEE Access*, **8**, 98027-98037
21. Trikande M.W., Karve N.K., Anand Raj R., Jagirdar W., Vasudevan R., 2018, Semi-active vibration control of an 8x8 armored wheeled platform, *Journal of Vibration and Control*, **24**, 2, 283-302
22. Tseng H.E., Hrovat D., 2015, State of the art survey: active and semi-active suspension control, *Vehicle System Dynamics*, **53**, 7, 1034-1062

23. Wang W., Song Y., Xue Y., Jin H., Hou J., Zhao M., 2015, An optimal vibration control strategy for a vehicle's active suspension based on improved cultural algorithm, *Applied Soft Computing*, **28**, 167-174

24. Zeng W., Jiang Q., Xie J., Yu T., 2020, A functional variable universe fuzzy PID controller for load following operation of PWR with the multiple model, *Annals of Nuclear Energy*, **140**, 107174

25. Zhang B., Zhao H., Zhang X., 2024, Adaptive variable domain fuzzy PID control strategy based on road excitation for semi-active suspension using CDC shock absorber, *Journal of Vibration and Control*, **30**, 3-4, 860-875

26. Zhang N., Zhao Q., 2017, Fuzzy sliding mode controller design for semi-active seat suspension with neuro-inverse dynamics approximation for MR damper, *Journal of Vibroengineering*, **19**, 5, 3488-3511

27. Zhong C., Li G., Meng Z., 2022, Beluga whale optimization: A novel nature-inspired metaheuristic algorithm, *Knowledge-Based Systems*, **251**, 109215

Manuscript received November 7, 2023; accepted for publication September 19, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 713-719, Warsaw 2024
https://doi.org/10.15632/jtam-pl/193829

ASYMMETRIC DUFFING OSCILLATOR: THE BIRTH AND BUILD-UP OF A PERIOD-DOUBLING CASCADE

Jan Kyzioł, Andrzej Okniński
Kielce University of Technology, Kielce, Poland
e-mail: kyziol@tu.kielce.pl; fizao@tu.kielce.pl

We investigate the period-doubling phenomenon in aperiodically forced asymmetric Duffing oscillator. We use the known steady-state asymptotic solution – the amplitude-frequency implicit function – and known criterion for the existence of period-doubling, also in an implicit form. Working in the framework of differential properties of implicit functions, we derive analytical formulas for the birth of period-doubled solutions.

Keywords: Duffing equation, period-doubling, implicit functions

1. Introduction and motivation

A period-doubling cascade of bifurcations is a typical route to chaos in nonlinear dynamical systems. We shall study this phenomenon in Duffing-type oscillator equations.

In this work, we study period-doubling in a forced asymmetric Duffing oscillator governed by the non-dimensional equation

$$\ddot{y} + 2\zeta\dot{y} + \gamma y^3 = F_0 + F\cos(\Omega t) \tag{1.1}$$

which has a single equilibrium position and a corresponding one-well potential (Kovacic and Brennan, 2011), where ζ, γ, F_0, F are parameters and Ω is the angular frequency of the periodic force.

The period-doubling scenario in dynamical system (1.1) was investigated by Szemplińska-Stupnicka in a series of groundbreaking papers (Szemplińska-Stupnicka and Bajkowski, 1986; Szemplińska-Stupnicka, 1987, 1988), see also (Kovacic and Brennan, 2011) for a review and further results, and (Xu and Luo, 2020) for another approach to period-doubling in a similar system.

The main idea introduced in (Szemplińska-Stupnicka and Bajkowski, 1986) consists of perturbing the main steady-state (approximate) asymptotic solution of Eq. (1.1), a $1:1$ resonance $y_0(t)$

$$\begin{aligned} y(t) &= y_0(t) + B\cos\Big(\tfrac{1}{2}\Omega t + \varphi\Big) \\ y_0(t) &= A_0 + A_1\cos(\Omega t + \theta) \end{aligned} \tag{1.2}$$

The perturbed solution $y(t)$ is substituted into Eq. (1.1), and the condition $B \neq 0$ is demanded. In papers (Kovacic and Brennan, 2011; Szemplińska-Stupnicka and Bajkowski, 1986; Szemplińska-Stupnicka, 1987, 1988) several conditions guaranteeing the formation and stability of solution (1.2) were found and used to study the period-doubling phenomenon.

For example, these authors were able to find intervals (Ω_1, Ω_2) in which solution $(1.2)_2$ destabilized via the formation of period-doubled solution $(1.2)_1$ (Kovacic and Brennan, 2011. They also demonstrated that a cascade of period doubling leading to chaos was formed (Kovacic

and Brennan, 2011; Szemplińska-Stupnicka and Bajkowski, 1986; Szemplińska-Stupnicka, 1987, 1988).

Our motivation is fueled by the observation that steady-state solution $(1.2)_2$ as well as period-doubling conditions found in (Kovacic and Brennan, 2011; Szemplińska-Stupnicka and Bajkowski, 1986; Szemplińska-Stupnicka, 1987, 1988) are in the form of some implicit functions. Therefore, it should be possible, within the framework of differential properties of implicit functions (Fikhtengolt's, 1965; Kyzioł ad Okniński, 2022), to obtain new results concerning the period-doubling mechanism.

The aim of the present work is thus to apply this formalism to implicit functions derived in (Kovacic and Brennan, 2011; Szemplińska-Stupnicka and Bajkowski, 1986; Szemplińska--Stupnicka, 1987, 1988).

The paper is organized as follows. In Section 2, a steady-state solution of Eq. (1.1) of form $(1.2)_2$ is reviewed and a period-doubling condition derived in (Kovacic and Brennan, 2011; Szemplińska-Stupnicka and Bajkowski, 1986; Szemplińska-Stupnicka, 1987, 1988) is described in Section 3. In Section 4, we derive new results concerning period-doubling applying the formalism of differential properties of implicit functions, and in Section 5 we verify our results. In Section 6, we summarize our findings.

2. The main resonance: steady-state solution

The steady-state solution of Eq. (1.1) of form $(1.3)_2$, describing the $1:1$ resonance, was computed in (Szemplińska-Stupnicka and Bajkowski, 1986; Jordan and Smith, 1999, Kovacic and Brennan, 2011). Proceeding as in (Kyzioł and Okniński, 2023), we get two implicit equations for A_0, A_1 and Ω

$$\begin{aligned} &A_1^2\Big(3\gamma A_0^2+\frac{3}{4}\gamma A_1^2-\Omega^2\Big)^2+4\Omega^2\zeta^2A_1^2=F^2\\ &\gamma A_0^3+\frac{3}{2}\gamma A_0A_1^2-F_0=0 \end{aligned} \tag{2.1}$$

Computing A_1^2 from Eq. $(2.1)_2$ for $A_0\neq 0$ and substituting into $(2.1)_1$, we obtain finally one implicit equation for A_0, Ω (Kovacic and Brennan, 2011; Kyzioł and Okniński, 2023)

$$f(A_0,\Omega,\gamma,\zeta,F_0,F)=\sum_{k=0}^{9}c_kA_0^k=0 \tag{2.2}$$

where the coefficients c_k are given in Table 1 (cf. Eq. (8.3.12) in (Kovacic and Brennan, 2011).

Table 1. Coefficients c_k of polynomial (2.2)

$c_0=-F_0^3$	$c_5=4\gamma\Omega^2(\Omega^2+4\zeta^2)$
$c_1=4\Omega^2F_0^2$	$c_6=-15\gamma^2F_0$
$c_2=-4F_0\Omega^4-16\zeta^2\Omega^2F_0$	$c_7=-20\Omega^2\gamma^2$
$c_3=-9\gamma F_0^2+6\gamma F^2$	$c_8=0$
$c_4=16\Omega^2\gamma F_0$	$c_9=25\gamma^3$

We can also obtain an implicit equation for A_1, Ω as in (Kyzioł and Okniński, 2023). Solving Eq.$(2.1)_2$ for A_0 (there is only one real root) and substituting to $(2.1)_1$, we get

$$g(A_1,\Omega,\gamma,\zeta,F,F_0)=A_1^2\Big(3\gamma A_0^2+\frac{3}{4}\gamma A_1^2-\Omega^2\Big)^2+4\Omega^2\zeta^2A_1^2-F^2=0 \tag{2.3}$$

where A_0 and Y are defined as

$$A_0 = -\frac{A_1^2}{2Y} + Y \qquad Y = \sqrt[3]{\sqrt{\frac{1}{8}A_1^6 + \frac{1}{4\gamma^2}F_0^2} + \frac{1}{2\gamma}F_0} \tag{2.4}$$

3. Birth of period-doubling

The stability of the steady-state solution $y_0(t) = A_0 + A_1\cos(\Omega t + \theta)$ is studied via substitution into Eq. (1.1): $y = y_0 + u(t)$, with $u(t)$ small. This substitution leads to (linear) Hill's equation for $u(t)$ (Kovacic and Brennan, 2011; Szemplińska-Stupnicka and Bajkowski, 1986)

$$\begin{aligned} &\widehat{L}u \equiv \ddot{u} + 2\zeta\dot{u} + [\sigma_0 + \sigma_1\cos(\Omega t + \theta) + \sigma_2\cos(2(\Omega t + \theta))]u = 0 \\ &\sigma_0 = 3\gamma A_0^2 + \frac{3}{2}\gamma A_1^2 \qquad \sigma_1 = 6\gamma A_0 A_1 \qquad \sigma_2 = \frac{3}{2}\gamma A_1^2 \end{aligned} \tag{3.1}$$

provided that higher powers of u are neglected.

To study destabilization of the $1:1$ resonance via the period-doubling scenario, one puts into Eq. (3.1)

$$u(t) = B\cos\Big(\frac{1}{2}\Omega t + \varphi\Big) \tag{3.2}$$

obtaining, after a harmonic balance method is used, a simple necessary condition for the onset of period-doubling (a condition for non-zero B)

$$h(A_0, A_1, \Omega; \gamma, \zeta) \equiv \Big(\sigma_0 - \frac{1}{4}\Omega^2\Big)^2 + \zeta^2\Omega^2 - \frac{1}{4}\sigma_1^2 = 0 \tag{3.3}$$

see Eq. (8.5.5) in (Kovacic and Brennan, 2011) or Eq. (4d) in (Szemplińska-Stupnicka and Bajkowski, 1986).

Equation (3.3) can be simplified. We compute A_1^2 from Eq. $(2.1)_2$ for $A_0 \neq 0$ and substitute it into Eq. (3.3), obtaining a simplified condition for the birth of period-doubling

$$\begin{aligned} k(A_0, \Omega; \gamma, \zeta, F_0) &= 10\gamma^2 A_0^6 - \gamma\Omega^2 A_0^4 - 2\gamma F_0 A_0^3 \\ &+ \Big(\frac{1}{16}\Omega^4 + \zeta^2\Omega^2\Big)A_0^2 - \frac{1}{2}F_0\Omega^2 A_0 + F_0^2 = 0 \end{aligned} \tag{3.4}$$

4. Differential condition for period-doubling

We are going to show that there is a differential condition that permits a further insight into the nature of the birth of period doubling.

We consider a quite obvious differential condition for the birth of period-doubling. More precisely, we investigate when equations (2.2), (3.4) have a common (real) root (A_0, Ω). Equation (2.2) guarantees that an isolated point of implicit function (3.4) lies in the $1:1$ resonance amplitude-frequency curve (2.2) – a double solution of Eqs. (2.2), (3.4). This corresponds to a birth of instability of the $1:1$ resonance with the creation of period-doubled solution (3.2), i.e. to a birth of period-doubling.

Accordingly, we consider the following equations

$$\begin{aligned} &f(A_0, \Omega; \gamma, \zeta, F_0, F) = 0 \qquad k(A_0, \Omega; \gamma, \zeta, F_0) = 0 \\ &\frac{\partial k(A_0, \Omega, \gamma, \zeta, F_0)}{\partial A_0} = 0 \qquad \frac{\partial k(A_0, \Omega, \gamma, \zeta, F_0)}{\partial \Omega} = 0 \end{aligned} \tag{4.1}$$

where Eq. $(4.1)_1$, equivalent to (2.2), is the steady-state condition for the $1:1$ resonance, Eq. $(4.1)_2$ is period-doubling condition (3.4) and equations $(4.1)_{3,4}$ mean that implicit function (3.4) has a singular point. We show below that this singular point is an isolated point of (3.4).

Acceptable solutions of Eqs. (4.1), i.e $\Omega > 0$, $A_0 > 0$, $\gamma > 0$, $F > 0$ are

$$\Omega_* = c_1\zeta \qquad A_{0*} = c_2\frac{F_0}{\zeta^2} \qquad \gamma_* = c_3\frac{\zeta^6}{F_0^2} \qquad F_* = c_4F_0 \tag{4.2}$$

and

$$\begin{aligned} c_1 &= 2\sqrt{2}\sqrt{\sqrt{2}+1} \cong 4.395 \qquad & c_2 &= \frac{3}{4}\Big(1-\frac{1}{2}\sqrt{2}\Big) \cong 0.220 \\ c_3 &= \frac{32}{27}(7\sqrt{2}+10) \cong 23.585 \qquad & c_4 &= \frac{3}{8}\sqrt{2}\sqrt{16\sqrt{2}+73} \cong 5.186 \end{aligned} \tag{4.3}$$

where $\zeta > 0$, $F_0 > 0$ are free parameters. These solutions have been computed using Maple from Scientific WorkPlace 4.0. Finally, we compute A_{1*}^2 from Eqs. $(2.1)_2$, (4.2) and (4.3) and check that inequality $A_{1*}^2 > 0$ is fulfilled for $F_0 > 0$ since $(3/2)\gamma A_{0*}A_{1*}^2 = F_0 - \gamma_* A_{0*}^3 = (3/4)F_0 > 0$.

Now we demonstrate that solution (4.2) and (4.3) corresponds to an isolated point. The determinant of the Hessian matrix, computed for the function $k(A_0, \Omega, \zeta, \gamma, F_0 t)$ at singular point (4.2) and(4.3), is positive

$$\det\begin{bmatrix} \dfrac{\partial^2 k(A_0,\Omega)}{\partial\Omega^2} & \dfrac{\partial^2 k(A_0,\Omega}{\partial\Omega\partial A_0} \\ \dfrac{\partial^2 k(A_0,\Omega)}{\partial A_0\partial\Omega} & \dfrac{\partial^2 k(A_0,\Omega)}{\partial A_0^2} \end{bmatrix} = (36\sqrt{2}+18)\zeta^2F_0^2 > 0 \tag{4.4}$$

and this means that this is an isolated point (Fikhtengolt's, 1965).

For example, if we choose $\gamma = 0.1$ and $F_0 = 0.02$ then we compute from Eqs. (4.2) and (4.3) other parameters of the isolated point, ζ_*, F_*, as well as Ω_*, A_{0*}, and A_{1*} from Eq. $(2.1)_2$, listed in Table 2.

Table 2. Variables and parameters computed from Eqs. (4.2) and (4.3), $\gamma = 0.1$, $F_0 = 0.02$

ζ_*	γ	F_0	F_*	Ω_*	A_{0*}	A_{1*}
0.109204	0.1	0.02	0.103721	0.479923	0.368403	0.521001

And indeed, if we solve Eqs. $(4.1)_{1,2}$ for this set of parameters, we obtain a double solution $(A_{0*}, \Omega_*) = (0.368403, 0.479923)$, see also Fig. 1, where the singular point (A_{0*}, Ω_*) – an isolated point of implicit function (3.4) – is shown as a red dot.

For $\zeta > \zeta_*$ solutions of Eq. $(4.1)_2$ are complex, for $\zeta = \zeta_*$ there is a real isolated point (a double root of Eqs. (4.1)) lying in curve $(4.1)_1$ – a red dot in Fig. 1, and for decreasing values of ζ curves $(4.1)_2$ are growing blue ovals. More exactly, in Fig. 1 we have $\gamma = 0.1$,$F_0 = 0.02$, $F = 0.103721$ while $\zeta = 0.109204$ (singular – isolated point), 0.109, 0.108, 0.105, 0.100.

Note that for decreasing ζ implicit function $(4.1)_1$, describing the $1:1$ resonance changes only slightly while implicit function $(4.1)_2$, destabilization condition of the resonance, changes significantly.

5. Numerical verification

It follows from Section 4 that for $\gamma = 0.1$, $F_0 = 0.02$, $F = 0.103721$, destabilization of the $1:1$ resonance occurs for $\zeta \leqslant \zeta_* = 0.109204$. Therefore, we have computed bifurcation diagrams solving Eq. (1.1) for $\gamma = 0.1$, $F_0 = 0.02$, $F = 0.103721$, and $\zeta \approx 0.109204$ looking for an onset of period-doubling. And indeed, the $1:1$ resonance becomes unstable for $\zeta \cong 0.1105$.

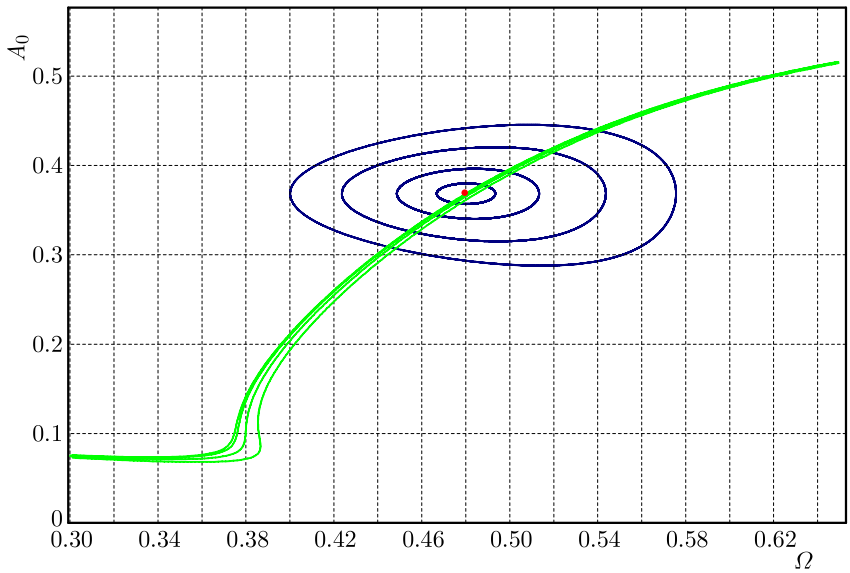


Fig. 1. Implicit functions: $(4.1)_1$ (green), $(4.1)_2$ (blue), isolated point (red); $\gamma = 0.1$, $F_0 = 0.02$, $F = 0.103721$ and $\zeta = 0.109204$ (singular), 0.109, 0.108, 0.105, 0.100

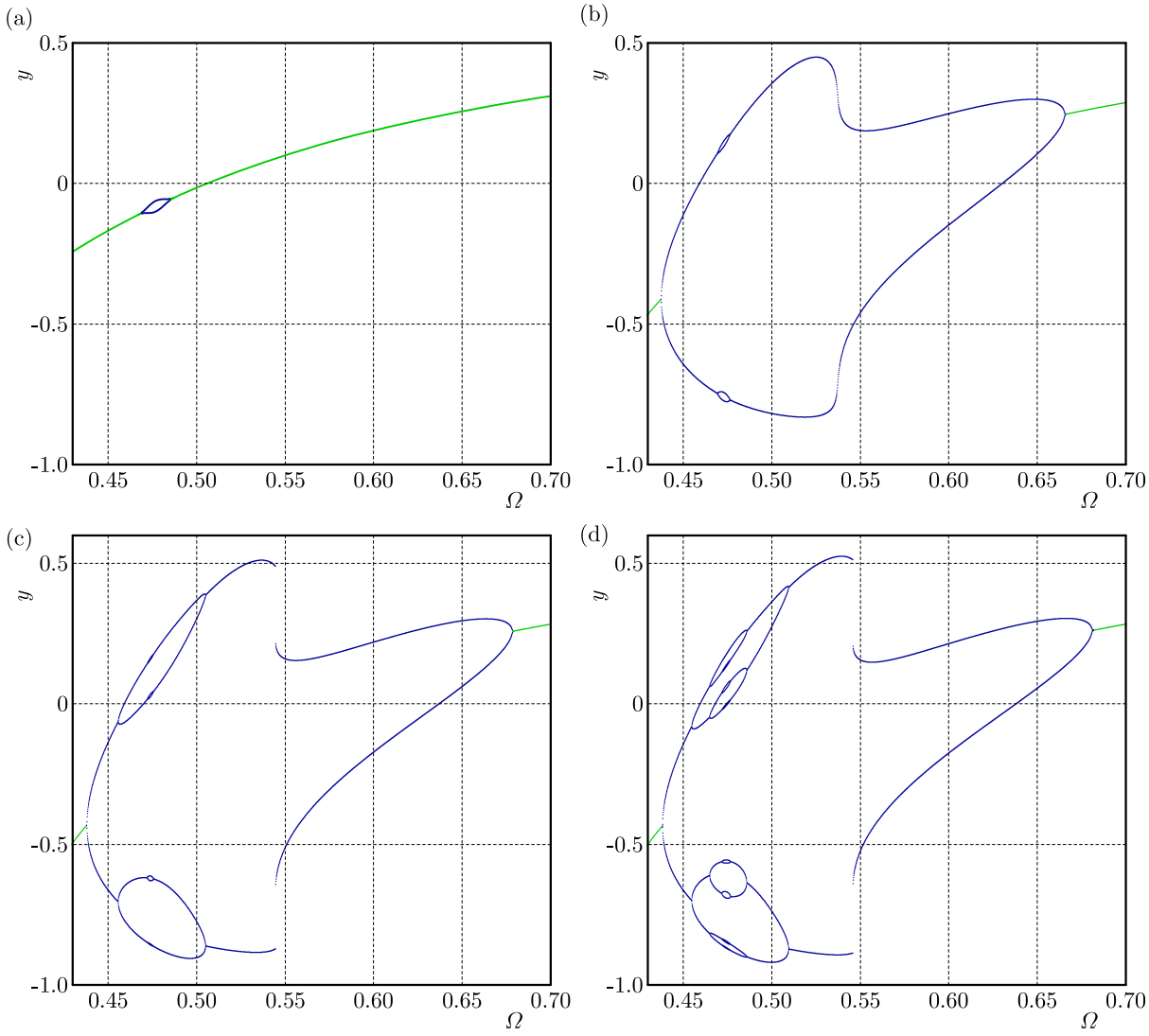


Fig. 2. Bifurcation diagrams: y against Ω, $\gamma = 0.1$, $F_0 = 0.02$, $F = 0.103721$ and $\zeta = 0.1105$, 0.0621, 0.0562, 0.0550 in diagrams 1, 2, 3, 4, respectively; $1:1$ (green) and period-doubled (blue) resonances

Numerical solutions of $y(t)$ from Eq. (1.1) (bifurcation diagrams) were computed running DYNAMICS (Nusse and York, 1998) and Wolfram Mathematica 12.1 (Wolfram, 2020) in the interval $\Omega \in (0.43, 0.70)$ for $\gamma = 0.1$, $F_0 = 0.02$, $F = 0.103721$ and $\zeta = 0.1105$, 0.0621, 0.0562, 0.0550, see Fig. 2. The initial conditions were $y(0) = 0$, $\dot{y}(0) = 0$.

We note that destabilization of the $1:1$ resonance with the formation of $1:2$ solution (1.2) (as well as other resonances) appears at $\Omega = 0.47\ldots$ in a good agreement with the analytical value $\Omega_* = 0.479923$ in Table 2.

Moreover, we have computed numerical values of parameter ζ at which the first and subsequent period-doubling bifurcations occur, see Table 3. More precisely, we have been computing bifurcation diagrams for decreasing values of ζ until the period-doubled solution was visible in the corresponding bifurcation diagram (see, for example, diagram 1 in Fig. 2). The first period--doubling takes place at $\zeta_1 = 0.1105334$ in good agreement with analytical value $\zeta_* = 0.109204$, see Table 2. We have also computed ratios $(\zeta_{i-1} - \zeta_{i-2})/(\zeta_i - \zeta_{i-1})$ which converge quite well to the Feigenbaum constant $\delta = 4.6692011609\ldots$ (Feigenbaum, 1978).

Table 3. Period-doubling bifurcations, $\gamma = 0.1$, $F_0 = 0.02$, $F = 0.103721$

i	period 2^i	Ω_i	ζ_i	$\dfrac{\zeta_{i-1} - \zeta_{i-2}}{\zeta_i - \zeta_{i-1}}$
1	2^1	0.477000	0.1105334	–
2	2^2	0.473200	0.0622045	–
3	2^3	0.474200	0.0562256	8.083
4	2^4	0.474550	0.0550668	5.160
5	2^5	0.474515	0.0548240	4.772
6	2^6	0.474515	0.0547719	4.660

6. Summary

Based on the known amplitude-frequency steady-state equation for $1:1$ resonance (2.2) and period-doubling condition (3.4), we have computed a two-parameter family of solutions of Eqs. (4.1) – see Eqs. (4.2) and(4.3). Equations $(4.1)_{2,3,4}$ guarantee that these solutions are singular points of period-doubling condition (3.4) while Eq. $(4.1)_1$, equivalent to Eq. (2.2), entails that solutions of equations (4.1) fulfill also the $1:1$ resonance condition.

We have demonstrated that these singular points are isolated points (cf. Eq. (4.4)) which due to Eq. $(4.1)_1$ lie in the amplitude-frequency curves of the $1:1$ resonance as in Fig. 1. The emergence of an isolated point, a double real root of Eqs. (4.1), corresponds thus to the onset of period-doubling.

Furthermore, we have obtained a good agreement between analytical value $\zeta = \zeta_* = 0.109204$, Table 2, and numerical value $\zeta = \zeta_1 = 0.1105334$ for the onset of period-doubling, Table 3. Moreover, the first period-doubling occurs for $\Omega_1 = 0.477000$ in agreement with analytical value $\Omega_* = 0.479923$.

It is possible to control destabilization of the $1:1$ resonance by decreasing ζ, $\zeta < \zeta_*$. Indeed, it follows from Fig. 2 and Table 3 that upon decreasing ζ, we observe a build-up of the Feigenbaum cascade of period-doubling, leading to chaos. We note that all period-doubling occur for $\Omega \in (0.47, 0.48)$, and this suggests that in the case of higher period-doubling there is a similar mechanism at work.

We hope, that our approach can be applied to other non-linear dynamical systems – periodically forced nonlinear ordinary differential equations.

A. Computational details

Nonlinear polynomial equations (4.1) were solved using the computational engine Maple 4.0 from Scientific WorkPlace 4.0. Figure 1 was plotted with the computational engine MuPAD 4.0 from Scientific WorkPlace 5.5. Bifurcation diagrams in Fig. 2 were computed by integrating numerically Eq. (1.1) running DYNAMICS (Nusse and York, 1998) and Wolfram Mathematica 12.1 (Wolfram, 2020).

References

1. Feigenbaum M.J., 1978, Quantitative universality for a class of nonlinear transformations, *Journal of Statistical Physics*, **19**, 25-52
2. Fikhtengol'ts G.M., 1965, *The Fundamentals of Mathematical Analysis*, I.N. Sneddon (Edit.), **2**, Chapter 19, Elsevier (translated from Russian)
3. Jordan D.W., Smith P., 1999, *Nonlinear Ordinary Differential Equations: An Introduction to Dynamical Systems*, Oxford University Press, New York
4. Kovacic I., Brennan M.J., 2011, Forced harmonic vibration of an asymmetric Duffing oscillator, [In:] *The Duffing Equation: Nonlinear Oscillators and Their Behavior*, I. Kovacic, M.J. Brennan (Edit.), John Wiley & Sons, Hoboken, New Jersey, 277-322
5. Kyzioł J., Okniński A., 2022, Localizing bifurcations in non-linear dynamical systems via analytical and numerical methods, *Processes*, **10**, 1, 127
6. Kyzioł J., Okniński A., 2023, Asymmetric Duffing oscillator: jump manifold and border set, *Nonlinear Dynamics and Systems Theory*, **23**, 1, 46-57
7. Nusse H.E., Yorke J.A., 1998, *Dynamics: Numerical Explorations. Accompanying Computer Program Dynamics*, Applied Mathematical Sciences, **101**, Springer
8. Szemplińska-Stupnicka W., 1987, Secondary resonances and approximate models of routes to chaotic motion in non-linear oscillators, *Journal of Sound and Vibration*, **113**, 1, 155-172
9. Szemplińska-Stupnicka W., 1988, Bifurcations of harmonic solution leading to chaotic motion in the softening type Duffing's oscillator, *International Journal of Non-Linear Mechanics*, **23**, 4, 257-277
10. Szemplińska-Stupnicka W., Bajkowski J., 1986, The 12 subharmonic resonance and its transition to chaotic motion in a non-linear oscillator, *International Journal of Non-Linear Mechanics*, **21**, 5, 401-419
11. Wolfram Research, Inc., 2020, Mathematica, Version 12.1, Champaign, IL
12. Xu Y., Luo A.C.J., 2020, Independent period-2 motions to chaos in a van der Pol-Duffing oscillator, *International Journal of Bifurcation and Chaos*, **30**, 15, 2030045

Manuscript received January 8, 2024; accepted for publication September 27, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 721-735, Warsaw 2024
https://doi.org/10.15632/jtam-pl/194237

BEARING LIFE PREDICTION MODEL FOR ELECTROMECHANICAL EQUIPMENT BY INTEGRATING DEEP NEURAL NETWORK AND K-NEAREST NEIGHBOR ALGORITHM AND ITS APPLICATION

Yiming Ma
Jilin Jianzhu University, Changchun, China
e-mail: mayiming0131@163.com

Current life prediction methods of Electromechanical equipment bearings have issues of low accuracy and lack of stability. To address these problems, firstly, indicators based on life degradation characteristics of bearings are selected. Then, a deep neural network-based life prediction model is constructed. Finally, the K-nearest neighbor algorithm is introduced to correct the deviation of the deep neural network prediction model, and a hybrid life prediction model is designed. Results show that effectiveness of the designed model was better, which was of great practical significance for detecting bearing failures in advance, reducing equipment losses and improving equipment reliability.

Keywords: bearings, deep neural network, K-nearest neighbor algorithm, deviation, life prediction model

1. Introduction

In recent years, electromechanical equipment plays a vital role in the modern industrial production. And one of the most common problems in equipment failure is life prediction of bearings. As an important part of electromechanical equipment, accurate prediction of the state of bearings is critical for normal operation and maintenance of the equipment. Traditional bearing life prediction methods mainly rely on empirical models and mathematical statistical methods, which can provide reliable prediction results in some cases, but have certain limitations in complex working conditions and variable environments (Zhang *et al.*, 2023). Deep Neural Network (DNN) is a multi-layer neuron based artificial intelligence model that can automatically extract features and perform complex pattern recognition by learning a large amount of data. It has powerful nonlinear modeling capability and high adaptivity, and is proper for processing high dimensional, nonlinear and large-scale data (Abdou, 2022). The K-Nearest Neighbors (KNN) algorithm, on the other hand, is a machine learning method that makes predictions by calculating the distances between samples and using information from the nearest neighbor samples. The algorithm is simple and effective, and has a good ability to handle small samples and nonlinear problems (Hatem, 2022). In this context, DNN and KNN are fused to design a DNN-KNN-based bearing life prediction model. Firstly, the indexes are selected with characteristics of bearing life degradation, then the remaining life prediction model is constructed by DNN, and finally KNN is introduced to correct the deviation of the model to improve the accuracy and stability of the prediction. The research is composed of five parts. The first section is the background of bearing life prediction. The second section is a review of the current research status of bearing life prediction. The third section is the construction of a bearing life prediction model, which includes three sections: index selection, model construction, and model optimization. The fourth part is the experimental outcome of the model, in which the first and the second sections are the performance and application effect analysis of the designed algorithm. The fifth section is the summary of the whole paper and shortcomings of the study.

2. Literature review

With the increasing importance of electromechanical equipment in the industrial production and life, bearing, as a core component of electromechanical equipment, its life prediction is of great significance for the operational stability and economic benefits of the equipment. In the field of bearing life prediction of electromechanical equipment, many scholars and researchers have proposed various methods and models to solve this problem. Nistane (2024) designed a rolling bearing remaining life optimization prediction model based on integrated optimization health indicators and hybrid machine learning algorithms to understand the degree of deterioration of rolling element bearings at any time. The model preprocesses the original signal data through wavelet transform and optimizes features using machine learning techniques. The results show that the prediction error of the model is low. Li and Wang (2024) designed a method that integrates time series window features and first prediction time recognition to predict the remaining life of rolling bearings using limited data. The method uses a multi-step rolling prediction strategy based on the first prediction time degradation factor to reveal the future degradation trend of bearings, and the results show a prediction error as low as 8.06%. Fei *et al.* (2024) have designed a life prediction method based on convolutional deep neural networks to improve the accuracy of turbocharger bearing life prediction. The method uses convolutional neural networks to extract feature convolutional layers and uses deep neural networks to regress and model the bearing life. The results show that this method has high efficiency and accuracy. Mao *et al.* (2019) designed a gated recurrent unit neural network for prediction of bearing life by providing sufficient feature representation and adaptive extraction for the bearing life prediction, and the findings indicated that the model had a high generalization ability and accuracy. To provide sufficient feature representation and adaptive extraction for bearing life prediction, Mao *et al.* (2019) designed a remaining life prediction method based on deep feature characterization and migration learning, which was divided into two phases to characterize the faults and correct the features to complete the prediction, and the outcomes expressed that the method had a better numerical stability and prediction accuracy. Sun *et al.* (2021) designed a method with vibration signal detection to predict the life of conventional low-voltage circuit breakers, which combined the designed signal processing method to extract effective vibration segments that characterized mechanical properties of the contact system to construct a model, and the findings illustrated that the method had a high average fit. Rezamand and other scholars determined the effect of operating conditions on the dynamics of bearing failure to achieve a hybrid prediction based on real-time monitoring and data acquisition and a vibration signal prediction method which used vibration signals to identify fault dynamics of the life of bearing prediction. A hybrid prediction method based on real-time monitoring and data acquisition and a vibration signal prediction was designed to identify the failure dynamics through vibration signals and Bayesian algorithm, which was shown to have a high prediction accuracy (Rezamand *et al.*, 2021).

To address the issue of non-existence of interpretability in bearing life prediction, Ding and other researchers designed a dynamic structure and adaptive notation method, which modeled the health indexes of multiple signals and tracked the real-time degradation of the machine by using dynamic coupling terms. The findings denoted that the method had better generalization ability and lower prediction error (Ding *et al.*, 2021). To solve the uncertainty of a recurrent neural network in predicting bearing life, Wang and other scholars designed a long and short-term memory network model based on residual convolution, which quantified the inaccuracy of prediction results by constructing an appropriate output layer obeying the normal distribution, and the findings denoted that the model could effectively predict the bearing life (Wang *et al.*, 2022). Rezamand *et al.* (2020) to accurately estimate the prediction of bearing life, designed a comprehensive prediction method based on the signal processing and adaptive Bayesian algorithm, which was based on feature extraction and feature selection. It detected, dynamics of

various faults through feature extraction, feature selection and signal denoising and predicted them by using the adaptive Bayesian algorithm, and the results showed that the method had a high prediction accuracy (Rezamand *et al.*, 2020). In order to achieve accurate prediction of remaining lifespan, Wu *et al.* (2024) designed a residual lifespan prediction model based on wavelet enhanced dual tree residual network. The model decomposed time series through wavelet transform and predicted remaining lifespan by concatenating dual tree features. The results showed that the prediction effect of that method was good (Wu *et al.*, 2024). Li *et al.* (2023) researchers designed a deep adversarial network-based residual service life prediction method for partial sensor failure to achieve a good electromechanical health assessment, which extracted generalized sensor invariant features through adversarial learning to make a full use of the information from different sensors, and the findings indicated that the method had a high robustness. Yang *et al.* (2024) designed a dynamic spatiotemporal graph driven bearing remaining life prediction method based on graph data expansion to maintain normal operation of the machine. The method captured hidden information using short-time Fourier transform and predicted through a graph embedding module based on an autoencoder. The results showed that the prediction performance of that method was high (Yang *et al.*, 2024).

In summary, many scholars have made significant contributions in the field of bearing life prediction for electromechanical equipment, however, there are still some limitations of these methods, such as limited universality, missing comparative analysis and insufficient interpretability. Therefore, the study is based on life degradation characteristics of bearings for indicator selection, followed by constructing a life prediction model using DNN. At the same time, to correct the error of the DNN prediction model, the K nearest neighbor algorithm is introduced and a hybrid life prediction model is designed to improve the model performance of measurement and generalization ability.

3. Construction of the bearing life prediction model for electromechanical equipment by integrating DNN and KNN

This Section focuses on the construction of the fusion bearing life prediction model. The first section is selection of feature indicators, the second section is construction of the DNN-based model, and the third section is introduction of KNN to optimize the model.

3.1. Selection of characteristic indicators for bearing life degradation of electromechanical equipment

As a key component of electromechanical equipment, the performance degradation of bearings will seriously affect the operation safety and stability of the whole equipment. In order to effectively monitor the performance status of bearings during operation and predict their life, it is crucial to extract characteristic quantities that can reflect the performance degradation indicators of bearings (Cao *et al.*, 2023). Therefore, one needs to select the life degradation index of the bearing. In actual working conditions, there are many nonlinear factors that affect the construction performance of health indicators for rolling bearings. Traditional indicator extraction methods are prone to losing the state information on rolling bearings, and therefore cannot extract true and effective state features from complex signals. Research is being carried out on introducing deep learning models to extract state indicators of bearings. The index extraction process is shown in Fig. 1.

In the operation of electromechanical equipment, the rolling bearing as a key support component. During its long time operation, due to the influence of various factors, there may be faults or an abnormal operation state. At that time, the vibration signal will be manifested out of the bearing running state corresponding to the characteristic quantity. These characteristics

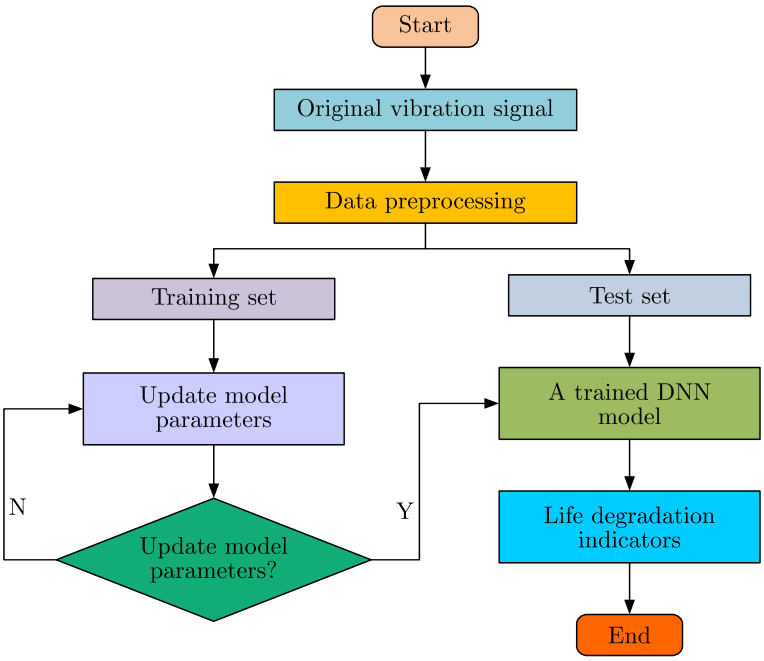


Fig. 1. Extraction process of bearing life degradation indicators

include time-domain features such as the mean value, peak value, RMS, craggyness, etc., as well as frequency-domain features such as center frequency, average energy, spectral partitioning and summation. In the study, the peak value, RMS, magnitude and spectral partition sum are selected as degradation characteristics of the bearing. Among them, the peak value is the maximum magnitude of the vibration signal, which can reflect the vibration intensity of the bearing, and its calculation method is expressed as

$$X_P = \max |x_i| \tag{3.1}$$

In Eq. (3.1), X_P represents the peak value, and x_i represents the input eigenvector, where $i = 1, 2, \ldots, N$. The RMS value of the vibration signal can reflect the vibration energy of the bearing, and is calculated as

$$X_R = \sqrt{\frac{1}{N}\sum_{i=1}^{n} x_i^2} \tag{3.2}$$

In Eq. (3.2), X_R indicates the RMS of the vibration signal, and N indicates the total amount of data in the sample. Sharpness of the vibration signal reflects the vibration frequency distribution of the bearing, and its calculation method is

$$X_K = \frac{1}{N}\frac{1}{X_R^4}\sum_{i=1}^{N}(|x_i - \overline{x}|)^4 \tag{3.3}$$

In Eq. (3.3), X_K represents the craggy index, and $\overline{x}$ represents the average value of the input eigenvectors. Spectral partitioning and summation is to divide the vibration signal spectrum into several intervals and calculate the sum of spectral energy in each interval, which can reflect the spectral distribution of the bearing vibration signal, and its calculation method is

$$X_F = \sum_{k=[V+N_\eta(r-1)]/V}^{N_\eta v/V} S(r) \tag{3.4}$$

In Eq. (3.4), X_F denotes the value of the sum of spectral partitions, which is a one-dimensional vector containing M elements, N_η denotes length of the signal spectrum, r means the amount of spectral lines in each spectrum, and $m = 1, 2, \ldots, M$. In the next step, data preprocessing is performed on the acquired feature parameters to raise the learning efficiency of the network and to prevent the occurrence of the gradient vanishing problem. That is, the data are standardized and normalized, and the calculation method is indicated by

$$o' = \frac{o - u}{\sigma} \qquad o'' = \text{round}\Big(\frac{o' - o_{min}}{o_{max} - o_{min}}\Big) \tag{3.5}$$

In Eq. (3.5), o denotes the collected data, o' denotes the normalized data, u means the mean value of all the data, σ refers to the standard deviation of all the data, o'' denotes the normalized data, which is dimensionless with the range of $[0, 1]$, o_{min} denotes the rounding function, o_{min} means the minimum value of the data, and o_{max} expresses the maximum value of the data. In summary, the peak value, RMS, craggyness and spectral partition summation can effectively reflect the performance degradation state, so they are selected as the degradation characterization indexes of the bearings.

3.2. DNN-based bearing life prediction model construction

With continuous growth of artificial intelligence technology, deep learning algorithms, as powerful tools, they have shown great potential in the field of bearing life prediction for electromechanical equipment (Bhosle and Musande, 2023). The study chooses DNN model to perform bearings life prediction based on the characteristic indicators of bearing life degradation. DNN consists of multiple layers of neurons, each with connections to adjacent layers, and can be used to handle complex nonlinear problems. It has been widely used in fields such as image and speech recognition, natural language processing, and recommendation systems. The DNN model architecture is shown in Fig. 2.

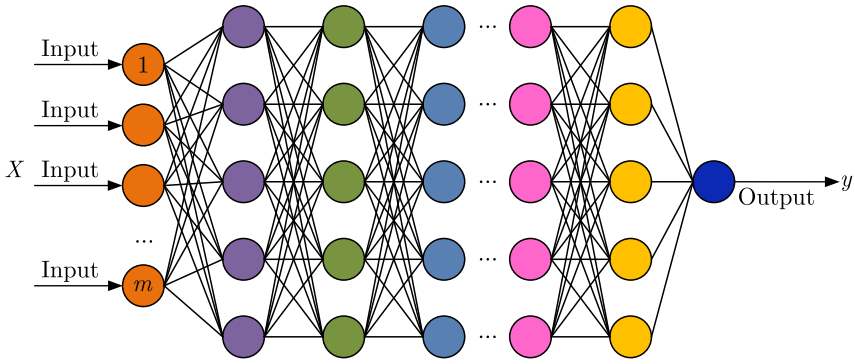


Fig. 2. DNN model architecture

The network structure of DNN is a typical deep learning model, which is composed of an inputting layer, multiple hidden layers and an outputting layer. The inputting layer is responsible for receiving m-dimensional input data and passing it to the next layer by linear transformation through an activation function. Repeating this process until the outputting layer is reached, the final outputting is obtained. The number of network layers of a DNN can vary depending on the application scenario, and some complex models can even reach 20 layers or more. This multi-layer structure can better capture complex features of the data and thus predict results more accurately. When training bearings using DNNs, the study first initializes all parameters by a random normal distribution, then uses the ReLu function in the middle layer of the network

and the Sigmoid function in the last layer to better match the normalized data. In terms of the loss function, an expression is shown as

$$F_M = \frac{1}{N}\sum_{i=1}^{N}(y_e^i - y_p^i)^2 \tag{3.6}$$

In Eq. (3.6), F_M represents the loss function, y_e^i represents the experimental value, and y_p^i represents the predicted value. The next step is to use the Adam function for optimization and train the DNN using the learning rate decay mechanism. The Adam optimizer is an optimization algorithm with an adaptive learning rate, which combines the advantages of momentum gradient descent and adaptive learning rate, and can adjust the learning rate of each parameter with the first-order and second-order moment estimation of the gradient, so as to optimize the model parameters effectively. When training the DNN, due to complexity of the network structure and uncertainty of the data, one needs to optimize the model performance by adjusting the learning rate, and the learning rate decay mechanism is able to raise the generalization ability and stability of the DNN model by gradually decreasing the learning rate, which is computed as

$$l_r = l_r(min) + [l_r(max) - l_r(min)]\mathrm{e}^{-\frac{itr}{d}} \tag{3.7}$$

In Eq. (3.7), l_r represents the learning rate. $l_r(max)$ and $l_r(min)$ mean the maximum and the minimum values of the learning rate, respectively. itr refers to the amount of iterations, and d represents the decay rate of the learning rate. After setting the parameters, the degradation characteristics are used as inputs to obtain the degradation trend of the bearing, and finally the remaining life of the bearing can be calculated by obtaining the moment of bearing failure according to the failure threshold, which is shown as

$$t_r = \{t' - t | t' > t, Z(t)\} \tag{3.8}$$

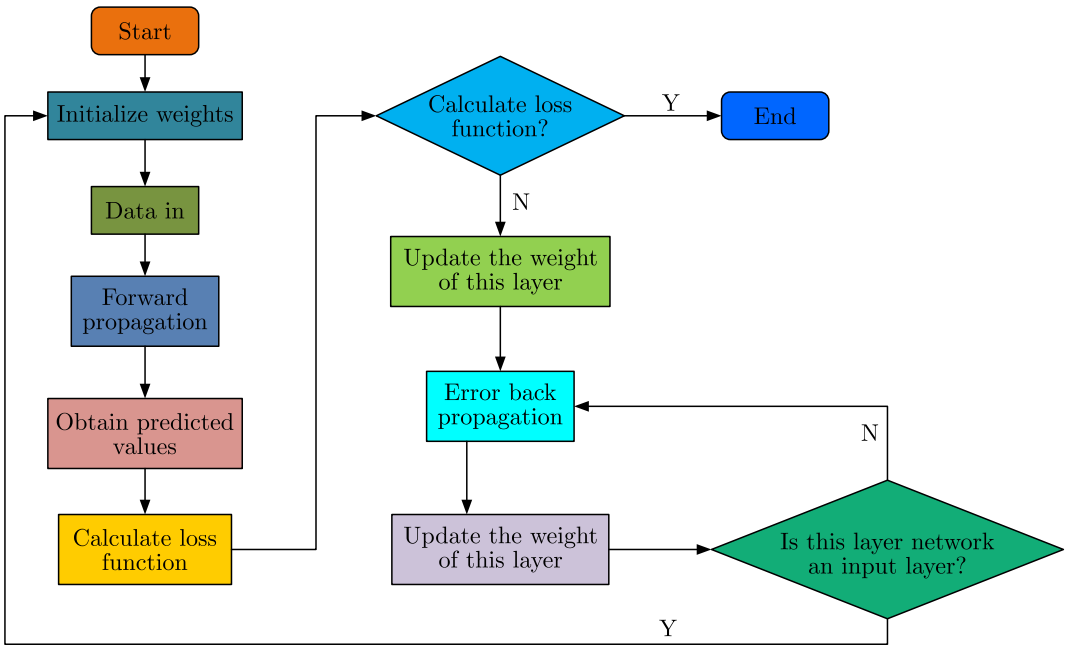


Fig. 3. DNN based bearing life prediction process

In Eq. (3.8), t_r represents the remaining life of bearing, t' represents the moment of failure, t represents the current operating moment of the bearing, and $Z(t)$ represents all historical

operating conditions of the bearing up to the current moment. The bearing life prediction process based on the DNN is shown in Fig. 3.

Firstly, the data preprocessing is performed to normalize life vibration signals of the original bearings and label them with health indicators, as well as the ash content training and testing sets. Then a DNN model is built, hyperparameters initialized, and the processed data input into the DNN model. Finally, based on the loss value, the weights, biases, and other parameters of each convolutional layer in the network are updated and optimized to achieve the optimal model, thereby obtaining the bearing health index construction model.

3.3. An improved bearing life prediction model based on DNN-KNN

Although the bearing life prediction model for electromechanical equipment has been built on the basis of DNN, the DNN cannot accurately describe all structural details of the complex mapping function, which leads to a certain deviation of the prediction output from the real data and is prone to local optimal solutions. And the KNN is a local model, which is good at dealing with nonlinear decision boundary and multi-classification problems (Althubaiti *et al.*, 2022). Therefore, it is studied to combine the DNN and KNN into a DNN-KNN model to correct the deviation value in prediction and raise the effectiveness of the model. The bearing life prediction process based on the DNN-KNN model is shown in Fig. 4.

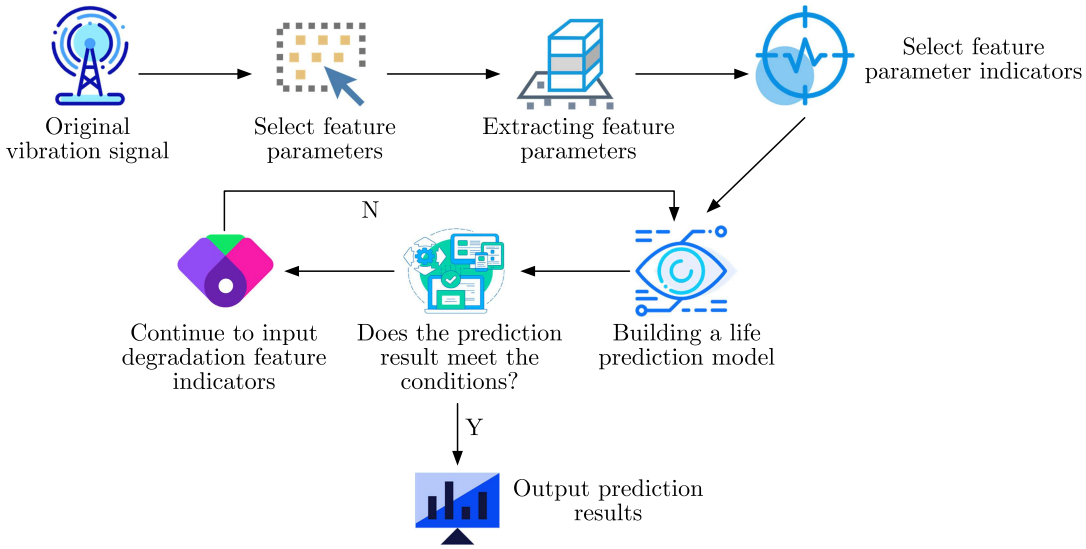


Fig. 4. Bearing life prediction based on DNN-KNN model

For bearing life prediction, the original vibration signals need to be input, and the feature parameters of the signals are extracted from them, which include frequency-domain features, time-domain features, and time-frequency-domain features. For any bearing, it is difficult for a single DNN model to describe the relationship between all feature parameters and bearing life. Therefore, the study first clusters multiple DNN models expressioned in a vector form

$$\mathbf{DNN}_m = \begin{bmatrix} DNN_1 \\ DNN_2 \\ \vdots \\ DNN_m \end{bmatrix} \tag{3.9}$$

In Eq. (3.9), $\mathbf{DNN}_m$ represents the set of m DNN models. Multiplying the input feature vectors with the clustering results of multiple DNN models, ome is able to obtain a set of vectors of the output predicted values, which is calculated as

$$\mathbf{x}_i \times \mathbf{DNN}_m = \begin{bmatrix} P_1^i \\ \vdots \\ P_j^i \\ \vdots \\ P_m^i \end{bmatrix} \tag{3.10}$$

In Eq. (3.10), P_M^i represents the first i m predicted value. At this point, the average of predicted values is generally calculated as the final result, and the calculation is denoted by

$$\overline{P} = \frac{1}{m} \sum_{j=1,m} P_j^i \tag{3.11}$$

In Eq. (3.11), $\overline{P}$ denotes the average prediction value. But the results calculated in this way ignore the prediction bias of each DNN model. Therefore, the study introduces a KNN algorithm by which the test data are predicted, which usually predicts the values of new input samples by analyzing the training samples. Specifically, the KNN algorithm selects a number of training samples to generate the nearest-neighbor dataset, and the KNN algorithm clusters the scatter points as shown in Fig. 5.

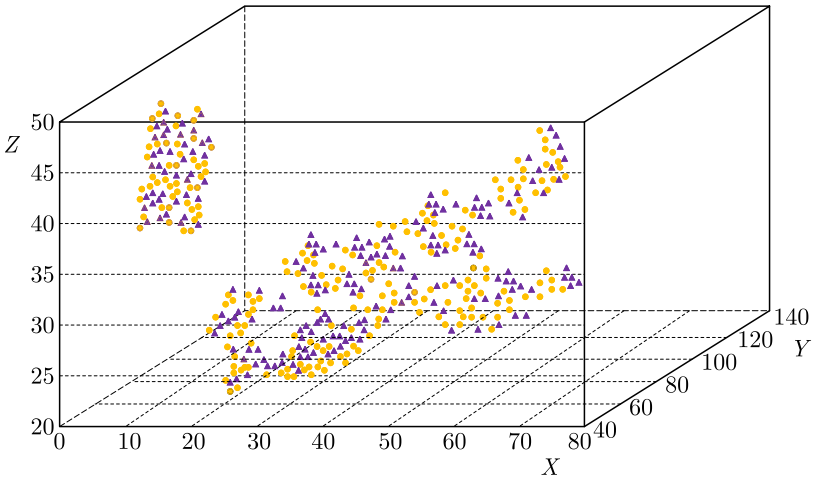


Fig. 5. KNN algorithm clustering a scatter plot

In Fig. 5, the KNN algorithm is able to find the nearest neighbors by calculating the distance between sample points and using the information from these neighbors to classify or regressively predict new data points. The algorithm has advantages of simplicity, being easy to understand and implement, and is suitable for small datasets and sample imbalance. The training dataset is first set to be $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where each sample contains an input feature vector $\mathbf{x}$ and a corresponding output prediction $\mathbf{y}$. The set expression is shown as

$$Y_K = \{y_i : \quad i = 1, 2, \ldots, k\} \tag{3.12}$$

In Eq. (3.12), Y_K denotes the nearest-neighbor dataset, and i expresses the i-th sample. In the next step, the predicted values of test samples can be calculated from these similar training samples with the fokkowing expression

$$Y_{PV} = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i} \qquad i = 1, 2, \ldots, k \tag{3.13}$$

In Eq. (3.13), Y_{PV} means the test value of the test sample, and w_i indicates the weights of the approximation sample. These weights can be determined based on the similarity with the

test sample, and the inverse of the distance is usually used as the weight. The next step is to calculate the input prediction value from the test value of the test sample, which is worked out and expressed by

$$P(x) = \frac{1}{k} \sum_{j \in N_k(x)} y(x_j) \tag{3.14}$$

In Eq. (3.14), $P(x)$ denotes the predicted value of the input feature vector $\mathbf{x}$, $N_k(x)$ denotes the set of k data similar to the feature vector $\mathbf{x}$, and $y(x_j)$ denotes the true value corresponding to the data x_j. The next step is to calculate the similarity between the predicted values of the test and training samples by using Pearson's correlation coefficient to correct the bias between the two, and the expression of Pearson's correlation coefficient is shown as

$$R(\alpha, \beta) = \frac{\sum_i (\alpha_i - \overline{\alpha})(\beta_i - \overline{\beta})}{\sqrt{\sum_i (\alpha_i - \overline{\alpha})^2} \sqrt{\sum_i (\beta_i - \overline{\beta})^2}} \tag{3.15}$$

In Eq. (3.15), R denotes the Pearson correlation coefficient, and α and β denote the two eigenvectors. The correction for bias is calculated from the equation

$$P'(x) = P(x) + \frac{1}{k} \sum_{j \in N_k(x_i)} [y(x_j) - P(x)] \tag{3.16}$$

In Eq. (2.16), $P'(x)$ denotes the real value, and $y(x_j) - P(x)$ denotes the deviation of the predicted value. At this point, the improvement of bearing life prediction on the basis of DNN is completed.

4. Analysis of the results of the DNN-KNN-based bearing life prediction model for electromechanical equipment

This Section focuses on the research findings of the designed fusion bearing life prediction model, with the first section analyzing the performance of the designed model and the second section analyzing the effectiveness of the designed model in practical applications.

4.1. Performance analysis of the bearing life prediction model based on DNN-KNN

To verify the performance of the designed DNN-KNN-based bearing life prediction model for electromechanical equipment, the study sets the size of convolution kernel to 5×5, sets the learning rate to 0.0001, and the iteration times to 200. At the same time, let the number of hidden layers in the network be 8, with the number of neurons in each layer being 300, 2001, 150, 100, 80, 50, 30, 1, and the last layer being the output layer. Since each input data outputs a predicted lifespan value, the output layer contains 1 neuron. Firstly, the accuracy and loss of the designed model are calculated and compared with the DNN, Support Vector Machines, and Decision Tree algorithms. The results are shown in Fig. 6.

From Fig. 6a, the accuracy of all four algorithms tended to increase as the amount of iterations increased. Among them, the accuracy of the designed DNN-KNN algorithm is 0.96 when it tends to level off, the accuracy of the DNN algorithm is 0.90 when it tends to stabilize, the accuracy of the SVM algorithm is 0.87 when it reaches stability, and the accuracy of the DT algorithm is 0.81 when it tends to stabilize. From Fig. 6b, the loss value of the four algorithms had a tendency to decrease gradually, and the four algorithms reached the maximum amount of iterations. The loss values when the four algorithms reached the max amount of iterations are

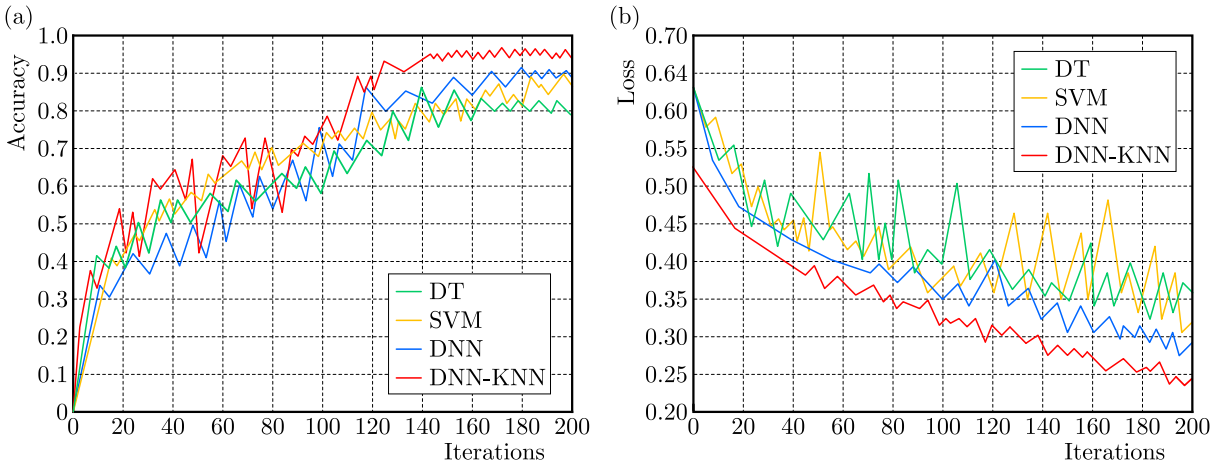


Fig. 6. Accuracy and loss of different models

0.23, 0.29, 0.32, and 0.35, respectively. The above outcomes denoted that the designed DNN-KNN algorithm has a high prediction accuracy and good convergence performance. In the next step, the acquired dataset is divided into six types of test sets, and the errors of different models are calculated separately. The outcomes are expressed in Table 1.

Table 1. Errors of different models

	DNN-KNN	DNN	SVM	DT
Test set 1	14.16	12.56	27.51	17.55
Test set 2	21.69	29.54	22.32	23.47
Test set 3	16.33	20.21	18.18	20.06
Test set 4	0.93	10.43	6.09	15.49
Test set 5	24.25	10.59	18.02	26.87
Test set 6	14.62	21.12	26.13	28.32
Average error	15.33	17.41	19.71	21.96

From Table 1, the average error of the designed DNN-KNN based life prediction model is 15.33%, the average error of the DNN-based model is 17.41%, and the average errors of the two prediction models, SVM and DT, are 19.71% and 21.96%, respectively. The error of the research-designed life prediction model based on the DNN-KNN is significantly lower than that of the other algorithms. The findings further demonstrated the high prediction accuracy of the designed model and also prove its reliability. Finally, the recall and $F1$ value are introduced to assess the comprehensive effectiveness of the designed algorithm and compared with the other four algorithms. The results are shown in Fig. 7.

In Fig. 7a, the recall of all four models tends to increase and level off as the iteration times increased. Among them, the recall of the designed DNN-KNN model is 97% when it reaches plateau, and the recall of the three models, DNN, SVM, and DT, are 93%, 88%, and 85% when they plateau, respectively. From Fig. 7b, the $F1$-score of all four models gradually increased as the iteration times increased. When the maximum amount of iterations was reached, the $F1$-score of DNN-KNN, DNN, SVM, and DT were 0.93, 0.80, 0.78, and 0.72, respectively. It can be found that the recall and $F1$-score of the designed DNN-KNN model are significantly higher than those of the other models, indicating that it possesses a better comprehensive performance, and meanwhile, proving that it has a good generalization ability.

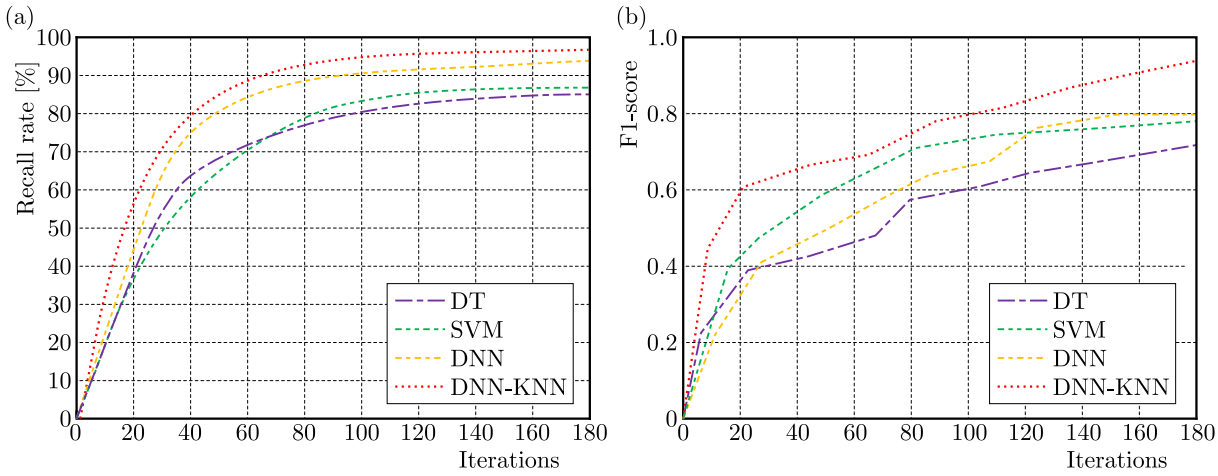


Fig. 7. Recall rates and $F1$-score of different algorithms

4.2. Analysis of the effect of practical application of the life prediction model based on DNN-KNN

To assess the performance of the designed DNN-KNN-based life prediction model for bearings in electromechanical equipment in practical applications, the study firstly selects the test bearings, and then calculates the residual life of the bearings using the DNN-KNN-based and DNN-based life prediction model, respectively. The scatter plots of the predicted life are shown in Fig. 8.

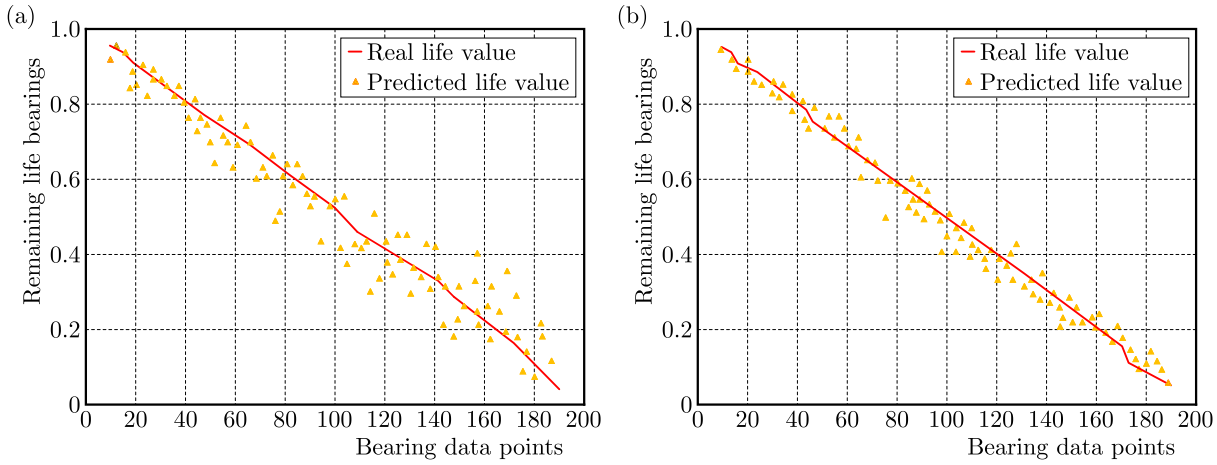


Fig. 8. Scatter plots of the residual life of test bearings based on different algorithms

In Fig. 8, among the two bearing residual life prediction models, the remaining life prediction value of the DNN-KNN-based model is obviously closer to the real life value, indicating that the prediction accuracy of the designed model is higher. In the next step of the study, six test bearings are selected, and the feature indexes are selected by different methods. The curves of the average value of degradation feature indexes obtained by each algorithm are compared with the cycle, and the findings are indicated in Fig. 9.

In Fig. 9, the values of degradation feature indicators extracted by both algorithms are between 0 and 1, which is conducive to determining the range of the indicator failure threshold. However, the curve of the mean feature data extracted by the DNN-KNN-based residual life prediction model is obviously smoother and less fluctuating, which proves that it has better stability and reliability in extracting feature indicators. This implies that the DNN-KNN-based

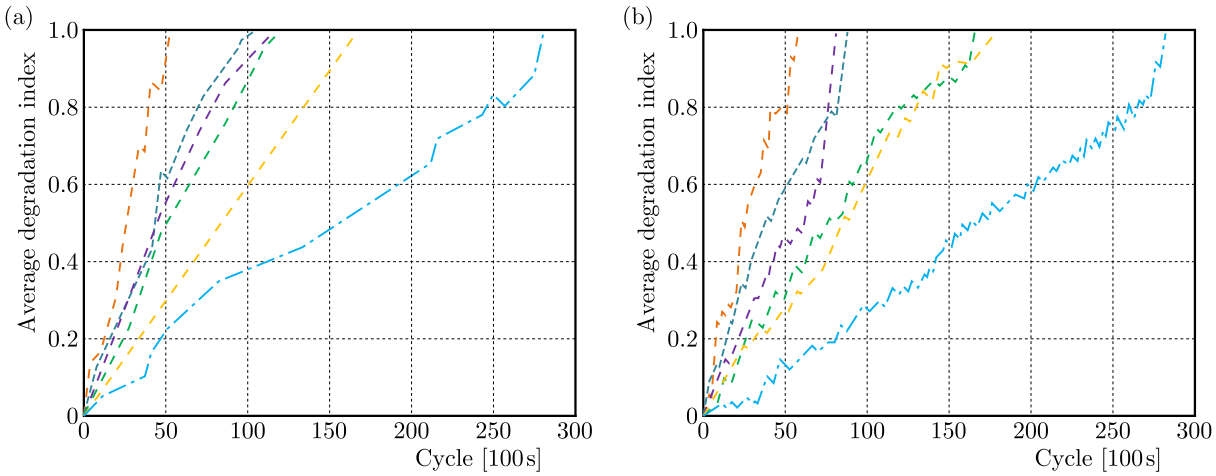


Fig. 9. The curve of the average value of degradation characteristic indicators changing with the cycle

model can more accurately capture the bearing degradation trend and provides more reliable life prediction results. The next step of the study introduces robustness and correlation to further evidence the effectiveness of the designed remaining life prediction model, and, at the same time, compares the results with those of the DNN-based residual life prediction model, which are given in Table 2.

Table 2. Comparison of robustness and correlation between different models

Testing bearings	DNN-KNN		DNN	
	Robustness	Correlation	Robustness	Correlation
Bearing 1	0.98	0.98	0.97	0.95
Bearing 2	0.98	0.99	0.98	0.96
Bearing 3	0.99	0.98	0.96	0.97
Bearing 4	0.98	0.98	0.98	0.98
Bearing 5	0.99	0.99	0.97	0.90
Bearing 6	0.99	0.99	0.92	0.94

From Table 2, among the six test bearings, the average robustness and average correlation of the DNN-KNN-based remaining life prediction model are 0.985 and 0.985, respectively, and the mean values of the robustness and correlation of the DNN-based remaining life prediction model are 0.963 and 0.95, respectively. The above findings indicated that the DNN-KNN-based remaining life prediction model performs better in terms of robustness and correlation. Finally, the study introduces the mean absolute error (MAE), mean square error (MSE), and RMS error to evaluate the prediction results, and compares the results with those of the DNN-based prediction model and the SVM-based prediction model. The results are shown in Fig. 10.

In Fig. 10, the MAE, MSE and RMS error of the SVM-based remaining life prediction model are 0.036, 0.071 and 0.039, respectively, and the values of the three indicators of the DNN-based remaining life prediction model are 0.024, 0.073 and 0.043, respectively, whereas the values of the three indicators of the designed DNN-KNN-based remaining life prediction model are 0.019, 0.066 and 0.025, respectively. It can be found that the three indicators of the designed model are significantly lower than the other models, which further proves that the prediction accuracy of the designed model is higher. The indicator values are 0.019, 0.066, and 0.025, respectively. To further verify the superiority of the DNN-KNN based electromechanical equipment bearing life prediction model, the average absolute percentage error and maximum absolute error were introduced to calculate the two indicator values of the design method and compared with the

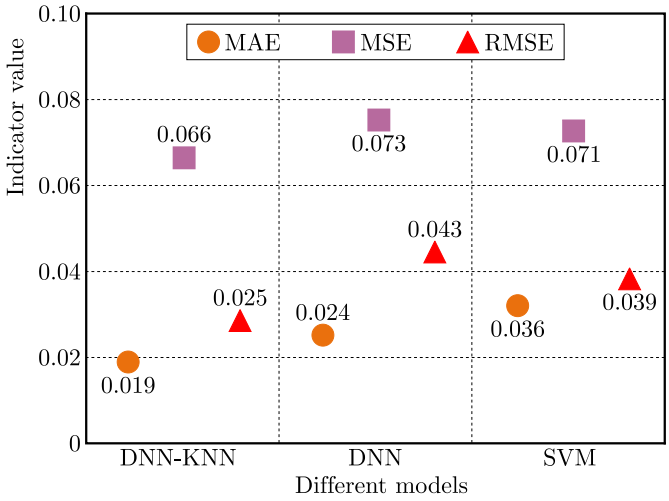


Fig. 10. Indicator values for different models

two indicator values of the latest methods in (Nistane, 2024; Li and Wang, 2024; Fei *et al.*, 2024; Yang *et al.*, 2024). The results are shown in Table 3.

Table 3. Comparison of average absolute percentage error and maximum absolute error of five methods

Model	Maximum absolute error	Maximum absolute percent error
Nistane (2024)	0.0967	0.1223
Li and Wang (2024)	0.1046	0.0955
Fei *et al.* (2024)	0.0953	0.0837
Yang *et al.* (2024)	0.0988	0.0846
DNN-KNN	0.0920	0.0732

From Table 3, it can be seen that the maximum absolute error and average absolute percentage error of the designed DNN-KNN model are 0.0920 and 0.0732, respectively. Their maximum absolute errors are 0.0047, 0.0079, 0.0033, and 0.0068 lower than those in (Nistane, 2024; Li and Wang, 2024; Fei *et al.*, 2024; Yang *et al.*, 2024), respectively. Its average absolute percentage errors are 0.0491, 0.0223, 0.0105, and 0.0114 lower than the other four methods, respectively. It can be found that the prediction error of DNN-KNN is smaller than in other methods, which proves that it can effectively improve the accuracy of predicting the remaining life of bearings.

5. Conclusion

In the field of engineering, accurate remaining life prediction of bearings in electromechanical equipment is of crucial significance for the reliability analysis and maintenance of the equipment. The traditional prediction methods are often difficult to meet the accuracy and real-time demands in the complex engineering environment. Therefore, the study firstly carries out selection of indicators of bearing degradation, then introduces DNN, constructs a remaining life prediction model based on these indicators, and finally introduces KNN to correct the deviation of the DNN model, and designs a DNN-KNN residual life prediction model. The results show that in the accuracy and loss calculation, the accuracy of the four algorithms when reaching the maximum number of iterations is 0.96, 0.90, 0.97, and 0.81, respectively, which means that the prediction accuracy of the designed algorithms is high. In the error calculation of different

models, the average errors of the four models are 15.33%, 17.41%, 19.71% and 21.96%, respectively, which further proves that the prediction accuracy of the proposed models is high and demonstrates their reliability. In the recall and $F1$ value calculations of different models, the designed DNN-KNN-based lifetime prediction model has a recall and $F1$ value of 97% and 0.93, respectively, which are significantly higher than in the other models, proving that it has a better overall performance, and also proving that it has a better generalization ability. The above findings prove the effectiveness of the designed life prediction model based on DNN-KNN, but the study still has some shortcomings when making predictions. A large number of historical data of bearings are used, which may have a certain impact in terms of computational efficiency, and the subsequent will continue to be improved in this aspect.

Acknowledgments

The research was funded by the Scientific and Technology Development Project of Jilin Province, China (Grant No. 20240304183SF).

References

1. Abdou M.A., 2022, Literature review: efficient deep neural networks techniques for medical image analysis, *Neural Computing and Applications*, **34**, 8, 5791-5812
2. Althubaiti A., Elasha F., Teixeira J.A., 2022, Fault diagnosis and health management of bearings in rotating equipment based on vibration analysis – a review, *Journal of Vibroengineering*, **24**, 1, 46-74
3. Bhosle K., Musande V., 2023, Evaluation of deep learning CNN model for recognition of Devanagari digit, *Artificial Intelligence and Applications*, **1**, 2, 114-118
4. Cao H., Wu Y., Bao Y., Feng X., Wan S., Qian C., 2023, UTrans-net: A model for short-term precipitation prediction, *Artificial Intelligence and Applications*, **1**, 2, 106-113
5. Ding P., Jia M., Wang H., 2021, A dynamic structure-adaptive symbolic approach for slewing bearings' life prediction under variable working conditions, *Structural Health Monitoring*, **20**, 1, 273-302
6. Fei C.W., Han Y.J., Wen J.R., Li C., Han L., Choy Y.S., 2024, Deep learning-based modeling method for probabilistic LCF life prediction of turbine blisk, *Propulsion and Power Research*, **13**, 1, 12-25
7. Hatem M.Q., 2022, Skin lesion classification system using a K-nearest neighbor algorithm, *Visual Computing for Industry, Biomedicine, and Art*, **5**, 7
8. Li H., Wang C., 2024, Combining first prediction time identification and time-series feature window for remaining useful life prediction of rolling bearings with limited data, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **238**, 2, 274-290
9. Li X., Xu Y., Li N., Yang B., Lei Y., 2023, Remaining useful life prediction with partial sensor malfunctions using deep adversarial networks, *IEEE/CAA Journal of Automatica Sinica*, **10**, 1, 121-134
10. Mao W., He J., Zuo M.J., 2019, Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning, *IEEE Transactions on Instrumentation and Measurement*, **69**, 4, 1594-1608
11. Nistane V., 2024, Optimum prediction model of remaining useful life for rolling element bearing based on integrating optimize health indicator (OHI) and machine learning algorithm, *World Journal of Engineering*, **21**, 1, 170-185
12. Rezamand M., Kordestani M., Carriveau R., Ting D.S.-K., Saif M., 2020, An integrated feature-based failure prognosis method for wind turbine bearings, *IEEE/ASME Transactions on Mechatronics*, **25**, 3, 1468-1478

13. Rezamand M., Kordestani M., Orchard M.E., Carriveau R., Ting D.S.-K., Saif M., 2021, Improved remaining useful life estimation of wind turbine drivetrain bearings under varying operating conditions, *IEEE Transactions on Industrial Informatics*, **17**, 3, 1742-1752

14. Sun S., Wen Z., Du T., Wang J., Tang Y., Gao H., 2021, Remaining life prediction of conventional low-voltage circuit breaker contact system based on effective vibration signal segment detection and MCCAE-LSTM, *IEEE Sensors Journal*, **21**, 19, 21862-21871

15. Wang W., Lei Y., Yan T., Li N., Nandi A., 2022, Residual convolution long short-term memory network for machines remaining useful life prediction and uncertainty quantification, *Journal of Dynamics, Monitoring and Diagnostics*, **1**, 1, 2-8

16. Wu T., Yao Y., Li Z., Chen B., Wu Y., Sun W., 2024, Remaining useful life prediction of circuit breaker operating mechanisms based on wavelet-enhanced dual-tree residual networks, *Journal of Power Electronics*, **24**, 1, 78-91

17. Yang C., Liu J., Zhou K., Li X., 2024, Dynamic spatial-temporal graph-driven machine remaining useful life prediction method using graph data augmentation, *Journal of Intelligent Manufacturing*, **35**, 1, 355-366

18. Zhang Y., Feng K., Ji J.C., Yu K., Ren Z., Liu Z., 2023, Dynamic model-assisted bearing remaining useful life prediction using the cross-domain transformer network, *IEEE/ASME Transactions on Mechatronics*, **28**, 2, 1070-1080

Manuscript received December 22, 2023; accepted for publication October 7, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 737-749, Warsaw 2024
https://doi.org/10.15632/jtam-pl/193391

DYNAMIC CHARACTERISTICS OF ROTOR-SFD SYSTEM WITH INERTIAL EFFECT OF SFD

Zhongyu Yang
Key Laboratory of Light Duty Gas Turbine, Institute of Engineering Thermophysics, Chinese Academy of Sciences, and School of Aeronautics and Astronautics, University of Chinese Academy of Sciences, and National Key Laboratory of Science and Technology on Advanced Light-duty Gas-turbine, Beijing, China

Jiaqi Li
Beijing Aerospace Propulsion Institute, Beijing, China

Jiali Chen
Key Laboratory of Light Duty Gas Turbine, Institute of Engineering Thermophysics, Chinese Academy of Sciences, and School of Aeronautics and Astronautics, University of Chinese Academy of Sciences, and National Key Laboratory of Science and Technology on Advanced Light-duty Gas-turbine, Beijing, China

Yinli Feng
Key Laboratory of Light Duty Gas Turbine, Institute of Engineering Thermophysics, Chinese Academy of Sciences, and National Key Laboratory of Science and Technology on Advanced Light-duty Gas-turbine, Beijing, China, and School of Engineering Sciences, University of Chinese Academy of Sciences
corresponding author Yinli Feng, e-mail: fengyl889@163.com

Squeeze film dampers (SFDs) are damping devices that have been widely used in rotating machines. SFDs can effectively suppress rotor vibration and reduce transmitted forces. This study established a dynamic characteristic model of the SFD considering inertia force and developed the finite element model of the rotor-SFD system by combining fluid mechanics and rotor dynamics. Furthermore, the influence of SFD structural parameters on cavitation, pressure distribution, stiffness, and damping coefficient of the SFD was investigated. The impact of the SFD clearance on the transient dynamic characteristics of the rotor-SFD system was also analyzed.

Keywords: squeeze film damper, inertial effect, rotordynamic

1. Introduction

Squeeze film dampers (SFDs) are widely utilized in aero-engines, gas turbines, steam turbines, etc., owing to their advantages of low cost, simple structure, and outstanding vibration attenuation effect. However, given the inherently nonlinear behavior of an SFD, a poorly designed damper could exacerbate rotor vibrations, significantly influencing the engine operation.

The SFD dynamic characteristics have been extensively investigated via both theoretical and experimental approaches. Gehannin *et al.* (2016) proposed an air ingestion model for SFDs based on the Rayleigh–Plesset equation and investigated the influence of vapor cavitation and air ingestion on the SFD damping characteristics. San Andrés *et al.* (1992) presented an innovative approach for analyzing the dynamic force response of an SFD considering the dynamic flow interaction between the squeeze film lands and the feeding groove. Their model could calculate the corrected values for damping and inertia force coefficients, highlighting the significance of groove volume and the fluid compressibility effect on the dynamic force response of SFDs. Moreover, Zhou *et al.* (2020) developed a three-dimensional model of an SFD considering air

ingestion employing the transient step-size solution method and dynamic mesh technology. They performed model validation through bidirectional excitation experiments. Additionally, they investigated the effects of whirl frequency and eccentricity on air ingestion. Gheller *et al.* (2022) developed a model for SFDs considering both cavitation and inertial forces. Using the finite difference method, they discretized the Rayleigh equation, including the effect of inertial forces, and studied the influence of different boundary conditions on the SFD damping characteristics. Fan *et al.* (2017) developed a cavitation model for SFDs using the Elrod algorithm and Gumbel's cavitation boundary condition, and performed comparisons with commonly used cavitation models. They investigated the influence of fluid inertia effects on the SFD cavitation behavior. Hamzehlouia and Behdinan (2019) used the momentum approximation and perturbation methods to analytically determine the pressure distribution, velocity distribution, and oil film force of SFDs considering lubricant inertia. They developed and experimentally validated a model for SFDs considering lubricant inertia. Drapatow *et al.* (2021) discretized the Navier-Stokes equations considering the local and convective inertia terms using the finite volume method, and they solved for the pressure and oil film reaction of SFDs. Moreover, they investigated the influence of inertial forces on SFD dynamic characteristics under high-temperature and high-speed conditions. Zhou *et al.* (2023) developed a computational fluid dynamics model for a sealed end SFD considering the effects of fluid inertia. They investigated the influence of eccentricity and whirl frequency on stiffness and damping coefficients of an SFD. San Andrés (2014) experimentally determined the dynamic forced response of an SFD with a simplified configuration, where the feedholes were completely eliminated. The measurement results exposed the restricted suitability of classical formulas and distinctly underscored the necessity of employing a more advanced model to precisely forecast the dynamic forced response of practical SFDs.

Chen *et al.* (2017) investigated the relationship between rigid body translation and precession for a rigid rotor-SFD system with nonsymmetric stiffness supports. They considered two cases: precession motion non-resonance and internal resonance when the translation motion was primarily resonant. To establish an optimization model for SFDs, Gupta *et al.* (2023) combined the response surface methodology and analysis of variance. The model considered the influence of journal precession, pressure of oil inlet, and oil-gas mixture ratio on the damping effect of the damper. They optimized SFDs for high-speed rotor systems. Younan *et al.* (2010) investigated the relationship between viscosity and temperature of the lubricating oil, corrected density and viscosity of the lubricating oil, and presented the effect of air pockets. El-Saeidy and Sticher (2010) derived the equation of motion of a system considering nonlinear supports based on the Lagrange equation and investigated dynamic characteristics of elastic support-rotor systems under unbalanced forces. Zapomél *et al.* (2019) proposed a control equation for oil film pressure distribution, solved the equation of motion of a rotor based on the Runge-Kutta method, and revealed nonlinear characteristics stemming from damping. El-Shafei (2002) established a steady-state precession model of the SFD and calculated the pressure distribution of the SFD without seals on both ends.

The traditional SFD models have neglected the influence of fluid inertia forces by assuming that the fluid inertia is completely dominated by viscous forces ($\mathrm{Re} = 0$), and the Reynolds equation has been used to represent the SFD pressure distribution. However, in applications involving high-speed turbomachinery and low-viscosity lubricants, the extrusion Reynolds number in SFDs can exceed one, and thus, the influence of lubricant inertia cannot be neglected (Hamzehlouia and Behdinan, 2019). Although the oil supply system significantly impacts the damping characteristics of SFDs, research on it has been relatively limited. To date, direct research on the influence of SFD clearance, oil inlet parameters, and SFD width on dynamic characteristics of rotors has been lacking. Indeed, most of the aforementioned studies have focused on individual research on SFDs. Hence, this study establishes a numerical model of SFD considering both inertial forces and cavitation. The influence of SFD clearance on cavitation

and damping characteristics are investigated. Moreover, a rotor system model considering SFD is established, and the influence of SFD clearance on the rotor-SFD system is analyzed.

2. Establishment and validation of the dynamic model

2.1. Establishment and validation of the numerical model for a SFD considering inertia

For a rotor system with a SFD, the damping properties of the oil film are crucial. Bearing 1 utilizes a combination of an elastic supporting structure comprising an SFD, a squirrel cage, and a rolling bearing, while bearing 2 adopts a rigid supporting structure with a ball bearing. Figure 1 displays a schematic of the SFD.

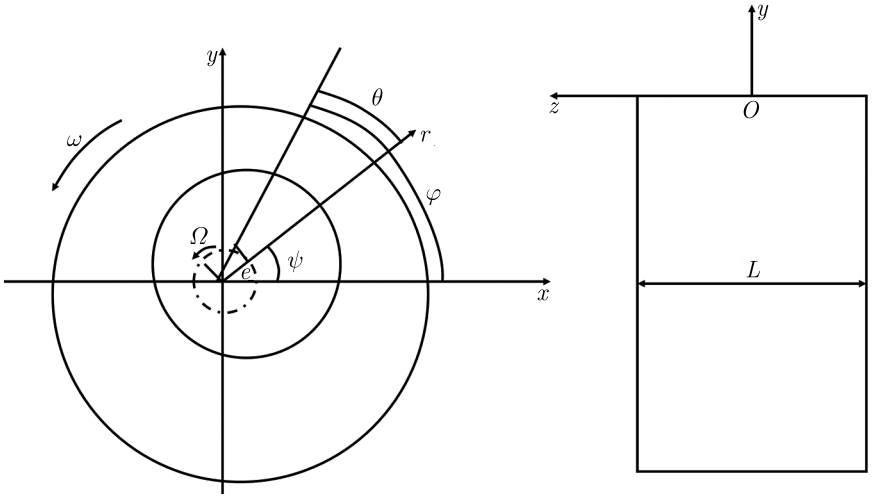


Fig. 1. Schematic of the SFD

To consider the oil film inertia, the axial pressure distribution equation is obtained based on the energy approximations by Crandall and El-Shafei (1991) as follows

$$\begin{aligned} &\frac{\partial p}{\partial z} = z\Big[2\frac{\mu}{h^3}\frac{\partial h}{\partial t} + \frac{6}{5}\frac{\rho}{h}\frac{\partial^2}{\partial t^2} - \frac{102}{35}\frac{\rho}{h^2}\Big(\frac{\partial h}{\partial t}\Big)^2\Big] \\ &p = p_{atm} = 0 \qquad\qquad z = \pm\frac{L}{2} \end{aligned} \tag{2.1}$$

where p is the oil film pressure, ρ is the lubricant density, h is the oil film thickness, μ is the lubricant viscosity, p_{atm} is the atmospheric pressure, and $h = c - e\cos\theta$. Additionally, e is the journal eccentricity, and θ is measured from the positive r-axis of the precessing coordinate system.

The short bearing theory hypothesis is available when no hermetic seals are present at both the ends, and the pressure at both the ends is similar to the external pressure. Due to the squeezing effect, the pressure in the central region of the damper is considerably higher than that at the ends, while the tangential pressure gradient is low

$$\frac{\partial p}{\partial z} \gg \frac{\partial p}{\partial \theta} \qquad p = \Big(\frac{z^2}{2} - \frac{L^2}{4}\Big)\Big[2\frac{\mu}{h^3}\frac{\partial h}{\partial t} + \frac{6}{5}\frac{\rho}{h}\Big(\frac{\partial^2 h}{\partial t^2}\Big) - \frac{102}{35}\frac{\rho}{h^2}\Big(\frac{\partial h}{\partial t}\Big)^2\Big] \tag{2.2}$$

By integrating equation $(2.2)_2$ over the circumferential and axial directions, the radial force F_r and tangential force F_t of the oil film are obtained

$$F_r = -\int_{L/2}^{L/2}\int_{\theta_1}^{\theta_2}(p\cos\theta)R\,d\theta\,dz \qquad F_t = -\int_{L/2}^{L/2}\int_{\theta_1}^{\theta_2}(p\sin\theta)R\,d\theta\,dz \tag{2.3}$$

where R is the SFD housing radius.

The expressions for the equivalent stiffness and equivalent damping coefficients of the SFD are as follows

$$K = -\frac{F_r}{e} \qquad C = -\frac{F_t}{e\Omega} \tag{2.4}$$

The above equations only exemplify the relationship between SFD damping characteristics and the oil film forces. To characterize the influence of inertia forces on vibration reduction characteristics of an SFD, the following dimensionless parameters are given (El-Shafei, 1991)

$$\begin{aligned} f_r &= -C_{rr}\varepsilon' - C_{rt}\varepsilon\psi' - Re\Big(M_{rrad}\varepsilon'' + M_{rnon}\frac{\varepsilon'^2}{\varepsilon} + M_{rtan}\varepsilon\psi'' - M_{rcen}\varepsilon\psi'^2\Big) \\ f_t &= -C_{tr}\varepsilon' - C_{tt}\varepsilon\psi' - Re\Big(M_{trad}\varepsilon'' + M_{tnon}\frac{\varepsilon'^2}{\varepsilon} + M_{ttan}\varepsilon\psi'' - M_{tcen}\varepsilon\psi'^2\Big) \end{aligned} \tag{2.5}$$

where C_{ij} and M_{ij} are the stiffness and inertia coefficients of the oil film force, respectively. $(\cdot') = d/d\tau$, $\tau = \omega t$, and $\varepsilon = e/C$.

The radial oil film force F_r and tangential oil film force F_t are as follows (El-Shafei, 1991)

$$F_r = \frac{\mu\Omega RL^3}{c^2} f_r \qquad F_t = \frac{\mu\Omega RL^3}{c^2} f_t \tag{2.6}$$

Based on the reference (El-Shafei, 1995), the expressions for C_{ij} and M_{ij} are presented in Tables 1 and 2, respectively.

Table 1. Damping coefficient C_{ij}

Damping coefficient of oil film C_{ij}	
$C_{rr} = \dfrac{\pi(1+2\varepsilon^2)}{2(1-\varepsilon^2)^{5/2}}$	$C_{rt} = \dfrac{2\varepsilon}{(1-\varepsilon^2)^2}$
$C_{tr} = \dfrac{2\varepsilon}{(1-\varepsilon^2)^2}$	$C_{tt} = \dfrac{\pi}{2(1-\varepsilon^2)^{3/2}}$

Table 2. Inertial force coefficient M_{ij}

Inertia coefficient of oil film M_{ij}	
$M_{rrad} = \dfrac{\pi}{10\varepsilon^2}[(1-\varepsilon^2)^{-1/2} - 1]$	$M_{tcen} = -\dfrac{27}{70\varepsilon}\Big[2 + \dfrac{1}{\varepsilon}\ln\Big(\dfrac{1-\varepsilon}{1+\varepsilon}\Big)\Big]$
$M_{trad} = -\dfrac{1}{10\varepsilon}\Big[2 + \dfrac{1}{\varepsilon}\ln\Big(\dfrac{1-\varepsilon}{1+\varepsilon}\Big)\Big]$	$M_{rnon} = \dfrac{17\pi}{70\varepsilon^2}\Big[2 - \dfrac{2-3\varepsilon^2}{(1-3\varepsilon^2)^{3/2}}\Big]$
$M_{rtan} = -\dfrac{1}{10\varepsilon}\Big[2 + \dfrac{1}{\varepsilon}\ln\Big(\dfrac{1-\varepsilon}{1+\varepsilon}\Big)\Big]$	$M_{tnon} = \dfrac{17}{70\varepsilon}\Big[2 + \dfrac{2}{\varepsilon}\ln\Big(\dfrac{1-\varepsilon}{1+\varepsilon}\Big) + \dfrac{2}{1-\varepsilon^2}\Big]$
$M_{ttan} = \dfrac{\pi}{10\varepsilon^2}[1 - (1-\varepsilon^2)^{1/2}]$	$M_{rcor} = -\dfrac{1}{70\varepsilon}\Big[20 + \dfrac{27}{\varepsilon}\ln\Big(\dfrac{1-\varepsilon}{1+\varepsilon}\Big) + \dfrac{31}{1-\varepsilon^2}\Big]$
$M_{ren} = \dfrac{\pi}{70\varepsilon^2}\Big[27 - \dfrac{27-17\varepsilon^2}{(1-\varepsilon^2)^{1/2}}\Big]$	$M_{tcor} = \dfrac{\pi}{10\varepsilon^2}\Big[-\dfrac{27}{7} - (1-\varepsilon^2)^{1/2}) + \dfrac{17}{7}\dfrac{2-\varepsilon^2}{(1-\varepsilon^2)^{1/2}}\Big]$

Next, C_{ij} and M_{ij} are substituted into equations (2.5) and (2.6). Then, the expression for the oil film forces considering the influence of the inertia force is

$$F_r = \frac{\mu\Omega r\varepsilon L^3}{c^2}(-C_{rt} + ReM_{rcen}) \qquad F_t = \frac{\mu\Omega r\varepsilon L^3}{c^2}(C_{tt} + ReM_{tcen}) \tag{2.7}$$

The numerical simulation results are validated based on the experimental data from the reference (Zhang, 2018). Table 3 lists the SFD structural parameters.

Table 3. SFD structural parameters

Diameter R [mm]	Clearance C [mm]	Width L [mm]	Eccentricity e	Frequency Ω Ω
43	0.1	18	0.2	80

Table 4 illustrates the comparison between the numerical simulation and experimental results of the SFD. Since the error is less than 5%, the numerical simulation method proposed in this study is considered to be reliable.

Table 4. Comparison between the numerical simulation and experimental results

Coefficient	Experimental	Numerical simulation	Error
Damping [Ns/m]	844	875	3.7%
Stiffness [N/m]	46000	44000	4.3%

2.2. Establishment and validation of the Finite Element model for the rotor system

A finite element model is established for a rotor with an SFD based on the high-pressure rotor of an aviation engine. In this model, the shaft length is 800 mm, the maximum diameter of the shaft is 88 mm, the disk diameter is 160 mm, and the shaft weight is 33 kg. Figure 2 depicts the specific structure of the model. Bearing 1 is a combined elastic support comprising a ball bearing, a squirrel-cage-type elastic support, and an SFD. The stiffness of the elastic support is $5 \cdot 10^6$ N/m. Bearing 2 is a rigid support constituting a roller bearing, with a stiffness of $1 \cdot 10^8$ N/m. Table 5 presents dimensions of each section of the rotor shaft.

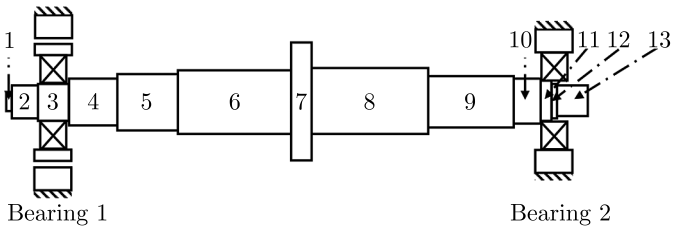


Fig. 2. Specific structure of the rotor-SFD system

Table 5. Dimensions of each section of the rotor shaft [mm]

	No.												
	1	2	3	4	5	6	7	8	9	10	11	12	13
D	25.4	45.5	48	64	76	88	160	86	70	62	57	52	46.5
L	4.5	40	40.5	66	84	158.5	16	170	119	35.5	17	4	45

The Newmark β method is adopted to solve the governing equations of the rotor system (Wang *et al.*, 2022), which can be written as follows

$$M\ddot{\mathbf{q}} + (\mathbf{G} + \mathbf{C})\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{Q} \tag{2.8}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{G}$ is the gyro moment matrix, $\mathbf{C}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix, and $\mathbf{Q}$ is the unbalanced excitation vector.

The damping matrix can be represented using Rayleigh damping

$$\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K} \tag{2.9}$$

where α and β are damping coefficients which are related to the modal damping ratio ξ of the structure.

Subsequently, modal dynamic experiments are conducted on the Jeffcott rotor in mounted states to validate the numerical method. Figure 3 exhibits the experimental setup of the mounted modal dynamic testing of the rotor. Figure 4 presents the nodal division for modal dynamic testing of the rotor. The rotor is divided into 13 nodes: node 7 is the pickup point, while the remaining nodes are excitation points.

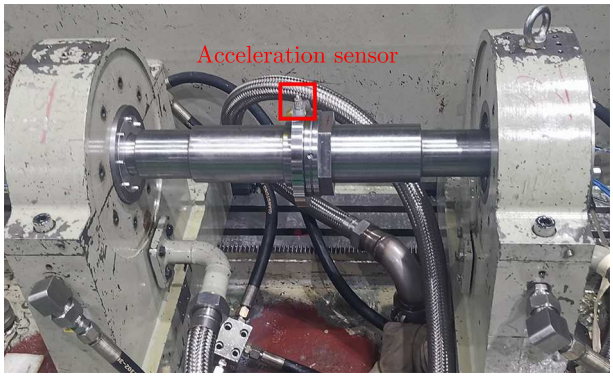


Fig. 3. Experimental setup of the mounted modal dynamic testing

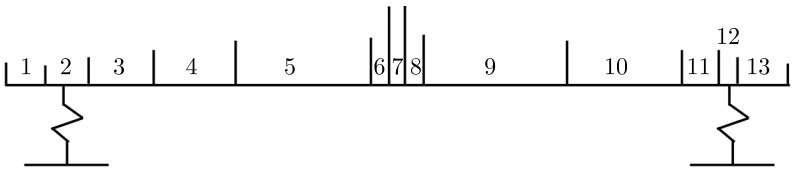


Fig. 4. Nodal division for modal dynamic testing of the Jeffcott rotor

Figure 5 displays the exponential form of transfer function curves for the mounted modal tests. Table 6 illustrates the comparison between the experimental and numerical simulation results for the mounted modal states. Since the mismatch is less than 5%, the numerical simulation results are reliable.

Table 6. Comparison between the experimental and numerical simulation results for the mounted modal state [Hz]

Order	Experiment	Numerical simulation	Error
1st	252	256	1.5%
2nd	1078	1067	1%
3rd	2585	2679	3.6%

Figure 6 depicts the transient experimental setup of the rotor-SFD system. In the system, acceleration sensors are installed at the front and rear bearing positions of the rotor system, while eddy current sensors are placed at the front end, disk, and rear end of the rotor shaft. The system accelerates to 9000 rpm with an acceleration rate of 4.5 rad·s^2.

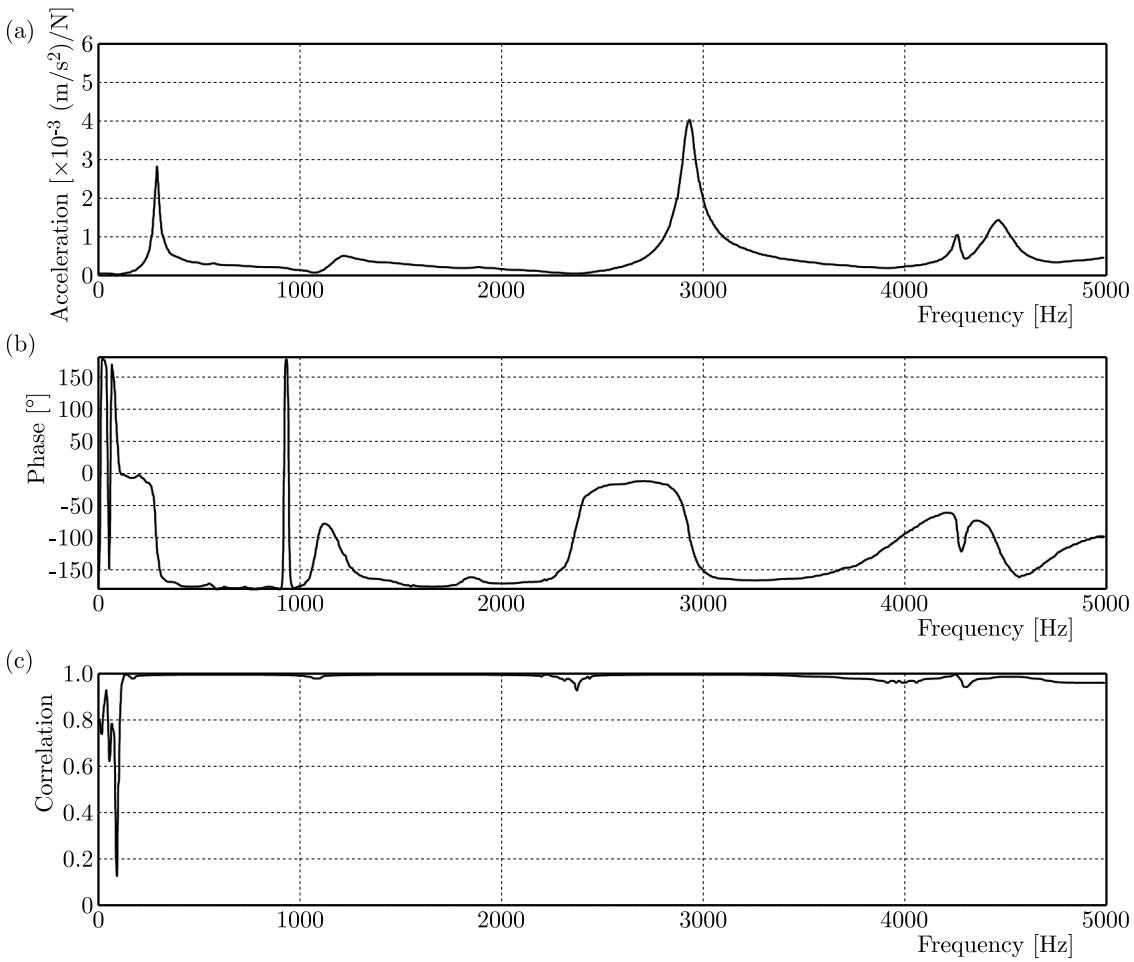


Fig. 5. Exponential form of transfer function curves for the mounted modal test: (a) transfer function curves for the mounted modal tests, (b) phase change, (c) correlation

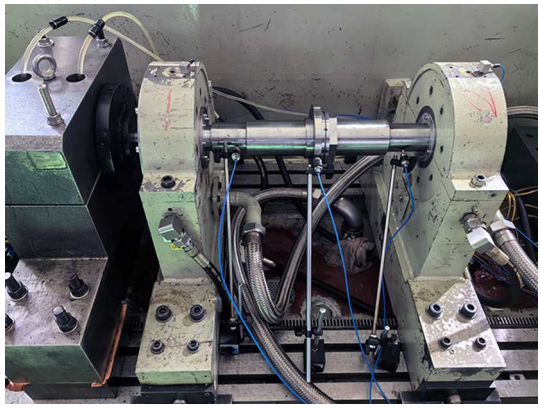


Fig. 6. Transient experimental of the rotor-SFD system

Figure 7 presents the comparison between the experimental and numerical simulation results of the transient acceleration response of the elastic support. The figure shows that compared to the experimental results, the simulation results are smoother with less fluctuations, lower acceleration values, and with an overall error of less than 20%. The simulated results are considered reliable as the experimental environment is unpredictable and complex.

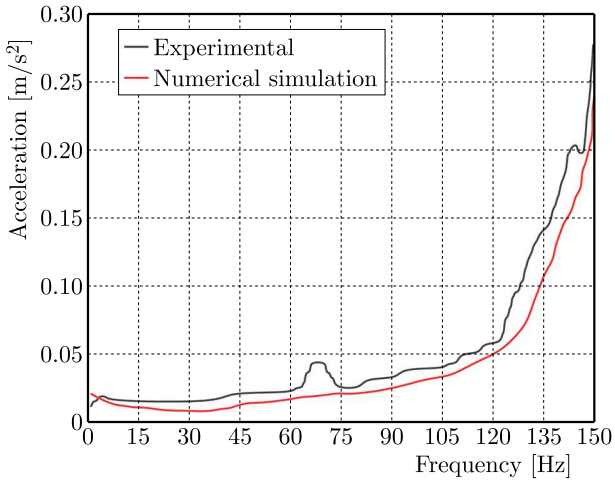


Fig. 7. Transient acceleration response of the rotor system

3. SFD dynamic characteristics

3.1. Cavitation

To investigate the influence of the oil film clearance on cavitation in the SFD, clearances of 0.08, 0.1, 0.12, 0.14, 0.16, 0.18 and 0.2 mm are selected, while the other structural parameters are kept constant.

Figure 8 exhibits the gas phase volume fraction range of the SFDs with different clearances. When the SFD clearance is increased from 0.08 mm to 0.12 mm, both the extent of cavitation and the gas volume fraction decrease. After the SFD clearance exceeds 0.14 mm, the extent of cavitation and the gas volume fraction gradually increase with the clearance. Cavitation includes two phenomena: liquid phase vaporization and air ingestion. Moreover, Figure 9 shows that the pressure distribution amplitude of the SFD decreases with increasing SFD clearance. When the SFD clearance is less than 0.14 mm, the pressure distribution amplitude of the SFD rapidly decreases, leading to a smaller range of liquid saturation vapor pressure and, thereby, to reduction of the extent of cavitation. However, when the SFD clearance exceeds 0.14 mm, the rate at which the pressure distribution amplitude of the SFD decreases slows down, and the occurrence of air ingestion becomes more prevalent with increasing SFD clearance, resulting in an increase of the extent of cavitation.

3.2. Damping characteristics

To investigate the effect of SFD clearance on the damping characteristics, different SFD clearances are selected: 0.08, 0.1, 0.12, 0.14, 0.16, 0.18 and 0.2 mm. The other SFD parameters are kept constant.

Figure 9 depicts the oil film pressure obtained through numerical simulations. In the axial direction, the amplitude of the positive pressure zone in the damper pressure distribution is high in the middle and low at both the ends. The amplitude of the negative pressure zone also exhibits the same pattern. Due to the presence of the oil supply hole, the pressure distribution exhibits fluctuations, and the waveform of the positive pressure zone is not completely smooth. Furthermore, the magnitude of pressure distribution decreases with increasing SFD clearance. When the SFD clearance is less than 0.14 mm, the magnitude of the pressure distribution rapidly decreases; when the SFD clearance exceeds 0.14 mm, the rate of decrease slows down. As the SFD clearance grows, the range of the positive pressure exhibits gradual transitions from 120°-

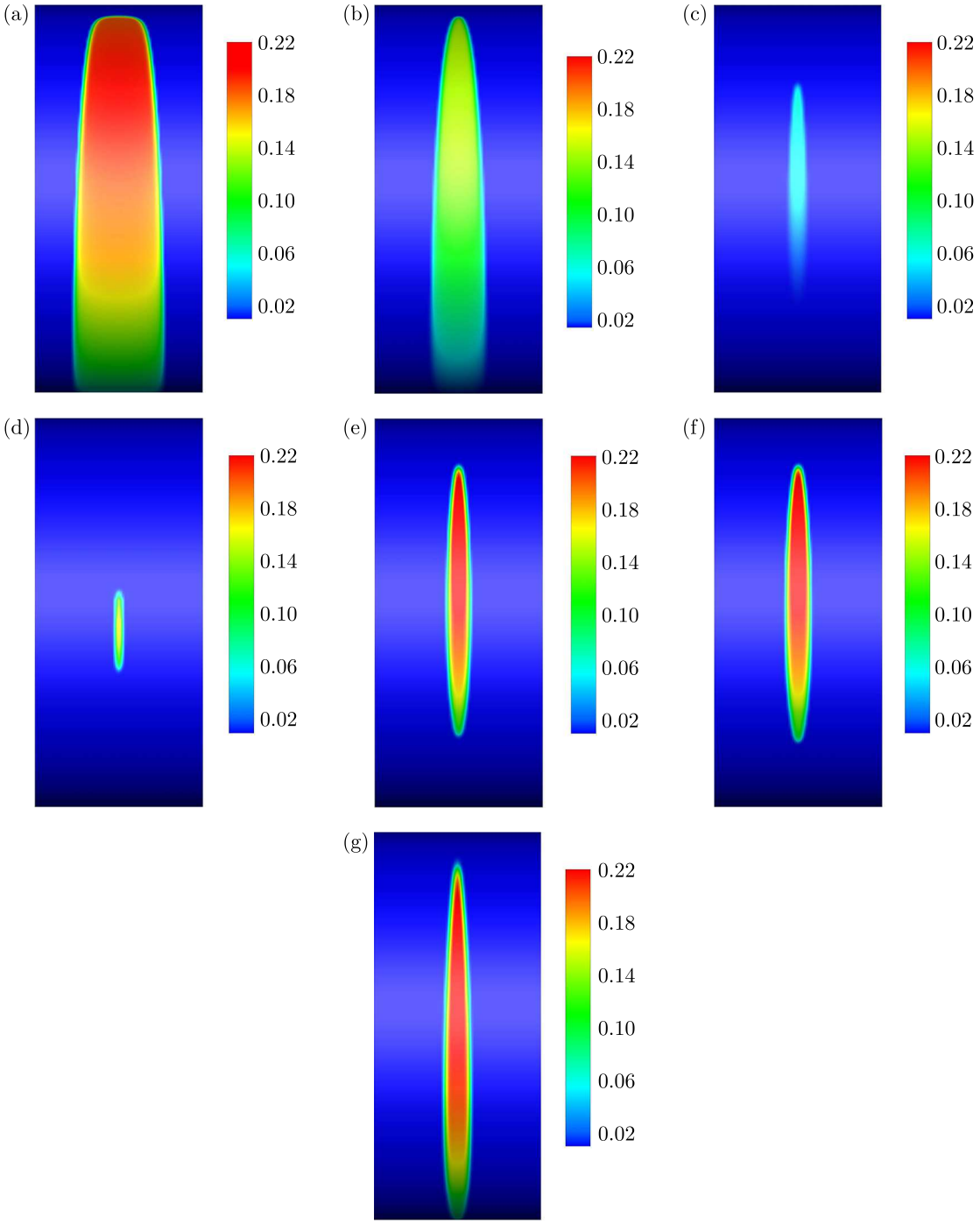


Fig. 8. Gas phase volume fraction range of SFDs with different clearances: (a) $c = 0.08\,\text{mm}$, (b) $c = 0.1\,\text{mm}$, (c) $c = 0.12\,\text{mm}$, (d) $c = 0.14\,\text{mm}$, (e) $c = 0.16\,\text{mm}$, (f) $c = 0.18\,\text{mm}$, (g) $c = 0.2\,\text{mm}$

240° in the circumferential direction to 180°-360°, i.e., and the range of the positive pressure area expands.

Figure 10 presents the SFD stiffness and damping coefficients. Both coefficients decrease with increasing SFD clearance. When the clearance is less than 0.14 mm, the pressure distribution amplitude rapidly decreases, leading to a quick reduction in the SFD stiffness and damping coefficients. When the clearance exceeds 0.14 mm, the reduction rate of the pressure distribution amplitude and the stiffness and damping coefficients slow down.

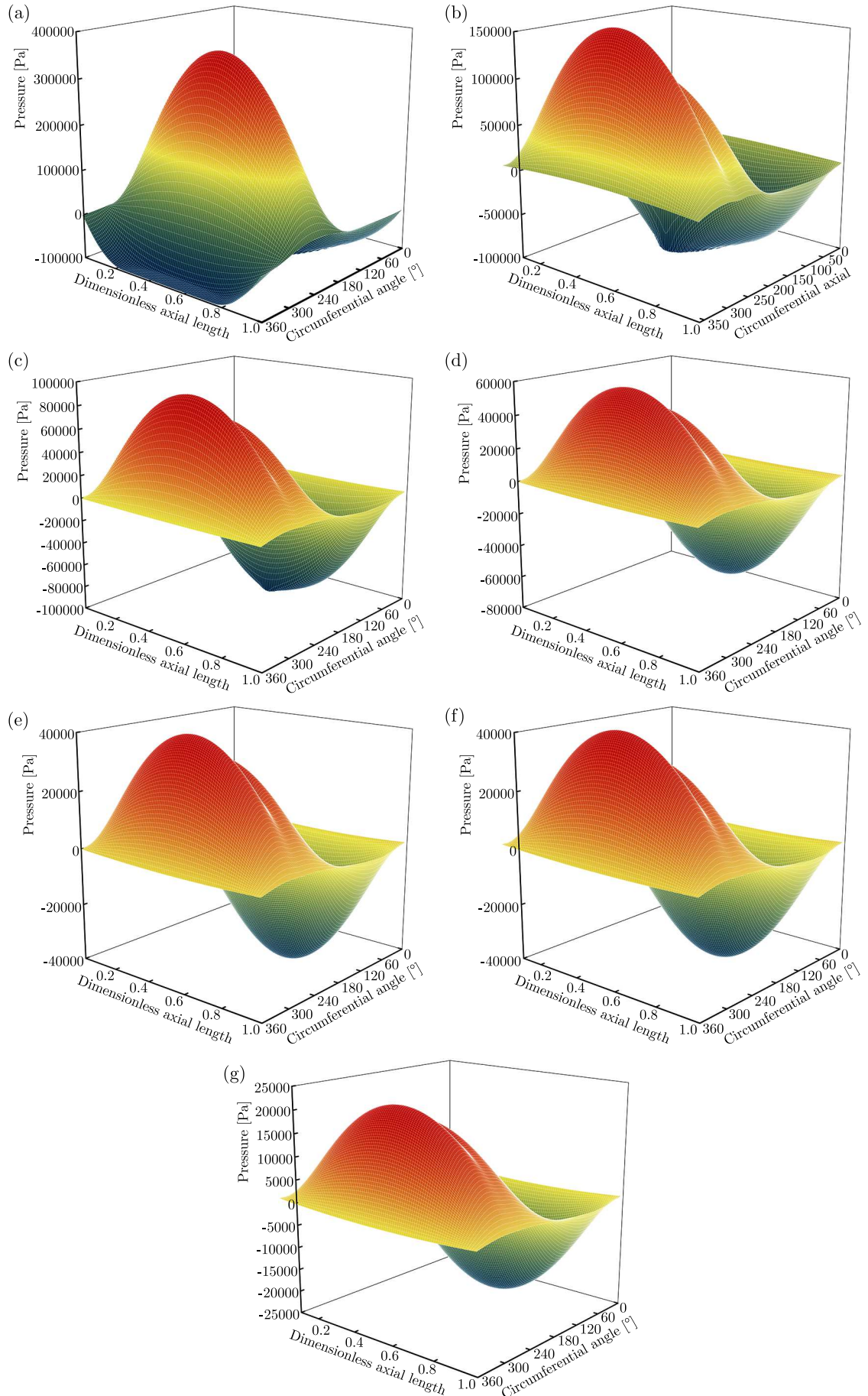


Fig. 9. SFD pressure distribution: (a) $c = 0.08\,\text{mm}$, (b) $c = 0.1\,\text{mm}$, (c) $c = 0.12\,\text{mm}$, (d) $c = 0.14\,\text{mm}$, (e) $c = 0.16\,\text{mm}$, (f) $c = 0.18\,\text{mm}$, (g) $c = 0.2\,\text{mm}$

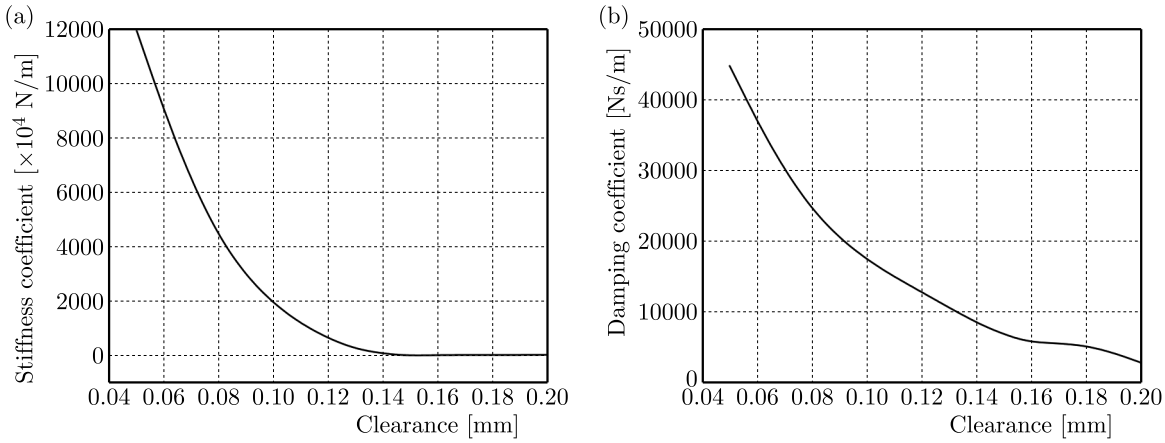


Fig. 10. (a) Stiffness and (b) damping coefficients

4. Dynamical characteristics of the rotor system

To investigate the influence of SFD clearance on the transient response of the rotor-SFD system, different clearances are selected: 0.05, 0.08, 0.1, 0.12, 0.14, 0.15, 0.16, 0.18 and 0.2 mm. The other damper parameters are kept constant. Figure 11 shows the transient responses of the elastic support of the rotor-SFD system under different clearances.

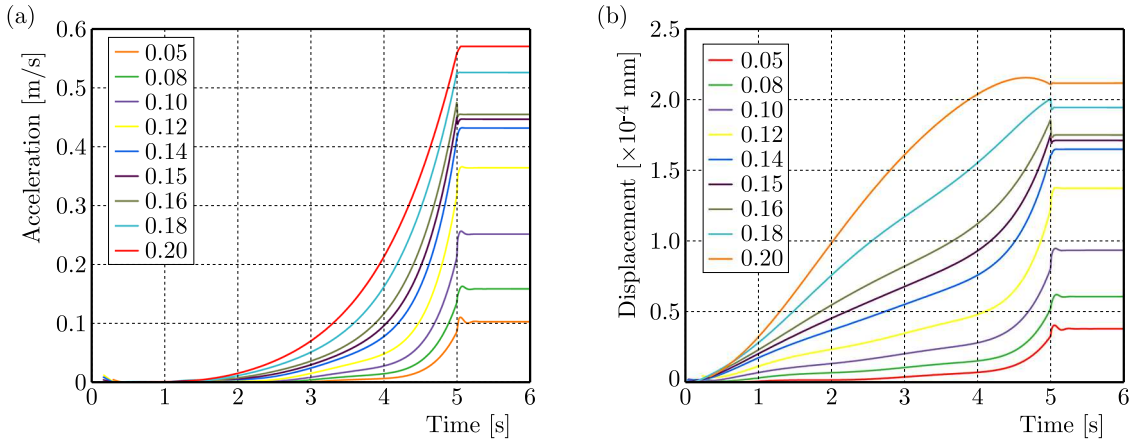


Fig. 11. Transient responses of the elastic support of the rotor-SFD system: (a) acceleration, (b) displacement

The figure depicts the acceleration and displacement response curves of the elastic support during the acceleration process of the rotor-SFD system from 0 rpm to 9000 rpm. The vibration amplitude of the rotor system increases with the SFD clearance. When the SFD clearance is less than 0.14 mm, the damping coefficient rapidly decreases, leading to a quick reduction in the vibration amplitude of the rotor system. Subsequently, the reduction rate of the vibration amplitude slows down. When the SFD clearance exceeds 0.16 mm, the rate of increase of the vibration amplitude accelerates. However, asynchronous displacement and acceleration amplitudes are observed.

5. Conclusions

This study established an SFD dynamic model considering both cavitation and fluid inertia, and analyzed the influence of SFD clearance on SFD cavitation and vibration reduction charac-

teristics. Moreover, a finite element model of the rotor-SFD system considering SFD damping characteristics was established. The influence of the SFD clearance on the transient response of the rotor-SFD system was studied. The accuracy of the dynamic model was experimentally verified. The following conclusions were obtained:

- The mismatch between the experimental and numerical simulation results was less than 5%, indicating that the numerical simulation method for the SFD and the finite element model of the rotor-SFD system proposed herein are reliable.
- When the SFD clearance was increased from 0.08 mm to 0.12 mm, both the extent of cavitation and the gas volume fraction decreased. As the SFD clearance exceeded 0.14 mm, the extent of cavitation expanded and the gas volume fraction gradually increased.
- The pressure distribution magnitude and the SFD stiffness and damping coefficients decreased with increasing SFD clearance. Additionally, the vibration damping effect of the SFD decreased.
- The vibration amplitude of the rotor system increased with the SFD clearance.

The proposed numerical simulation method of SFD allows for the study of the influence of structural and operating parameters of dampers on their vibration reduction characteristics, considering the effects of cavitation and inertia. Moreover, the finite element model of the rotor-SFD system presented in this study enables investigation of the influence of structural and operating parameters of the damper on the transient response of the rotor system. The proposed method enables better selection of the structural and operating parameters of an SFD during the design phase of the rotor system effectively reducing the vibration amplitude and improving stability of the rotor system during operation.

References

1. Chen H., Hou L., Chen Y., 2017, Bifurcation analysis of a rigid-rotor squeeze film damper system with unsymmetrical stiffness supports, *Archive of Applied Mechanics*, **87**, 8, 1347-1364
2. Crandall S.H., El-Shafei A., 1993, Momentum and energy approximations for elementary squeeze-film damper flows, *Journal of Applied Mechanics*, **60**, 3, 728-736
3. Drapatow T., Alber O., Woschke E., 2021, Transient simulation of a squeeze film damped turbocharger rotor under consideration of fluid inertia and cavitation, *Bulletin of the Polish Academy of Sciences-Technical Sciences*, **69**, 6, 1-8
4. El-Saeidy F.M.A., Sticher F., 2010, Dynamics of a rigid rotor linear/nonlinear bearings system subject to rotating unbalance and base excitations, *Journal of Vibration and Control*, **16**, 3, 403-438
5. El-Shafei A., 1991, Unbalance response of a Jeffcott rotor incorporating long squeeze film dampers, *Journal of Vibration and Acoustics*, **113**, 1, 85-94
6. El-Shafei A., 1995, Modeling fluid inertia forces of short journal bearings for rotordynamic applications, *Journal of Vibration and Acoustics*, **117**, 4, 462-469
7. El-Shafei A., 2002, A finite difference model for squeeze film dampers, *Journal of Engineering and Applied Science*, **49**, 1, 159-174
8. Fan T., Hamzehlouia S., Behdinan K., 2017, The effect of lubricant inertia on fluid cavitation for high-speed squeeze film dampers, *Journal of Vibroengineering*, **19**, 8, 6122-6134
9. Gehannin J., Arghir M., Bonneau O., 2016, A volume of fluid method for air ingestion in squeeze film dampers, *Tribology Transactions*, **59**, 2, 208-218
10. Gheller E., Chatterton S., Vania A., Pennacchi P., 2022, Squeeze film damper modeling: a comprehensive approach, *Machines*, **10**, 9, 1-26

11. Gupta R.K., Singh R.C., 2023, Comprehensive experimental analysis of a squeeze film damper for flexible rotor applications: utilizing Box-Behnken design with desirability optimization, *Journal of Vibration Engineering and Technologies*, **12**, 3, 5267-5290

12. Hamzehlouia S., Behdinan K., 2019, Squeeze film dampers supporting high-speed rotors: fluid inertia effects, *Proceedings of the Institution of Mechanical Engineers, Part J – Journal of Engineering Tribology*, **234**, 1, 18-32

13. San Andrés L., 1992, Analysis of short squeeze film dampers with a central groove, *Journal of Tribology*, **114**, 4, 659-664

14. San Andrés L., 2014, Force coefficients for a large clearance open ends squeeze film damper with a central feed groove: experiments and predictions, *Tribology International*, **71**, 17-25

15. Wang H., Zhao Y., Luo Z., Han Q., 2022, Analysis on influences of squeeze film damper on vibrations of rotor system in aeroengine, *Applied Sciences*, **12**, 2, 615

16. Younan A.A., Cao J., Dimond T.W., Allaire P.E.,, 2010, Nonlinear analysis of squeeze film damper with entrained air in rotordynamic systems, *Tribology Transactions*, **54**, 1, 132-144

17. Zapoměl J., Ferfecki P., Forte P., 2019, Vibrations attenuation of a Jeffcott rotor by application of a new mathematical model of a magnetorheological squeeze film damper based on a bilinear oil representation, *Acta Mechanica*, **230**, 1625-1640

18. Zhang M., 2018, *Simulation and Experimental Study on Dynamic Characteristics of Squeeze Film Damper* (in Chinese), Mechanical Engineering, Shenyang Aerospace University, Shenyang

19. Zhou H.L., Cang Y.G., Zhang Y.Q, Guo C., 2023, Analysis of dynamic characteristics of a sealed ends squeeze film damper considering the fluid inertia force, *Journal of Theoretical and Applied Mechanics*, **61**, 3, 441-452

20. Zhou H.L., Chen X., Zhang Y.Q., Ai Y.T., Sun D., 2020, An analysis on the influence of air ingestion on vibration damping properties of squeeze film dampers, *Tribology International*, **145**, 106168

Manuscript received March 30, 2024; accepted for publication September 14, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 751-761, Warsaw 2024
https://doi.org/10.15632/jtam-pl/194505

DYNAMICS MODELING OF VARIABLE MASS SYSTEMS – A CASE STUDY OF AN UNDERWATER INERTIA BASED PROPELLED GLIDER PERFORMANCE

Zbigniew Kostka, Elżbieta Jarzębowska

Warsaw University of Technology, Faculty of Power and Aeronautical Engineering, Warsaw, Poland

e-mail: zbigniew.kostka.dokt@pw.edu.pl; elzbieta.jarzebowska@pw.edu.pl

Underwater gliders are autonomous underwater vehicles that are widely used in oceanography and coastal surveillance due to their low manufacturing costs and long operation time. This paper addresses the development of a dynamical model of such vehicles which are inertia propelled. The dynamical model is based upon the Boltzmann-Hamel equations modified to variable mass and inertia systems. It yields dynamics in a body-fixed frame using non-inertial coordinates. The theoretical development of the vehicle dynamics based upon the modified Boltzmann-Hamel equations is validated by the longitudinal dynamics model of the underwater glider and its performance resulted from the mass change.

Keywords: underwater gliders, AUV, variable-mass systems, Boltzmann-Hamel equations

1. Introduction

An underwater glider is a type of autonomous underwater vehicle (AUV) that is characterized by its specific way of propulsion. It changes its depth in water using different volume of the liquid in ballast tanks or an external blader. The net buoyancy or mass change is the factor that propels the glider forward. The greater the change from the equilibrium, the faster the vehicle traverses the given distance. Additionally, it is equipped with a moveable internal mass whose change of positions is used to enable vehicle turning. The vehicle thus characterizes by low manufacturing costs and exhibits long operation time. It can traverse thousands of kilometers during its mission as opposed to the range of a few kilometers by a typical vehicle with thrusters. Due to its characteristics, this class of underwater vehicles excels in oceanographic research (Rudnick, 2016), mostly in observation of oceanic fronts and water-mass properties, e.g. salinity or temperature (Wagawa *et al.*, 2020). It is also capable of long coastal surveillance missions, where its excels other underwater vehicles. The inertia based propelled underwater glider concept dates back to 1989, when Henry Stommel came up with an idea of a machine that operates without any external propulsion drives (Stommel, 1989). Then, many solutions came to life – starting from ALBAC (Kawaguchi *et al.*, 1995), the first prototype of a glider which could perform only one glide cycle. Next, three AUV's that share a very similar design: Slocum (Schofield *et al.*, 2007), Seaglider (Eriksen, 2001) and Spray (Rudnick *et al.*, 2016), were designed and built. There are also hybrid gliders which combine the attributes of the inertia based propelled glider and a non-inertially propelled one like Petrel (Wang *et al.*, 2011), Fölaga (Alvarez *et al.*, 2009) or blended body vehicle Zray (Brodsky and Luby, 2013). A different approach to the inertia based propelled system was adopted by Liquid Robotics, which developed a two-part system – a wave glider made of a surface float and an underwater glider (Hine *et al.*, 2009).

An inertia based propelled glider is a variable mass or buoyancy system that modifies its depth using different volume of the liquid in ballast tanks or in the external blader. The net buoyancy or mass change is the factor that propels the glider forward. The greater the change from the equilibrium, the faster the vehicle traverses the given distance. Moveable mass that may change its location inside the glider, is used to enable turning motions (Mahmoudian *et al.*, 2007). Hydrodynamic forces acting upon the glider are often calculated using the CFD approach (Sun *et al.*, 2021). They include lift, drag and the added mass calculations. However, the underwater gliders move slowly, and the classical hydrodynamics is enough to determine the forces and the Magnus effect is usually omitted due to its insignificance for slowly moving objects.

The most popular method to model underwater vehicle dynamics is the one based on Euler--Newton, which was used by Graver (2005) and Sun *et al.* (2023) or the Lagrange equations (Cruz, 2011) or fluid-multibody coupling (Wang *et al.*, 2023). Usually, the Lagrange based dynamics models use the Euler angles and they are equations in the body fixed frames identical to the flight dynamics models (Fossen, 1994). The dynamics model derivation based upon the Boltzmann--Hamel equations, which we present in the paper, is one of the most recent approaches that can be used to handle the variable mass system and be convenient for the controller design for the glider. The Boltzmann-Hamel equations were originally formulated for constant mass systems in non-inertial coordinates, often referred to as quasi-velocities, see e.g. (Neimark and Fufaev, 1972). Contrary to popular approaches like the Lagrange equations method, they are free of multipliers for constraints, which are incorporated into the equations of motion instead of being an additional algebraic relation (Müller, 2021). They enable handling nonholonomic systems due to arbitrary selection of quasi-velocities that may satisfy the nonholonomic constraints. There are a few examples of using the Boltzmann-Hamel equations approach to modelling and control of multibody systems, see e.g. (Jarzębowska, 2012). However, in the paper (Jarzębowska and Cichowski, 2018) the underwater vehicle dynamics is developed under some design and performance constraints that limit its usability, i.e. centers of masses are assumed to be in one axis and the change of position of the center of masses of ballast tanks are not taken into consideration.

In this paper, the concentricity of the centers of masses of the vehicle components and the center of buoyancy requirement was removed, and an additional parameter describing the position of the mass center of the ballast tank was added. To confirm the improvement of performance of the glider model with respect to performance of real gliders, simulation studies are presented. A longitudinal model of the glider has been derived and tested. The glider motion in the longitudinal plane is its typical motion, so it was selected for testing. A test that validates the propelling efficiency was performed using an open-loop control algorithm. The glider moved the so-called saw-tooth motion pattern, a traversal motion method that is typical for it in real-life motion scenarios. The control inputs and outputs were inspected and proposals for upgrading the performance through a targeted control strategy have been made. The contribution of this paper is the development of the modified Boltzmann-Hamel equations that account for the change of mass and inertia in the underwater vehicle. The model can have the constraints put upon the mass change. Assumptions about specific locations of CB and CG were removed, thus the dynamic model yields results closer to the reality. Based upon this framework, the longitudinal glider model is developed and validated.

The paper is organized as follows. After Introduction, Section 2 reports the Boltzmann-Hamel equations development for a constant mass system and presents their modification for variable mass systems. Section 3 details the underwater glider physical and dynamical models. Section 4 validates the glider dynamics and demonstrates its performance in longitudinal motion. The paper closes with conclusions, future research prospects and the list of references.

2. Boltzmann-Hamel equations for constant and variable mass systems

Boltzmann-Hamel equations originally developed for constant mass systems proved to be more efficient method of dynamics derivation for both holonomic and nonholonomic system models, the ones like underwater gliders, than the most widely used approaches like Newton-Euler or Lagrange equations. They are derived in quasi-velocities that represent velocities in a non-inertial, usually body fixed, reference frame. Due to the derivation method of the Boltzmann--Hamel equations and arbitrary selection of the quasi-velocities, the constraint equations can be taken as the quasi-velocities and, as such, incorporated into the system dynamics and reduce the size of the resulting system dynamics. In this Section we briefly report the derivation of the Boltzmann-Hamel equations. The detailed derivation and discussion of the equations can be found in e.g. (Neimark and Fufaev, 1972). Herein, we recall the development of the form of the equations as they were originally developed and we recall the concepts of quasi-velocities. These are needed to develop dynamics of the underwater vehicle.

The main motivation to reach for the Boltzmann-Hammel modeling method is that it proved to be effective in control applications for both unconstrained and constrained system models (Jarzębowska, 2012).

2.1. Quasi-velocities and generalized coordinates

Quasi-velocities ω_σ with σ equal to the number of states can be selected arbitrarily as a linear or also as nonlinear combination of the generalized velocities $\dot{q}_i$ as in (2.1). Usually, they are the parameters associated with velocities in the body fixed frame, therefore they are referenced as non-inertial velocities

$$\omega_\sigma = a_{\sigma,0}(q) + \sum_{i=1}^{n} a_{\sigma,i}(q)\dot{q}_i \tag{2.1}$$

Quasi-coordinates are defined by their differentials as in (2.2). These terms are not integrable, therefore the parameters π_σ do not exist physically. If relations (2.2) can be integrated, then they are generalized velocities, whose derivatives can also serve as quasi-velocities.

$$d\pi_\sigma = a_{\sigma,0}(q)dt + \sum_{i=1}^{n} a_{\sigma,i}(q)dq_i \tag{2.2}$$

Usually, for selection of quasi-velocities we assume that $a_{\sigma,0}(q) = 0$, so expression (2.1) can be rewritten as $(2.3)_1$. Assuming that relations $(2.3)_1$ are invertible, the generalized velocities can be presented as linear combinations of quasi-velocities $(2.3)_2$. Based upon (2.2) and $(2.3)_2$, relation $(2.3)_3$ can be written. Combining expressions $(2.3)_{1,2}$, we can determine relations (2.4) between the coefficients $a_{l,\sigma}$ and $b_{\sigma,j}$

$$\omega_\sigma = \sum_{i=1}^{n} a_{\sigma,i}(q)\dot{q}_i \qquad \dot{q}_\sigma = \sum_{j=1}^{n} b_{\sigma,j}(q)\omega_j \qquad \delta q_\sigma = \sum_{j-1}^{n} b_{\sigma,j}(q)\delta\pi_j \tag{2.3}$$

and

$$\sum_{\sigma=1}^{n} a_{l,\sigma}b_{\sigma,j} = \delta_{l,j} = \begin{cases} 1 & \text{for} \quad l = j \\ 0 & \text{for} \quad l \neq j \end{cases} \tag{2.4}$$

In the case of the m nonholonomic constraint equations imposed upon a system, relations (2.3) consist of m quasi-velocities identically satisfying the constraint equations, and the rest $(n-m)$ of quasi-velocities are selected arbitrarily, which is practically suitable to the considered system modling.

2.2. The Boltzmann-Hamel equations for constant mass systems

The Boltzmann-Hamel equations can be derived, roughly speaking, from Lagrange equations $(2.5)_1$, where the generalized coordinates are replaced by quasi-velocities (2.1) and quasi-coordinates (2.2). The Lagrange equations, assuming independent δq_σ $(2.5)_1$ can be presented as in $(2.5)_2$

$$\begin{aligned}
&\sum_{\sigma=1}^{n}\Big[\frac{d}{dt}\Big(\frac{\partial T}{\partial \dot{q}_\sigma}\Big)-\frac{\partial T}{\partial q_\sigma}-Q_\sigma\Big]\delta q_\sigma=0\\
&\sum_{\sigma=1}^{n}\frac{d}{dt}\Big(\frac{\partial T}{\partial \dot{q}_\sigma}\Big)b_{\sigma,j}-\sum_{\sigma=1}^{n}\frac{\partial T}{\partial q_\sigma}b_{\sigma,j}=\sum_{\sigma=1}^{n}Q_\sigma b_{\sigma,j}
\end{aligned}\tag{2.5}$$

After denoting the kinetic energy written in quasi-velocities as T^* and taking into account $(2.3)_2$, a relation between $T^*(\omega,q,t)$ and $T(\dot{q},q,t)$ can be derived as in $(2.6)_1$. Relation $(2.6)_1$ is then used in transformations of the first term of $(2.5)_2$. It yields $(2.6)_2$

$$\begin{aligned}
&\frac{\partial T^*}{\partial \omega_j}=\sum_{\sigma=1}^{n}\frac{\partial T}{\partial \dot{q}_\sigma}\frac{\partial \dot{q}_\sigma}{\partial \omega_j}=\sum_{\sigma=1}^{n}\frac{\partial T}{\partial \dot{q}_\sigma}b_{\sigma,j}\\
&\sum_{\sigma=1}^{n}\frac{d}{dt}\Big(\frac{\partial T}{\partial \dot{q}_\sigma}\Big)b_{\sigma,j}=\frac{d}{dt}\sum_{\sigma=1}^{n}\frac{\partial T}{\partial \dot{q}_\sigma}b_{\sigma,j}-\sum_{\sigma=1}^{n}\frac{\partial T}{\partial \dot{q}_\sigma}\frac{\partial b_{\sigma,j}}{\partial t}=\frac{d}{dt}\Big(\frac{\partial T^*}{\partial \omega_j}\Big)\\
&\quad+\sum_{\sigma=1}^{n}\sum_{l=1}^{n}\sum_{\lambda=1}^{n}\sum_{\alpha=1}^{n}\frac{\partial T^*}{\partial \omega_l}\omega_\alpha\frac{\partial a_{l,\sigma}}{\partial q_\lambda}b_{\sigma,j}b_{\lambda,\alpha}
\end{aligned}\tag{2.6}$$

The derivative of T with respect to generalized coordinates can be expanded as in $(2.7)_1$. It is then used for transformation of the second term of $(2.5)_2$ and it results in $(2.7)_2$

$$\begin{aligned}
&\frac{\partial T}{\partial q_\sigma}=\frac{\partial T^*}{\partial q_\sigma}+\sum_{l=1}^{n}\frac{\partial T^*}{\partial \omega_l}\frac{\partial \omega_l}{\partial q_\sigma}=\frac{\partial T^*}{\partial q_\sigma}+\sum_{l=1}^{n}\sum_{\lambda=1}^{n}\sum_{\alpha=1}^{n}\frac{\partial T^*}{\partial \omega_l}\frac{\partial a_{l,\lambda}}{\partial q_\sigma}b_{\lambda,\alpha}\omega_\alpha\\
&\sum_{\sigma=1}^{n}\frac{\partial T}{\partial q_\sigma}b_{\sigma,j}=\frac{\partial T^*}{\partial \pi_j}+\sum_{\sigma=1}^{n}\sum_{l=1}^{n}\sum_{\lambda=1}^{n}\sum_{\alpha=1}^{n}\frac{\partial T^*}{\partial \omega_l}\frac{\partial a_{l,\lambda)}}{\partial q_\sigma}b_{\lambda,\alpha}\omega_\alpha b_{\sigma,j}
\end{aligned}\tag{2.7}$$

The right hand side term in $(2.5)_2$ is rewritten as $(2.8)_1$, and Q_j^* is the generalized force related to quasi-velocity ω_j. Relations $(2.6)_1$, $(2.7)_1$ and $(2.8)_1$ when inserted into $(2.5)_2$ result in equation $(2.8)_2$

$$\begin{aligned}
&Q_j^*=\sum_{\sigma=1}^{n}Q_\sigma b_{\sigma,j}\\
&\frac{d}{dt}\Big(\frac{\partial T^*}{\partial \omega_j}\Big)-\frac{\partial T^*}{\partial \pi_j}+\sum_{\alpha=1}^{n}\sum_{l=1}^{n}\frac{\partial T^*}{\partial \omega_l}\omega_\alpha\sum_{\sigma=1}^{n}\sum_{\lambda=1}^{n}\Big(\frac{\partial a_{l,\sigma}}{\partial q_\lambda}-\frac{\partial a_{l,\lambda}}{\partial q_\sigma}\Big)b_{\sigma,j}b_{\lambda,\alpha}=Q_j^*
\end{aligned}\tag{2.8}$$

After introducing the so called Boltzmann-Hamel coefficients $(2.9)_1$, equation $(2.8)_2$ can be presented as $(2.9)_2$

$$\begin{aligned}
&\gamma_{j\alpha}^{l}=\sum_{\sigma=1}^{n}\sum_{\lambda=1}^{n}\Big(\frac{\partial a_{l,\sigma}}{\partial q_\lambda}-\frac{\partial a_{l,\lambda}}{\partial q_\sigma}\Big)b_{\sigma,j}b_{\lambda,\alpha}\\
&\frac{d}{dt}\Big(\frac{\partial T^*}{\partial \omega_j}\Big)-\frac{\partial T^*}{\partial \pi_j}+\sum_{\alpha=1}^{n}\sum_{l=1}^{n}\gamma_{j\alpha}^{l}\frac{\partial T^*}{\partial \omega_l}\omega_\alpha=Q_j^*
\end{aligned}\tag{2.9}$$

The third term of equation $(2.9)_2$ can be further simplified, i.e. writing the Boltzmann-Hammel coefficients into a matrix form $(2.10)_1$ we can obtain a single sum as $(2.10)_2$. Recalling that the

quasi-velocity vector is $(n \times 1)$ dimensional, the final form of equation $(2.9)_2$ can be presented as n equations $(2.10)_3$ referred to as the Boltzmann-Hamel equations

$$\begin{aligned}
&\boldsymbol{\Gamma}_\alpha = \begin{bmatrix} \gamma_{1\alpha}^1 & \gamma_{1\alpha}^2 & \cdots & \gamma_{1\alpha}^n \\ \gamma_{2\alpha}^1 & \gamma_{2\alpha}^2 & \cdots & \gamma_{2\alpha}^n \\ \vdots & \vdots & \vdots & \ddots \\ \gamma_{n\alpha}^1 & \gamma_{n\alpha}^2 & \cdots & \gamma_{n\alpha}^n \end{bmatrix} \\
&\mathbf{G}(\boldsymbol{\omega}) = \sum_{\alpha=1}^{n} \boldsymbol{\Gamma}_\alpha \omega_\alpha \\
&\frac{d}{dt}\Big(\frac{\partial T^*}{\partial \boldsymbol{\omega}}\Big) - \frac{\partial T^*}{\partial \boldsymbol{\pi}} + \mathbf{G}(\boldsymbol{\omega})\frac{\partial T^*}{\partial \boldsymbol{\omega}} = \mathbf{Q}^*
\end{aligned} \tag{2.10}$$

2.3. Boltzmann-Hamel equations for variable-mass systems

The original Boltzmann-Hamel equations cannot be directly applied to derive dynamical models of variable mass and variable inertia systems. Therefore, some modifications are needed. The kinetic energy of the system written in quasi-velocities depends on the form of the inertia matrix $\mathbf{M}$ (2.11). It is time dependent or may also depend upon other parameters, e.g. velocity, if mass or inertia are prescribed to change according to some constraint equations. It impacts the form of equations $(2.10)_3$.

In this Section, we present a modification of the Boltzmann-Hamel equations to encompass the mass and inertia changes. The main idea is to explicitly account for the change of mass of water in the vehicle water tanks and motion the movable mass. The need of doing so is to obtain an explicit exposition of these terms in the system dynamical equations. This in turn is essential for a design of the control system. In this paper, we design a feedforward controller only to regulate mass of water in the water tank, however, the future work aims at the design of feedback control for both the regulation of water amount and for movable mass motion. Control of the movable mass is essential for performing maneuvers by the vehicle. Concluding then, the dynamic underwater vehicle model which we present herein will be transformed into a dynamic control model with the control inputs being the change of mass of water in the water tank and the positions of the movable mass.

To develop the variable inertia dynamics, we assume a modified form of the inertia matrix that includes explicitly a time derivative of the $\mathbf{M}$ in (2.11)

$$\frac{d}{dt}\Big(\frac{\partial T^*}{\partial \boldsymbol{\omega}}\Big) = \mathbf{M}(\dot{\boldsymbol{\omega}}) + \dot{\mathbf{M}}\boldsymbol{\omega} \tag{2.11}$$

Thus the variable mass and inertia Boltzmann-Hamel equations can be written in a compact form

$$\mathbf{M}(\dot{\boldsymbol{\omega}}) + \dot{\mathbf{M}}\boldsymbol{\omega} - \frac{\partial T^*}{\partial \boldsymbol{\pi}} + \mathbf{G}(\boldsymbol{\omega})\frac{\partial T^*}{\partial \boldsymbol{\omega}} = \mathbf{Q}^* \tag{2.12}$$

Equation (2.12) is applied to develop dynamical equations of the longitudinal motion for the underwater vehicle model.

3. Dynamics model of the underwater glider

To validate dynamics of the variable inertia system (2.13), this Section presents development of a longitudinal model of the underwater glider to confirm its saw-tooth motion pattern and tests through simulations the inertia based propelling efficiency using an open-loop controller. The

longitudinal motion is a dominant one in real gliders so it can be taken for model validation. The glider saw-tooth motion pattern is a traversal method which is typical in real-life scenarios. The control inputs and outputs were inspected and propositions for upgrading the performance through a targeted control strategy have been made.

In this paper, we adopt the assumption about the design of a typical underwater glider. In contrary to (Jarzębowska and Cichowski, 2018), an assumption about concentricity of centers of masses of the vehicle components and the center of buoyancy were removed, and an additional parameter describing the position of the mass center of the ballast tank was added. The underwater glider is assumed to be composed of two main components: a hull and a ballast tank. The buoyancy center (CB) of the hull is the same as the CB of the whole system, but its mass center (CM) does not coincide with the center of mass of the vehicle. The other mentioned components are treated as point masses

$$M = m_h + m_{tank} \tag{3.1}$$

The movable mass is not present in our model, since it is installed for the space maneuvers of the glider. Also, the second general assumption is that we do not assume that all vehicle components are located along the vehicle longitudinal axis, i.e. they can be displaced from x axis; see Fig. 1. Additionally, the location of the CM of the ballast tank changes according to the change of mass of the liquid rested in it. One more assumption enables positioning the CM of the vehicle below its CB, thus increasing its stability. All these assumptions make our glider model close to real designs of such kind of vehicles.

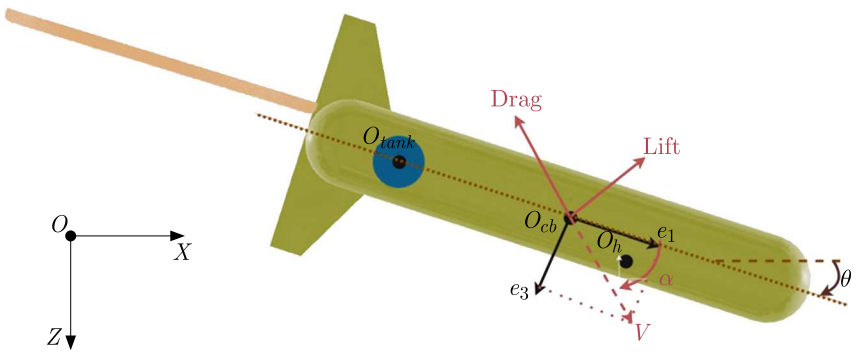


Fig. 1. A glider model and its components

Calculations which we present are performed in two reference frames: inertial and the body frame. The z and e_3 axes point downward instead of upward, as it is traditionally assumed in many works. The change was done to emphasize that the glider moves below the water surface. The glider model is presented in Fig. 1 and the data representing its parameters are shown in Table 1.

To describe the glider longitudinal motion, a state vector consists of 3 quasi-velocities and 2 generalized velocities $(3.2)_3$ and $(3.2)_4$. The quasi-velocities are the linear velocities along the body axis e_1, e_3, and Q refers for the pitch angle. Notice, that Q is the generalized velocity adopted as the quasi-velocity. The generalized velocities are derivatives of the position coordinates along inertia axes x and z $(3.2)_2$. Equations $(3.2)_5$ and $(3.2)_6$ present transformations between the quasi-velocities and generalized velocities

Table 1. Glider data

Parameter	Significance
$m_h = 26.9\,\mathrm{kg}$	mass of the hull
$z_{CB,tank} = 0\,\mathrm{m}$	z position of CM of ballast tank with respect to CB of the glider
$x_{CB,h} = 0\,\mathrm{m}$	x position of CM of hull with respect to CB of the glider
$z_{CB,h} = 0.1\,\mathrm{m}$	z position of CM of hull with respect to CB of the glider
$g = 9.81\,\mathrm{m/s^2}$	acceleration due to gravity
$v = 0.0274\,\mathrm{m^3}$	volume of glider
$\rho = 1023.6\,\mathrm{kg/m^3}$	density of seawater
$K_{D0} = 7.19\,\mathrm{kg/m}$	hydrodynamic constant impacting linear velocity U
$K_D = 386.29\,\mathrm{kg/(m{\cdot}rad^2)}$	hydrodynamic constant associated with square of α impacting linear velocity U
$K_\alpha = 440.99\,\mathrm{kg/(m{\cdot}rad)}$	hydrodynamic constant associated with α impacting linear velocity W
$K_M = 65.84\,\mathrm{kg/rad}$	hydrodynamic constant associated with α impacting angular velocity Q
$K_Q = 205.64\ \mathrm{kg{\cdot}s/rad^2}$	hydrodynamic constant associated with angular velocity Q impacting Q
$I_{CM,h} = 5.919\,\mathrm{kg{\cdot}m^2}$	moment of inertia of hull

$$
\begin{aligned}
&\boldsymbol{\omega}_v = \begin{bmatrix} U \\ W \end{bmatrix} \qquad \dot{\mathbf{q}}_v = \begin{bmatrix} \dot{x}_{CB} \\ \dot{z}_{CB} \end{bmatrix} \\
&\boldsymbol{\omega}^{\mathrm{T}} = \begin{bmatrix} \omega_v^{\mathrm{T}} & Q & \dot{x}_{CB,tank}^{\mathrm{T}} \end{bmatrix} \qquad \dot{\mathbf{q}}_v^{\mathrm{T}} = \begin{bmatrix} \dot{q}_v^{\mathrm{T}} & \dot{\theta} & \dot{x}_{CB,tank}^{\mathrm{T}} \end{bmatrix} \\
&\boldsymbol{\omega}_v = \begin{bmatrix} U \\ W \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \dot{x}_{CB} \\ \dot{z}_{CB} \end{bmatrix} = \boldsymbol{\Lambda}_v \dot{\overline{\mathbf{q}}}_v \qquad Q = \dot{\theta}
\end{aligned}
\tag{3.2}
$$

The model of the glider was created with the use of the Boltzmann-Hamel equations for variable inertia (2.12). Due to the change of mass in the ballast tank, the mass M and the position of the CM of the whole glider with respect to its CB $\mathbf{r}_{CB,CM}$ is depicted by $(3.3)_1$ and the inertia matrix $\mathbf{M}$ of the system are both time-dependent. The moment of inertia $(3.3)_2$ includes the moment of inertia of the hull $I_{CM,h}$, masses of the ballast tank m_{tank} and hull m_h as well as the positions of the CM of the hull $r_{CB,h}$ and of the ballast tank $\mathbf{r}_{CB,tank}$ $(3.3)_3$ with respect to the CB of the glider. They are all time-dependent. Therefore, inertia matrix of the glider $(3.3)_4$ depends not only on the position changes of the mass centers, but also on the change of mass in the ballast tank

$$
\begin{aligned}
&\mathbf{r}_{CB,CM} = \begin{bmatrix} x_{CB,CM} \\ z_{CB,CM} \end{bmatrix} \\
&I = I_{CM,h} + n_{tank} r_{CB,tank}^{\mathrm{T}} r_{CB,tank} + m_h r_{CB,h}^{\mathrm{T}} r_{CB,h} \\
&\mathbf{r}_{CB,tank} = \begin{bmatrix} x_{CB,tank} \\ z_{CB,tank} \end{bmatrix} \\
&T = \frac{1}{2} \overline{\boldsymbol{\omega}}^{\mathrm{T}} \begin{bmatrix} M & 0 & M z_{CB,CM} & m_{tank} \\ 0 & M & -M x_{CB,CM} & 0 \\ M z_{CB,CM} & -M x_{CB,CM} & I & m_{tank} z_{CB,tank} \\ m_{tank} & 0 & m_{tank} z_{CB,tank} & m_{tank} \end{bmatrix} \overline{\boldsymbol{\omega}} = \frac{1}{2} \overline{\boldsymbol{\omega}}^{\mathrm{T}} \mathbf{M} \overline{\boldsymbol{\omega}}
\end{aligned}
\tag{3.3}
$$

Three external forces act upon the glider model $(3.4)_1$. They are gravity $(3.4)_2$, buoyancy $(3.4)_3$ and hydrodynamic forces $(3.4)_4$. The hydrodynamic forces account only for a normalized drag and lift. The added mass effect and the Magnus effect are excluded from the calculations. The Magnus effect is irrelevant due to small velocity of the glider. The added mass effect is assumed to be constant for a specific pitch angle. The hydrodynamic forces are transformed to the body-fixed frame from the non-inertial wind frame using the attack angle α $(3.5)_1$ and the total velocity of the vehicle V_0 $(3.5)_2$

$$
\begin{aligned}
&\mathbf{F}_{ext} = \mathbf{F}_{G,CB} + \mathbf{F}_{B,CB} + \mathbf{F}_{H,CB} \\
&\mathbf{F}_{G,CB} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \\ z_{CB,CM}\cos\theta & x_{CB,CM}\sin\theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ Mg \end{bmatrix} \\
&\mathbf{F}_{B,CB} = \begin{bmatrix} \mathbf{1}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0}_{2\times 2} \end{bmatrix} \begin{bmatrix} 0 \\ -\rho v g \end{bmatrix} \\
&\mathbf{F}_{H,CB} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 & 0 \\ \sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -K_{D0} - K_D(\alpha^2) \\ -K_\alpha \alpha \\ -K_M \alpha - K_Q Q \\ 0 \end{bmatrix} V_0^2
\end{aligned}
\tag{3.4}
$$

and

$$
\alpha = \arctan 2(W, U) \qquad V_0 = \sqrt{U^2 + W^2} \tag{3.5}
$$

4. Simulation of the saw-tooth motion of the underwater glider

The so called saw-tooth motion is a typical motion of an inertia-propelled underwater vehicle. In this Section we develop the longitudinal dynamics model of the vehicle and validate its performance. Specifically, the model validates the self-propelling capabilities of the vehicle due to the change of its mass and inertia. The underwater vehicle increases its net weight when it is close to the surface and reduces it at the maximum desired depth. The cycle repeats throughout the entire operation. The bigger the net weight change, the faster the vehicle moves. The mass change is achieved through changes of the mass or volume of the liquid in the ballast tank. The process occurs near the surface or at the maximum depth only, therefore the inertia propelled glider is much more energy efficient than the one equipped with other propulsion means like thrusters, which need a constant influx of energy to keep floating and moving forward.

During the simulation studies, an open-loop control algorithm was used to execute forward motion by the saw-tooth pattern. Using the open-loop control, we get an insight into the glider dynamics with respect to the future feedback control design. In the simulation study, there was one control input, i.e. the mass change in the ballast tank. The mass of water in the tank changes between 0.56 kg and 2.2 kg as seen in Fig. 2a. The resulting saw-tooth motion is shown in Fig. 2b. The kind of periodicity of the saw-tooth motion pattern depends on how fast is the change of water amount in the water tanks. Initial values for quasi-velocities were set to $U_0 = 0.5\,\mathrm{m/s}$ and $W_0 = 0.25\,\mathrm{m/s}$.

The underwater glider controlled by the control input function shown in Fig. 2a is able to achieve approximately constant velocity. The small velocity deviations came from the net weight change process at the surface or the maximum depth and a slight difference of the absolute value of the net weight during sinking and surfacing. The traversal distance in the x direction in time is shown in Fig. 3a, while depth in time is in Fig. 3b.

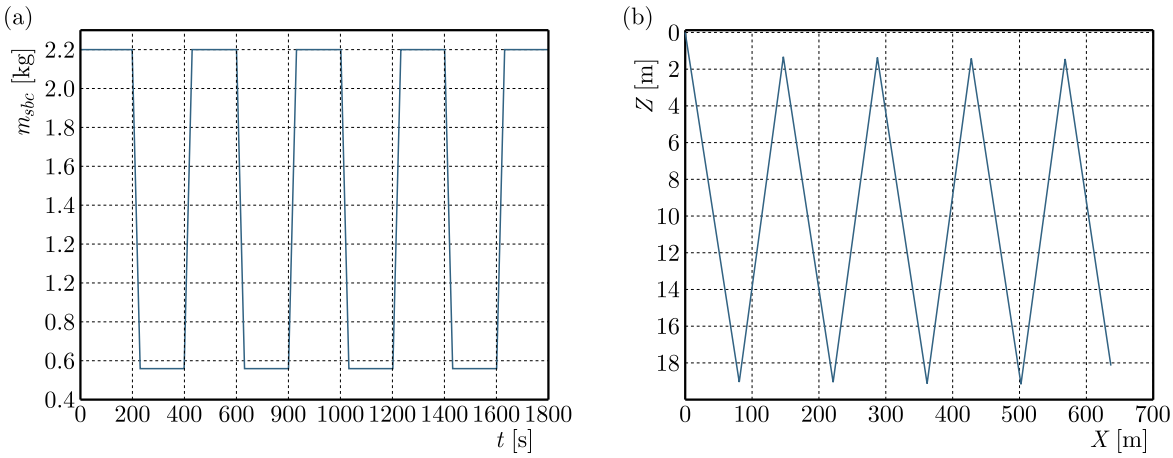


Fig. 2. (a) Control input time history for saw-tooth motion – mass change in the front ballast tank. (b) Saw-tooth trajectory pattern of the underwater glider

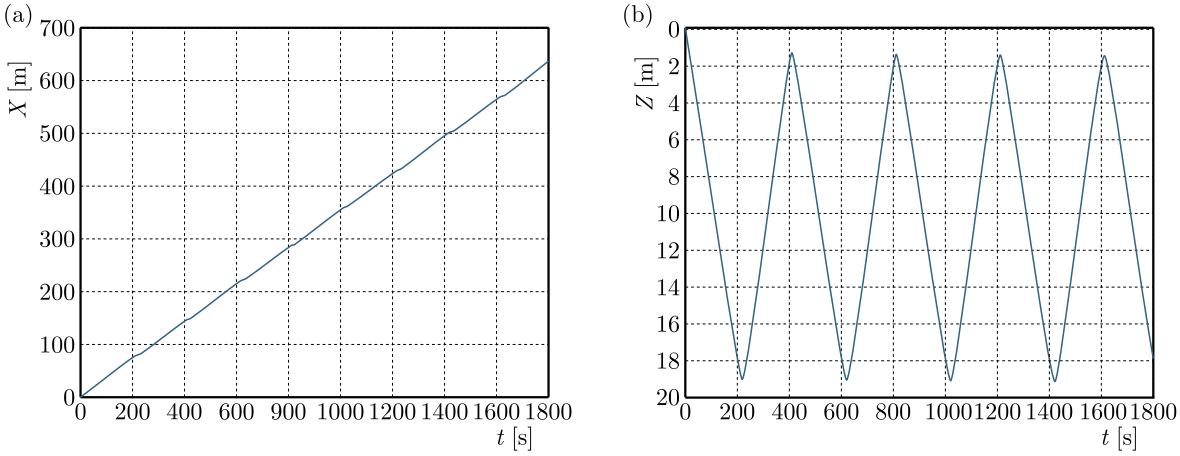


Fig. 3. (a) The traversal distance in the x direction in time covered by the glider. (b) The vertical distance in the z direction in time covered by the glider

The angle in the longitudinal motion that changed in time was the pitch angle. The changes of the pitch angle in time are presented in Fig. 4. The pitch angle initial value is set to be equal to $\pi/8$.

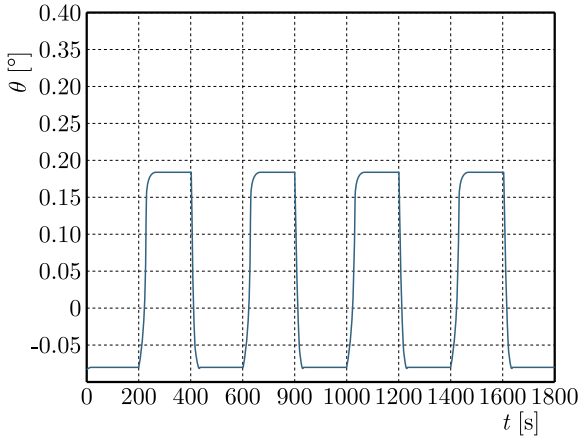


Fig. 4. The changes of the underwater glider pitch angle in time

5. Conclusion and perspective

The problem of modeling of the underwater glider dynamics is addressed in the paper with the aid of the Boltzmann-Hamel equations modified to encompass variable mass and inertia of the system. The modified Boltzmann-Hamel equations can serve for modeling of a system with any desired change of mass in water tanks and any motion of a movable mass that enables vehicle turning maneuvers.

In our modeling approach, the constraints that are assumed as in many references, i.e. concentricity of mass centers of the glider body components and the buoyancy center, and no influence of the change of mass on the location of the mass center of the ballast tank were relaxed.

The advantages of the modified modeling approach may serve both for modeling of constrained variable mass underwater vehicles, including constraints on the mass change, and for simplifing control applications like moving along desired trajectories or controling the mass change required for some maneuvers.

The presented dynamics model still requires some extensions, specifically as it is intended to be used for control applications. For example, it would need more precise calculation of hydrodynamic forces acting upon the vehicle. Future research aims at calculating the forces using the CFD method to get results closer to real ones than simplified outcomes in form of constant values for a specific attack angle what is often assumed in many references.

References

1. Alvarez A., Caffaz A., Caiti A., Casalino G., Gualdesi L., *et al.*, 2009, Fòlaga: a low-cost autonomous underwater vehicle combining glider and AUV capabilities, *Ocean Engineering*, **36**, 1, 24-38

2. Brodsky P., Luby J., 2013, *Flight Software Development for the Liberdade Flying Wing Glider*, Office of Naval Research, Arlington

3. Cruz N. (Ed.), 2011, *Autonomous Underwater Vehicles*, InTech, DOI: 10.5772/923

4. Eriksen C.C., Osse T.J., Light R.D., Wen T., Lehman T.W., *et al.*, 2001, Seaglider: a long-range autonomous underwater vehicle for oceanographic research, *IEEE Journal of Oceanic Engineering*, **26**, 4, 424-436

5. Fossen T.I., 1994, *Guidance and Control of Ocean Vehicles*, 1st ed., Wiley

6. Graver J.G., 2005, Underwater gliders: dynamics, control and design, Ph.D. Thesis, Princeton University, Princeton

7. Hine R., Willcox S., Hine G., Richardson T., 2009, The wave glider: a wave-powered autonomous marine vehicle, *Oceans 2009, IEEE Conference*, 1-6

8. Jarzębowska E., 2012, *Model-Based Tracking Control of Nonlinear Systems*, CRC Series: Modern Mechanics and Mathematics, CRC Press, Taylor & Francis Group, Boca Raton

9. Jarzębowska E., Cichowski M., 2018, Dynamics modeling and performance analysis of underwater vehicles based on the Boltzmann-Hamel equations approach, *MATEC Web of Conferences*, **148**, 03005

10. Kawaguchi K., Ura T., Oride M., Sakamaki T., 1995, Development of shuttle type AUV "ALBAC" and sea trials for oceanographic measurement, *Journal of the Society of Naval Architects of Japan*, **178**, 657-665

11. Mahmoudian N., Geisbert J., Woolsey C., 2007, *Dynamics and Control of Underwater Gliders I: Steady Motions*, Virginia Center for Autonomous Systems, Blacksburg, Virginia

12. Mahmoudian N., Geisbert J., Woolsey C., 2010, *Dynamics and Control of Underwater Gliders II: Steady Motions*, Virginia Center for Autonomous Systems, Blacksburg, Virginia

13. Müller A., 2021, On the Hamel coefficients and the Boltzmann-Hamel equations for the rigid body, *Journal of Nonlinear Science*, **31**, 2, 40

14. Neimark J.I., Fufaev N.A., 1972, *Dynamics of Nonholonomic Systems*, American Mathematical Society, Providence, RI

15. Rudnick D.L., 2016, Ocean research enabled by underwater gliders, *Annual Review of Marine Science*, **8**, 519-541

16. Rudnick D.L., Davis R.E., Sherman J.T., 2016, Spray Underwater Glider Operations, *Journal of Atmospheric and Oceanic Technology*, **33**, 6, 1113-1122

17. Schofield O., Kohut J., Aragon D., Creed L., Graver J., *et al.*, 2007, Slocum gliders: robust and ready, *Journal of Field Robotics*, **24**, 6, 473-485

18. Stommel H., 1989, The slocum mission, *Oceanography*, **2**, 1, 22-25

19. Sun C., Tian J., Huang R., Dong H., Li H., Ma Y., 2023, Internal layout optimization of the blended-wing-body underwater glider based on a range target, *Ocean Engineering*, **280**, 114450

20. Sun W., Zang W., Liu C., Guo T., Nie Y., Song D., 2021, Motion pattern optimization and energy analysis for underwater glider based on the multi-objective artificial bee colony method, *Journal of Marine Science and Engineering*, **9**, 3, 327

21. Wagawa T, Kawaguchi Y., Igeta Y., Honda N., Okunishi T., Yabe I., 2020, Observations of oceanic fronts and water-mass properties in the central Japan Sea: Repeated surveys from an underwater glider, *Journal of Marine Systems*, **201**, 103242

22. Wang H., Chen J., Feng Z., Li Y., Deng C., Chang Z., 2023, Dynamics analysis of underwater glider based on fluid-multibody coupling model, *Ocean Engineering*, **278**, 114330

23. Wang S., Sun X., Wang Y., Wu J., Wang X., 2011, Dynamic modeling and motion simulation for a winged hybrid-driven underwater glider, *China Ocean Engineering*, **25**, 1, 97-112

24. Zhang S., Yu J., Zhang A., Zhang F., 2013, Spiraling motion of underwater gliders: Modeling, analysis, and experimental results, *Ocean Engineering*, **60**, 1-13

Manuscript received December 10, 2023; accepted for publication October 11, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 763-767, Warsaw 2024
https://doi.org/10.15632/jtam-pl/194242

SHORT RESEARCH COMMUNICATION

RANDOM EIGENVIBRATIONS OF BEAMS WITH VISCOELASTIC LAYERS[1]

Marcin Kamiński
Łódź University of Technology, Faculty of Civil Engineering, Architecture and Environmental Engineering, Łódź, Poland
Maciej Przychodzki, Magdalena Łasecka-Plura, Michał Guminiak, Agnieszka Lenartowicz
Poznan University of Technology, Institute of Structural Analysis, Poznań, Poland
corresponding author M. Przychodzki, e-mail: maciej.przychodzki@put.poznan.pl

This paper is devoted to the study of the influence of random variation of model parameters of a beam with viscoelastic layers on probabilistic characteristics of its natural frequencies and dimensionless damping coefficients. The relationships between the model parameters and the dynamic characteristics of the beam were approximated by quartic polynomials based on the results of calculations using FEM, where beam finite elements were used, taking into account lamination of the beam. The nonlinear eigenproblem was solved using the continuation method. The calculation results for an example laminated beam are presented.

Keywords: viscoelasticity, probabilistic characteristics, dynamic characteristics, layered beam, fractional Zener model

1. Introduction

Correct determination of basic dynamic characteristics of structures is one of the important engineering problems. However, uncertainties or variability of model parameters may have a significant impact on obtained results. Therefore, research on the influence of random variability of model parameters on probabilistic characteristics of natural vibration frequencies and dimensionless damping coefficients seems justified. This paper presents results of research conducted for a beam with layers of a viscoelastic material described by the fractional Zener model. The basic numerical tool used in this type of analysis is the Stochastic Finite Element Method – SFEM (Arregui-Mena *et al.*, 2016; Kamiński, 2013; Stefanou, 2009). To determine the probabilistic characteristics, it was necessary to perform a number of deterministic calculations for variable parameters of the beam model. The Finite Element Method was applied to these calculations, using the layered beam finite element proposed by Lewandowski and Baum (2015). This type of finite element was also used in the studies by Łasecka-Plura (2023). The dynamic characteristics of the beam were obtained by solving a nonlinear generalized eigenproblem, for which the continuation method described by Pawlak and Lewandowski (2013) was used. Research on the influence of variability of model parameters on the dynamic characteristics of plates resting on viscoelastic supports was carried out by Kamiński *et al.* (2023).

2. Finite element formulation

The dynamic analysis of the beam with viscoelastic layers was performed based on the Finite Element Method (FEM), using beam finite elements, but taking into account the layered structure

[1]Paper presented during PCM-CMM 2023, Cracow, Poland

of the beam. The finite element of the layered beam was formulated in the frequency domain. The following assumptions were made in this formulation: each viscoelastic layer is located between two elastic layers, all layers are perfectly glued to each other, the material of each layer is isotropic and homogeneous, the Euler-Bernoulli beam theory was used for elastic layers and the Timoshenko beam theory for viscoelastic layers. The corresponding layers of adjacent finite elements must have the same thickness, and damping in elastic layers is neglected. The assumptions made in this way lead to the formulation of kinematic relations for the finite element of a laminated beam, which are described in detail in the paper by Lewandowski and Baum (2015).

The fractional Zener model was chosen to describe the viscoelastic material. This model describes the actual behavior of a viscoelastic material very well, requiring only four material parameters. The constitutive equation of this model is as follows

$$\sigma(x,t) + \tau^{\alpha} D_t^{\alpha} \sigma(x,t) = E_0 \varepsilon(x,t) + E_{\infty} \tau^{\alpha} D_t^{\alpha} \varepsilon(x,t) \tag{2.1}$$

This equation employs commonly used notations, namely $\sigma(x,t)$ and $\varepsilon(x,t)$ are functions of stress and strain, respectively, and τ is the relaxation time. The symbols E_0 and E_{∞} indicate the relaxed and non-relaxed elastic modules, respectively. The fractional derivative of order α is marked with the symbol $D_t^{\alpha}(\cdot)$. In the presented research, the definition of the Riemann-Louville fractional derivative was used.

After some mathematical operations, the following matrix equation of beam motion in the frequency domain can be written (Lewandowski and Baum, 2015)

$$[s^2\mathbf{M} + \mathbf{K} + \mathbf{K}_v(s)]\overline{\mathbf{q}}(s) = \overline{\mathbf{p}}(s) \tag{2.2}$$

where s is the Laplace variable, wherein $\overline{\mathbf{q}}(s)$ and $\overline{\mathbf{p}}(s)$ are the Laplace transforms of the nodal displacement vector and the external force vector, respectively. The symbols $\mathbf{M}$, $\mathbf{K}$ and $\mathbf{K}_v(s)$ denote, respectively, the mass matrix, the stiffness matrix and the so-called viscoelastic matrix, the elements of which are functions of the variable s.

3. Dynamic characteristics of the beam with viscoelastic layers

The natural frequencies and dimensionless damping coefficients of the beam with viscoelastic layers are calculated based on the solution of the nonlinear eigenproblem, which is obtained by assuming a zero vector of the external forces in Eq. (2.2)

$$[s^2\mathbf{M} + \mathbf{K} + \mathbf{K}_v(s)]\overline{\mathbf{q}}(s) = \mathbf{0} \tag{3.1}$$

This eigenproblem can be solved by the continuation method (Pawlak and Lewandowski, 2013). Since the obtained eigenvalues are complex numbers, they can be written as

$$s_i = \mu_i + \mathrm{i}\eta_i \tag{3.2}$$

The natural frequencies and dimensionless damping coefficients can be calculated accordingly

$$\omega_i^2 = \mu_i^2 + \eta_i^2 \qquad \gamma_i = -\frac{\mu_i}{\omega_i} \tag{3.3}$$

4. Influence of model parameters variation on dynamic characteristics of a layered beam

The main goal of the analyzes was to examine the impact of variability of specific parameters of the beam model on its dynamic characteristics. The analyzed characteristics are natural frequencies of the beam and dimensionless damping coefficients. In order to determine the relationship

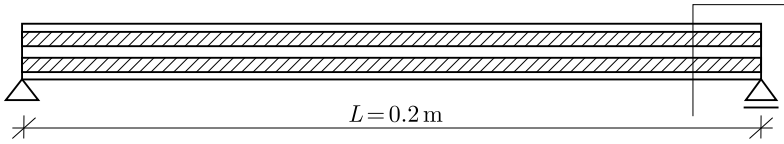
between a specific model parameter and the appropriate dynamic characteristic of the beam, a polynomial approximation using the least squares method was used based on the results of FEM calculations, i.e. a finite number of deterministic results. Quartic polynomials were adopted as approximating functions.

Inference on the influence of variability of model parameters on the natural frequencies and dimensionless damping coefficients of the laminated beam was made on the basis of the following probabilistic characteristics: expected value, standard deviation, coefficient of variation, skewness and kurtosis.

Calculations of probabilistic characteristics are, by their nature, complex tasks, which is why approximate methods are generally used in these tasks. In the conducted research, the calculation of probabilistic characteristics was carried out using three computational methods, and consistency of the obtained results was compared. The methods used were semi-analytical method – SAM (based on symbolic calculation procedures in the Maple program), stochastic perturbation technique – SPT (tenth order) and Monte-Carlo simulation – MCS (number of trials is 10^5).

5. Numerical example

This paper presents results of sample calculations performed for a beam consisting of five layers: three elastic and two viscoelastic. The model of such a beam is shown in Fig. 1.



Aluminium	$h_{e1} = 0.001\,\mathrm{m}$
Viscoelastic material	$h_{v1} = 0.002\,\mathrm{m}$
Aluminium	$h_{e2} = 0.004\,\mathrm{m}$
Viscoelastic material	$h_{v2} = 0.002\,\mathrm{m}$
Aluminium	$h_{e3} = 0.001\,\mathrm{m}$

Fig. 1. Model of an examplary layered beam

The total length of the beam is 20 cm. The lower and upper elastic layers have the same thickness, which is $h_{e1} = h_{e3} = 0.001\,\mathrm{m}$, while the thickness of the middle elastic layer is $h_{e3} = 0.004\,\mathrm{m}$. The remaining parameters of the elastic layers are as follows: Young's modulus $E_e = 70.3\,\mathrm{GPa}$, Poisson's ratio $\nu_e = 0.3$, density $\rho_e = 2690\,\mathrm{kg/m^3}$. The parameters of both viscoelastic layers are the same and their values are as follows: layer thickness $h_{v1} = h_{v2} = 0.002\,\mathrm{m}$, density $\rho_e = 1600\,\mathrm{kg/m^3}$, relaxed elastic modulus $E_0 = 1.5\,\mathrm{MPa}$, non-relaxed elastic modulus $E_\infty = 70\,\mathrm{MPa}$, Poisson's ratio – $\nu_v = 0.5$, relaxation time $\tau = 1.4 \cdot 10^{-5}\,\mathrm{s}$, parameter describing the order of derivative fractional $\alpha = 0.8$. The viscoelastic material parameters were taken from the literature (Galucio *et al.*, 2004) and describe the polymer 3M ISD112.

The analysis of the influence of variability of the parameters of the layered beam model on the probabilistic characteristics of its natural vibration frequency and dimensionless damping coefficients was performed for four parameters: thickness of the elastic layer h_{e3}, thickness of the viscoelastic layer h_{v2}, relaxation time τ and the parameter describing the order of the fractional derivative α. These parameters were treated as random variables with a Gaussian probability distribution, for which the coefficient of variation was assumed to have values ranging from 0.025 to 0.25.

As previously mentioned, the calculations were performed using three different methods: semi-analytical, perturbation technique and Monte-Carlo simulation. Figure 2 shows the results obtained using these methods. As can be seen, the graphs of the expected value, coefficient of variation, skewness and kurtosis of the first natural vibration frequency depending on the coefficient of variation of the parameter α obtained by these three methods show high agreement.

As a part of the research, a summary of the probabilistic characteristics obtained for the model parameters was made for the coefficient of variation varying within the assumed range

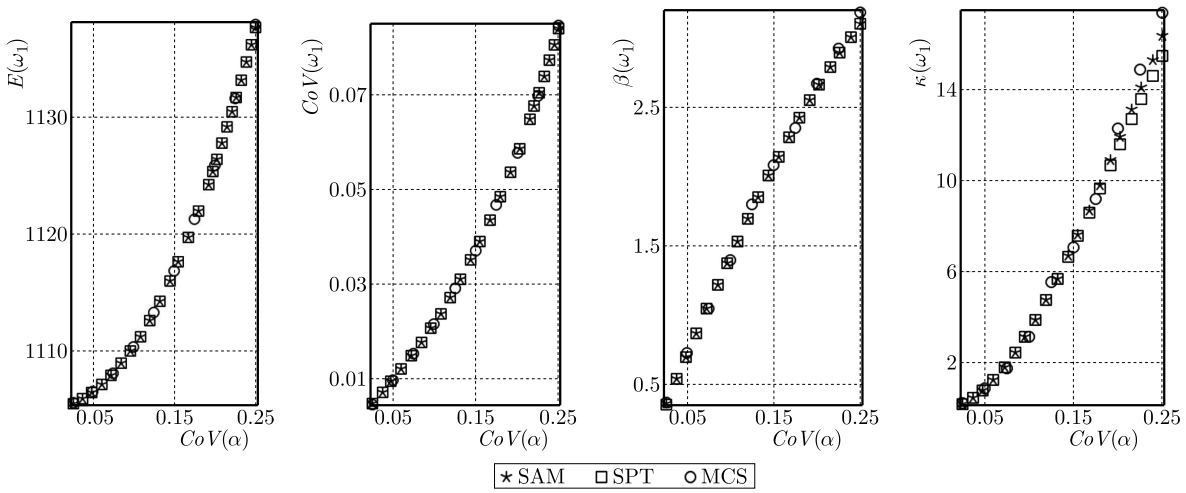


Fig. 2. Validation of methods for calculating probabilistic characteristics: expected value $E(\omega_1)$, coefficient of variation $CoV(\omega_1)$, skewness $\beta(\omega_1)$, kurtosis $\kappa(\omega_1)$ (SAM – semi-analytical method, SPT – stochastic perturbation technique, MCS – Monte Carlo simulation)

and is shown in Fig. 3. The main observation is that in the case of the first natural frequency, an increase in the variability of the parameter α has by far the greatest impact on the probabilistic characteristics. However, it should be emphasized that even in this case, the coefficient of variation of the first natural frequency ω_1 is significantly smaller than the corresponding coefficient of variation of this parameter. The impact of the variability of the remaining parameters on the probabilistic characteristics of ω_1 is significantly smaller.

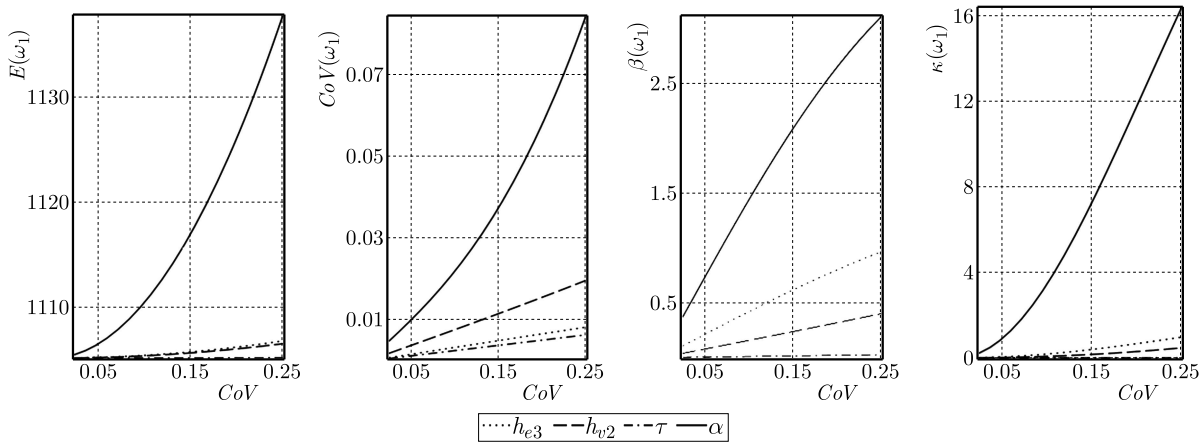


Fig. 3. Comparison of probabilistic characteristics of ω_1: expected value $E(\omega_1)$, coefficient of variation $CoV(\omega_1)$, skewness $\beta(\omega_1)$, kurtosis $\kappa(\omega_1)$ for variable coefficient of variation (CoV) of beam model parameters (results obtained by a semi-analytical method)

6. Conclusions

The calculation results presented in the previous Section allow the following conclusions to be formulated:

- Analogous analyzes as in the case of natural frequencies were performed for dimensionless damping coefficients. The results of these studies will be presented in the extended version of the article.

- Comparison of the results obtained from the semi-analytical method, stochastic perturbation technique and Monte-Carlo simulation shows that all these methods generally give similar results, but in some cases they may differ, so to be sure of the correctness of the solution, it is worth using at least two of them in parallel.
- An increase in the variability of the viscoelastic material model parameter α has by far the greatest impact on the probabilistic characteristics of natural vibration.
- Only selected research results are presented, the scope of which was much wider. In the remaining analyses, the conclusions presented above were confirmed.

Acknowledgments

This work contains the results obtained in the framework of research grant OPUS No. 2021/41/B/ST8/02432 entitled "Probabilistic entropy in engineering computations", and sponsored by the National Science Center in Cracow, Poland, 2022-2025.

References

1. Arregui-Mena J.D., Margetts L., Mummery P.M., 2016, Practical application of the stochastic finite element method, *Archives of Computational Methods in Engineering*, **23**, 1, 171-190
2. Galucio A.C., Deü J.-F., Ohayon R., 2004, Finite element formulation of viscoelastic sandwich beams using fractional derivative operators, *Computational Mechanics*, **33**, 4, 282-291
3. Kamiński M., 2013, *The Stochastic Perturbation Method for Computational Mechanics*, John Wiley & Sons, Ltd., Hoboken, NJ, USA
4. Kamiński M., Guminiak M., Lenartowicz A., Łasecka-Plura M., Przychodzki M., Sumelka W., 2023, Stochastic nonlinear eigenvibrations of thin elastic plates resting on time-fractional viscoelastic supports, *Probabilistic Engineering Mechanics*, **74**, 103522
5. Lewandowski R., Baum M., 2015, Dynamic characteristics of multilayered beams with viscoelastic layers described by the fractional Zener model, *Archive of Applied Mechanics*, **85**, 12, 1793-1814
6. Łasecka-Plura M., 2023, Dynamic characteristics of a composite beam with viscoelastic layers under uncertain-but-bounded design parameters, *Applied Sciences*, **13**, 11, 6473
7. Pawlak Z., Lewandowski R., 2013, The continuation method for the eigenvalue problem of structures with viscoelastic dampers, *Computers and Structures*, **125**, 53-61
8. Stefanou G., 2009, The stochastic finite element method: Past, present and future, *Computer Methods in Applied Mechanics and Engineering*, **198**, 9-12, 1031-1051

Manuscript received November 30, 2023; accepted for publication October 7, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 769-786, Warsaw 2024
https://doi.org/10.15632/jtam-pl/195069

INFLUENCE OF GEOMETRICAL DEFECTS ON AERODYNAMIC DRAG OF NON-WATERTIGHT AHMED BODY

Xuelong Liu
Tianjin University, Tianjin, China; and
CATARC (Tianjin) Automotive Engineering Research Institute Co., Ltd., Tianjin, China
e-mail: lixuelong@catarc.ac.cn

Haidong Yuan, Qing Qin
CATARC (Tianjin) Automotive Engineering Research Institute Co., Ltd., Tianjin, China
e-mail: yuanhaidong@catarc.ac.cn; qinqing@catarc.ac.cn

Songze Chen
Shenzhen Tenfong Technology Co., Ltd., Shenzhen, China
e-mail: chensz@tenfong.cn

A non-watertight Ahmed body is adopted to examine the influence of geometrical defects on the aerodynamic drag. The Ahmed body, a standard automotive test model, is widely used in wind tunnel experiments to study vehicle aerodynamics. However, in real-world applications, vehicles often exhibit deviations from the ideal geometry due to manufacturing tolerances, wear, and other factors. The aim of this study is to quantify the impact of such geometrical defects on the aerodynamic drag. The research employs the lattice Boltzmann method (LBM) software to analyze the flow field around the Ahmed body with and without the introduction of geometrical defects. These defects are modeled as small holes in the shape of the body. The simulations are performed under the varying location and size of the holes to explore the aerodynamics phenomena. The results indicate that the location and size of geometrical defects can significantly alter the aerodynamic drag of the Ahmed body. The holes located at the rear of the Ahmed body rarely effect the aerodynamic drag. The holes at the top or bottom are found to have the most pronounced effect. The study also reveals that the influence of defects varies with their size, with large areas leading to more substantial changes in the aerodynamic drag. The results demonstrate that non-watertight geometry with small defects can be used to produce a reasonable drag coefficient compared to the results of watertight geometry.

Keywords: Ahmed body, non-watertight geometry, aerodynamic drag

1. Introduction

The Ahmed body is a highly recognized testing apparatus in automotive aerodynamics. It serves as a simplified yet representative model of a passenger car for studying aerodynamic drag and flow characteristics. It was originally designed by Ahmed *et al.* (1984). The Ahmed body exhibits a distinct shape that closely resembles the fundamental contours of a real-world car. Its front section is flat and connected with the rectangle body by a curved surface. The most notable feature of the Ahmed body is its truncated rear end, which is characterized by a slant angle that can be adjusted to simulate different rear geometries of cars.

Many scholars and researchers conducted wind tunnel experiments on the Ahmed body with different rear slant angles. Ahmed *et al.* (1984) measured the drag coefficient and pressure distribution for different rear slant angles from 0° to 40°, showing complex flow phenomena at the rear. Bayraktar *et al.* (2001) conducted experiments with different Reynolds numbers in the

range of $2.2M$ to $13.2M$ for models with different rear slant angles of 0°, 12.5° and 25° and yaw angles between −15° to 15°, and compared their effects on the force and flow field of the model. Lienhart and Becker (2003) provided supplementary statistical data on average velocity distribution, average pressure distribution, and turbulent high-order quantities for their model experiments with rear slant angles of 25° and 35°. Strachan *et al.* (2007) compared the effects of static and moving ground on the drag coefficient and flow field of the model with a rear slant angle of 0° to 40°.

In addition to the above mentioned and many other experimental studies (ERCOFTAC, 2023), plenty of numerical simulations (Han, 1989; Gilliéron and Chometon, 1999; Keating *et al.*, 2008; Holman *et al.*, 2012) including LBM (Keating *et al.*, 2008; Fares, 2006; Kotapati *et al.*, 2009) were conducted for further exploration of the Ahmed body.

Han (1989), Gilliéron and Chometon (1999), Keating *et al.* (2008), Holman *et al.* (2012) comprehensively replicated the distinct drag coefficient characteristics of the Ahmed body in details. From 0° to approximately 9°, the drag coefficient of the Ahmed body decreases as the slant angle increases. Then it increases with an increase of the slant angle, and the drag coefficient reaches the highest value when it reaches 30°. When the slant angle is greater than 30°, the wake changes greatly, resulting in a sudden drop in the drag coefficient. In numerous numerical simulations, the Ahmed body serves as a simplified model for verification. Nonetheless, literature and technical reports rarely discuss real-world geometric imperfections in numerical studies. Perhaps, required by many numerical methods, the geometry had to be watertight. Any geometric imperfections or defects are corrected in advance. The aerodynamics of non-watertight geometry with geometrical defects was rarely explored in the past.

It is worth noting that the aerodynamic feature of non-watertight geometry also plays a crucial role in automotive design. Unlike simple geometries, say, sphere or cylinder, most geometries of the vehicle have many components. The assembly errors are common, and will make many small cracks or holes in the surface. Some of these defects are solely computational, but some are really generated in the manufacturing process. At present, almost all cracks and holes that prevent the geometry from being watertight are repaired or fixed in the preprocessing phase, regardless of whether they are real or computational.

In this article, we delve into the drag coefficient of the non-watertight Ahmed body. To assess its sensitivity, we introduce various types of openings into the surface of the standard Ahmed body, adjusting their sizes and locations. Our objective is to discern the key factor that significantly impacts the drag coefficient and identify which imperfections have a minimal effect. By analyzing the simulation results, we aim to determine the primary influence on the drag coefficient.

To achieve our goal, it is imperative to employ a CFD solver capable of handling non--watertight geometries. Therefore, we opt for TF-Lattice, a CFD software that utilizes the Lattice Boltzmann Method (LBM) (Qian *et al.*, 1992; Chen and Doolen, 1998; Jacob *et al.*, 2018). The grid system in TF-Lattice provides a flexible framework to represent non-watertight surfaces. A concise overview of TF-Lattice, along with its validation, is provided in the Appendix for further reference.

2. Geometric model

2.1. Standard Ahmed body with different rear slant angles

Reproduce the baseline results and validate the software.

Ahead of the non-watertight geometry, the standard Ahmed bodies with different rear slant angles are simulated by TF-Lattice to reproduce the baseline results achieved in the previous research to ensure the consistency and reliability of the software. By validating the software

against these established benchmarks (Ahmed *et al.*, 1984; Keating *et al.*, 2008; Hinterberger *et al.*, 2004), we aim to provide confidence of the software accuracy, robustness, and ability to perform as expected. A sketch of the Ahmed body (Hinterberger *et al.*, 2004) is shown in Fig. 1, which exhibits different vortices, separation, and reattachment phenomena at different rear slant angles. The length of the model is $L = 1.044\,\mathrm{m}$, the width $W = 0.389\,\mathrm{m}$, the height $H = 0.288\,\mathrm{m}$, and the ground clearance is $0.05\,\mathrm{m}$. There are four cylindrical support rods with a diameter of $0.03\,\mathrm{m}$ located between the bottom of the model and the ground. The simulation results of the drag coefficient and flow state will be present at four slant angles of 0°, 12.5°, 25° and 35°.

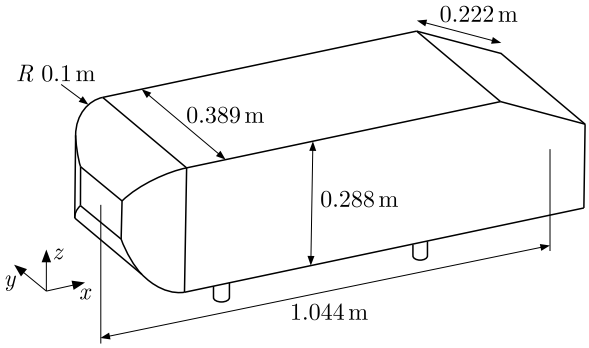


Fig. 1. Sketch of the Ahmed body

2.2. Non-watertight Ahmed body of 0° slant angle

To investigate the effect of geometrical defects on the drag coefficient, we introduced openings at various locations in the standard Ahmed body with a 0° slant angle, specifically in its upper surface, lower surface, lateral surface, and rear surface. We perform a series of numerical simulations, varying the area and distribution of the openings, with a particular emphasis on the lower surface. This focus is based on the recognition that the chassis section is the most intricate component of a vehicle and is where geometrical defects are most likely to occur. Therefore, our case studies concentrate on examining various configurations of the openings in the lower surface. The specific opening locations and corresponding parameters are shown in Table 1, and 3D views of various opening situations are shown in Fig. 2.

Table 1. Setup for the non-watertight Ahmed bodies

Case	Hole location	Area ratio (hole area/face area)
1	None	–
2	Top	10%
3	Lateral	10%
4	Rear	10%
5	Bottom	10%
6	Bottom	5%
7	Bottom	1% × 5
8	Bottom	1%

3. Computational settings

The computational domain is established as shown in Fig. 3. The total length of the computational domain is $10L$, where L represents length of the Ahmed body. The inlet of computational domain is located at $2L$ in front of the Ahmed body, and the outlet is located at $7L$ behind the

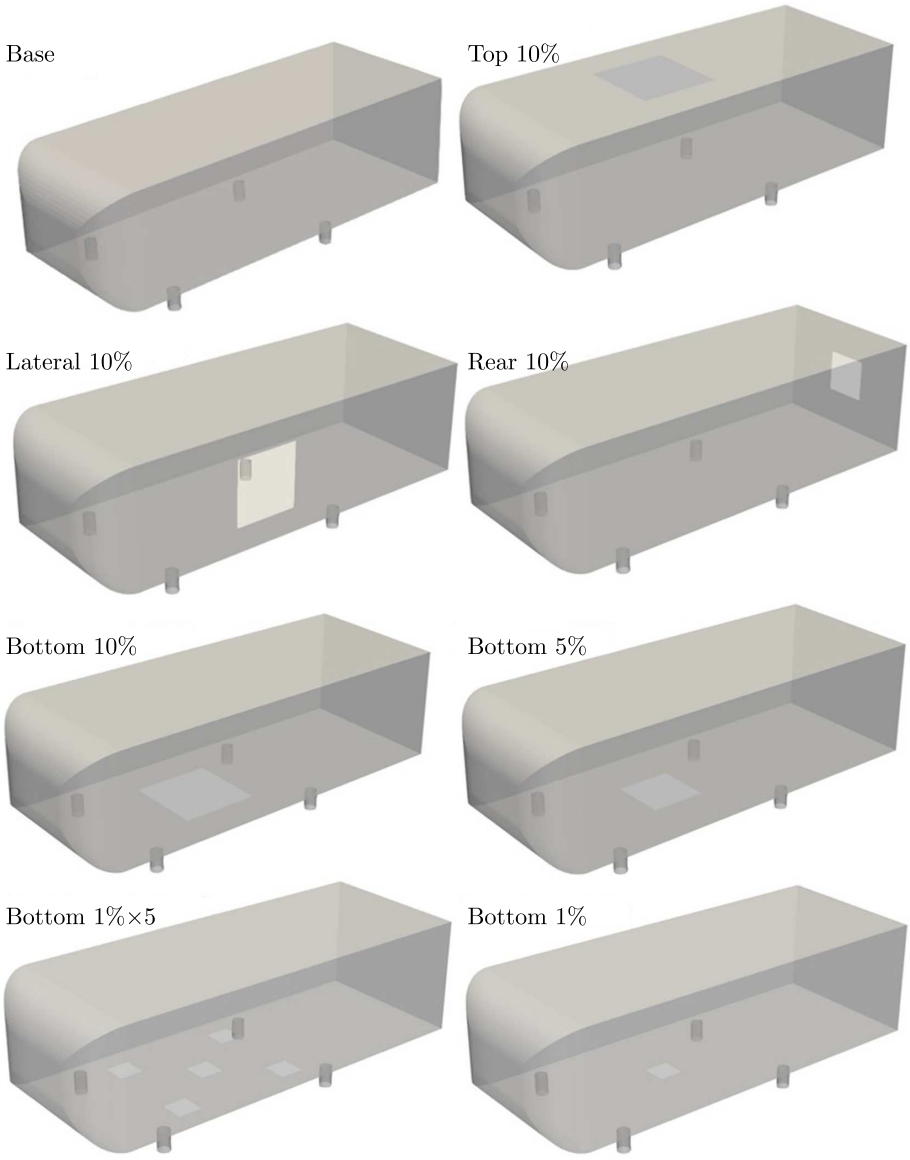


Fig. 2. Illustrations of the non-watertight Ahmed bodies (slant angle = 0°)

Ahmed body. The rear wake area is long enough to ensure that the wake is fully developed. The total domain width is $8W$, the height of computational domain is $9H$, and the blockage ratio of the whole model is 1.4%, which is a common blockage ratio in a wind tunnel.

Boundary conditions are exhibited in Table 2. The Reynolds number of the flow, which is about 4×10^5, is estimated with the incoming flow velocity of 20 m/s, air viscosity and the height H of the Ahmed body.

Table 2. Boundary conditions

Boundary name	Boundary condition
Outlet	Uniform velocity inlet of 20 m/s
Inlet	Far field
Spanwise and upper faces	Symmetry
Ground	Fixed no-slip wall
Ahmed body	Fixed no-slip wall

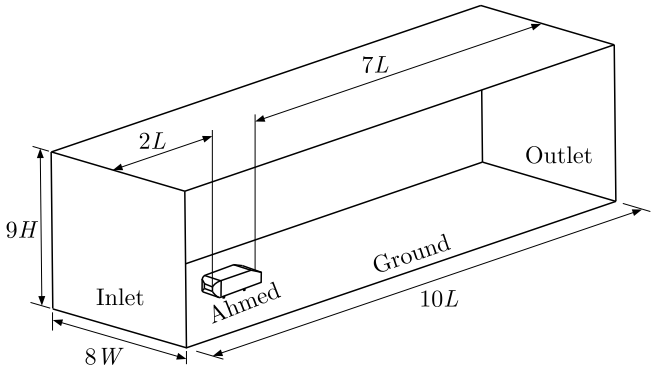


Fig. 3. Illustration of the simulation domain

A multilayer grid is generated to ensure that near wall areas are discretized with a fine grid size, and the outer areas are discretized with a relatively coarse grid size to save computation resources. Specifically, the computational domain is divided into six layers. The coarsest grid size among these layers is 0.0768 m, and is located at the outer areas of the computational domain. The finest grid distributes around the surface of the Ahmed body, its size is 0.0024 m. The total grid volume is about 11 million. Figure 4 shows the grid of the computational domain in $Y = 0$.

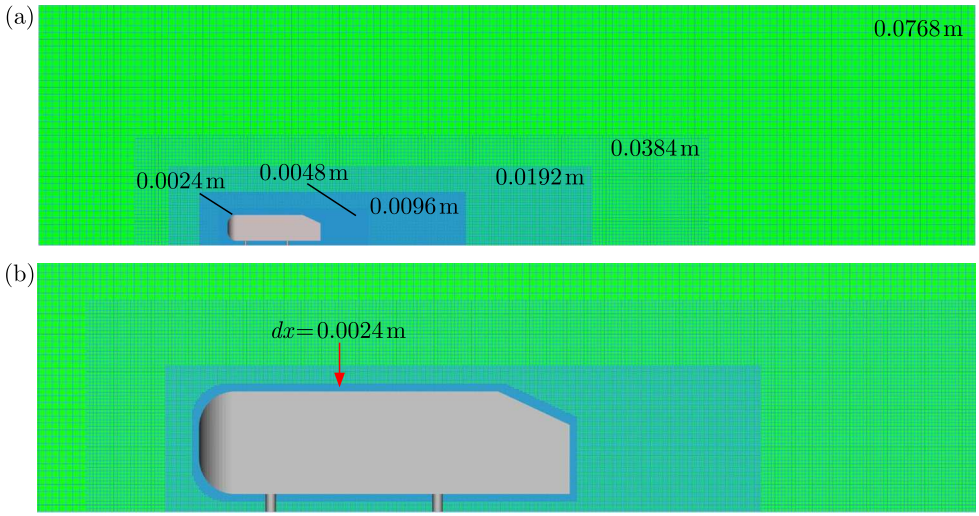


Fig. 4. Grid of the computational domain in $Y = 0$: (a) grid of the domain in $Y = 0$ slice, (b) multilayer grid around the Ahmed body in $Y = 0$ slice

The simulations were conducted on a 128-core CPU (AMD EPYC 7H12 64-Core Processor x2). The initial flow field is set as a uniform flow with a constant velocity and density, except for a box area enclosing the Ahmed body. In the small box around the Ahmed body, the initial flow is set to zero velocity to avoid the transient oscillation inside the Ahmed body. The calculated physical duration is 1 s (the fluid sweeps about 10 body lengths).

4. Results and discussion

4.1. Standard Ahmed body with different rear slant angles

4.1.1. Velocity distribution and pressure distribution

Figure 5 and 6 show the average velocity field, pressure distribution and streamlines of the central section of the model with different rear slant angles $Y = 0$ in the period of 0.75 s-1.0 s, and

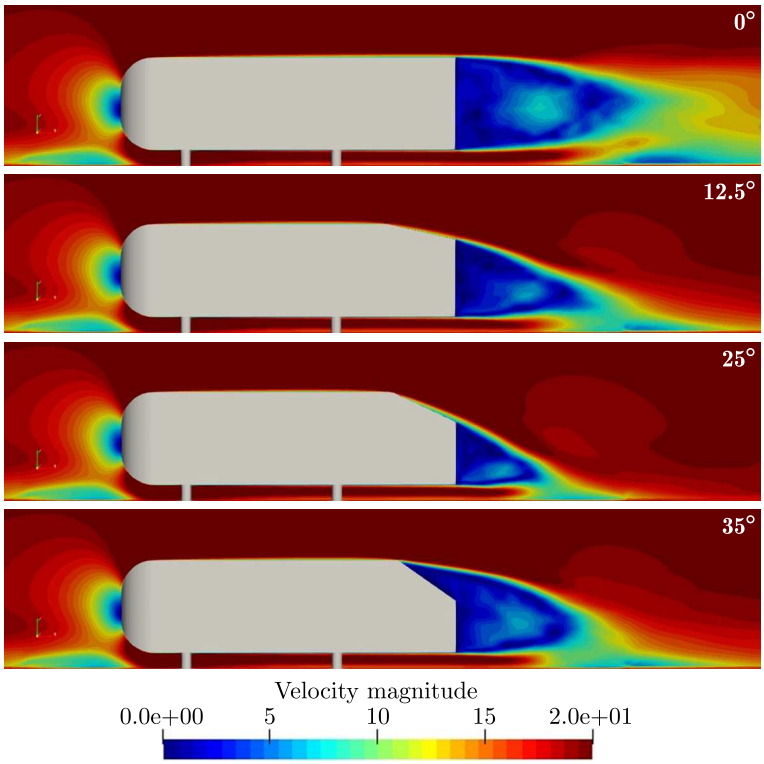


Fig. 5. Time averaged velocity field in $Y = 0$ slice

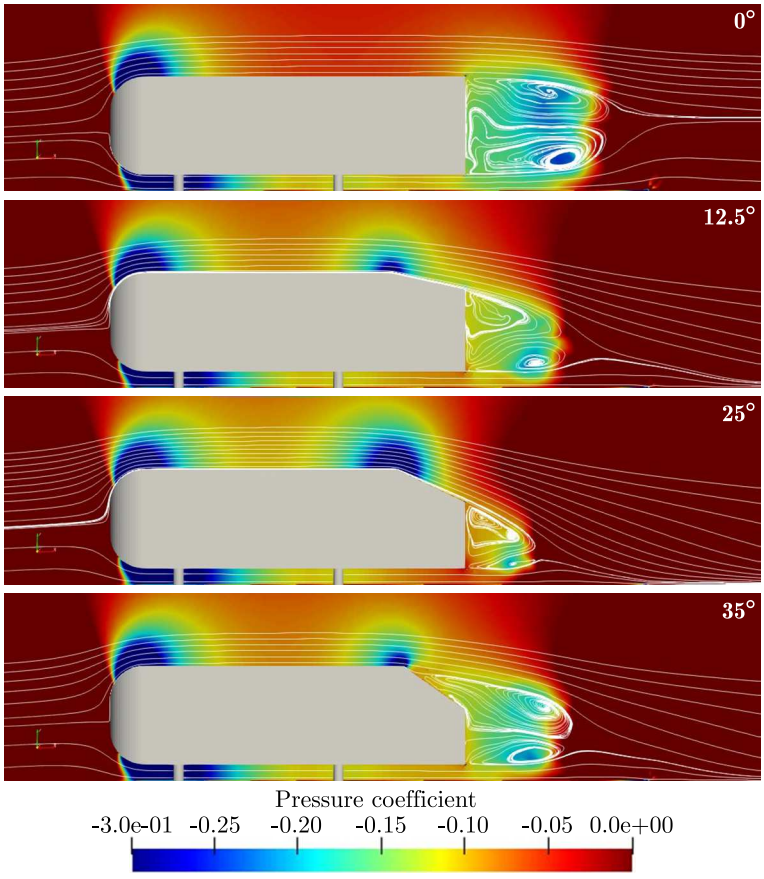


Fig. 6. Pressure coefficient and streamline in $Y = 0$ slice

Fig. 7 shows the iso-surfaces with the total pressure of 0. The velocity and pressure distributions from the head of the model to the rear slant are basically the same in the four simulations. Under the condition of the rear slant angle of 0°, our simulation results show that the rear presents a relatively stable wake vortex far away from the rear part. The flow at a slant angle of 12.5° is similar to the experimental results (Ahmed, *et al.* 1984) and simulation results (Keating *et al.*, 2008): the airflow remains attached at the top corner, and only separates at the end of the rear; the wake vortex at the rear becomes smaller than that of the 0° slant angle. The pressure decreases in a small area near the top corner. As the slant angle increases to 25°, experiments (Bayraktar *et al.*, 2001) and other simulation results (Keogh *et al.*, 2016) show that the airflow separates at the top and then reattaches the surface at about 1/2 way down the slant. A large area of low pressure appears near the top corner of the model. When the slant angle is 35°, the simulation results show that the air flow at the top of the model is cleanly separated, and the wake vortex is far away from the rear.

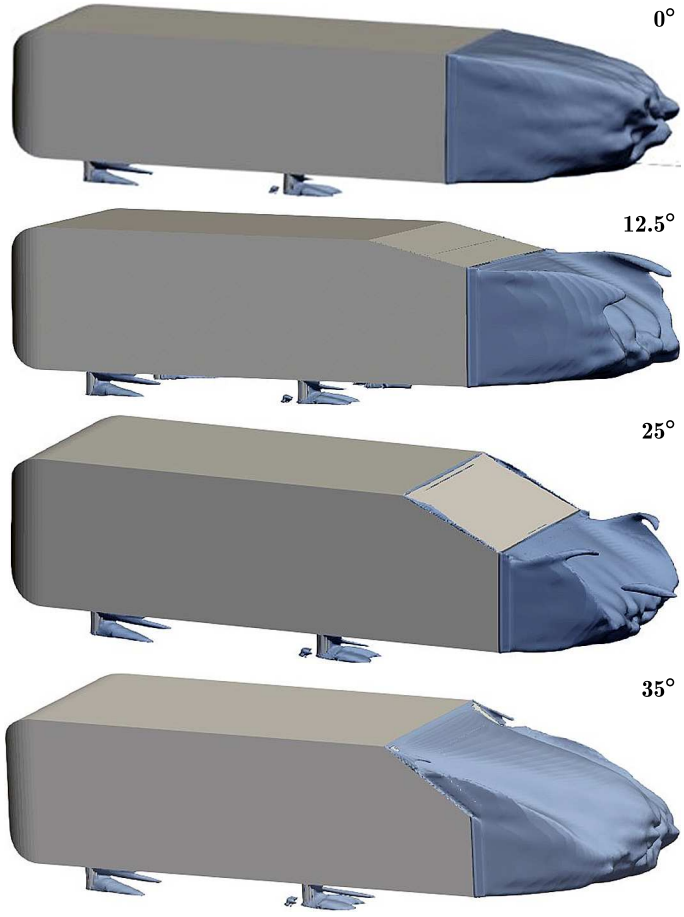


Fig. 7. Zero total pressure contour

4.1.2. Drag coefficient

According to the experimental findings reported by Ahmed *et al.* (1984), it is evident that under identical conditions, the drag coefficient exhibits a distinct trend with respect to the rear slant angle of the Ahmed body. Specifically, as the slant angle increases from a smaller value, the drag coefficient also increases, exhibiting a positive correlation up to a threshold of 30°. However, once the slant angle surpasses this critical point of 30°, the drag coefficient undergoes a sharp decline, plummeting to a significantly lower value. This indicates a significant drop in the drag resistance as the slant angle exceeds 30°. Table 3 lists the drag coefficients of the Ahmed body

with different slant angles obtained by TF-Lattice, PowerFLOW v4.0 (Kotapati *et al.*, 2009) and the experiment conducted by Ahmed *et al.* (1984). These data are also plotted in Fig. 8 for comparison. We can find that the trend of the drag coefficient obtained by TF-Lattice with the slant angle is consistent with the experiment and former simulation, and the error to the experimental results is within 5%.

Table 3. Drag coefficient of Ahmed bodies

Slant angle	0°	12.5°	25°	35°
TF-Lattice C_D	0.247	0.238	0.271	0.249
Ahmed *et al.* (1984) C_D	0.250	0.230	0.285	0.260
PowerFLOW C_D	–	0.237	0.316	0.278

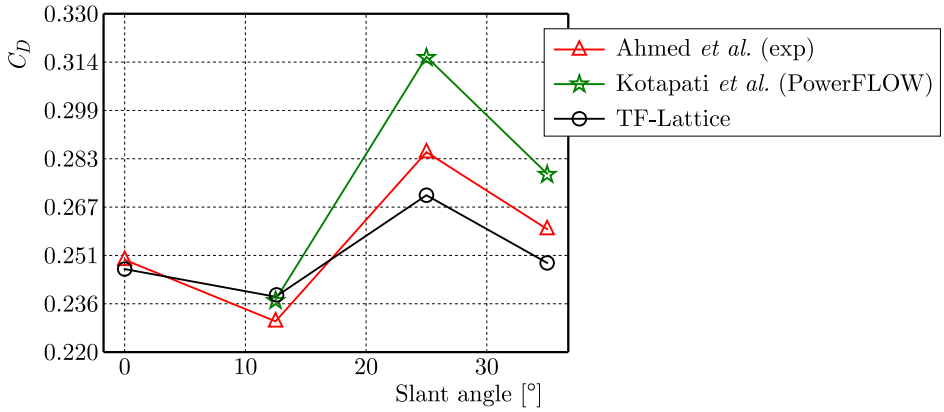


Fig. 8. Drag coefficient of Ahmed bodies with different slant angles

4.2. Non-watertight Ahmed body of 0° rear slant angle

This Section focuses on exploring the aerodynamics of a non-watertight Ahmed body with the aim of identifying the most significant factor that influences its drag coefficient.

4.2.1. Pressure distribution

When simulating the aerodynamics of a non-watertight Ahmed body, a significant alteration is observed in the pressure distribution within the structure. This change in pressure distribution is a notable consequence of the opening in the surface. Figure 9 shows the average pressure distribution and streamline of the $Y = 0$ slice of the Ahmed body in the 0.75 s-1.0 s time periods. The first row in Fig. 9 shows that the internal pressure keeps constant during the simulation for the watertight Ahmed body. It can be seen from the figures that the streamlines enter the interior of the model from the opening. Different opening location leads to different flow patterns inside the model. As the opening exists, the internal pressure drops, even the area ratios of the opening decrease to 1%, the pressure also drops dramatically. In fact, the pressure balance to the adjacent outer flow pressure will establish through the openings.

Besides the pressure drop inside the Ahmed body, the uniformity of the pressure field is obviously different from each other. As the base line simulation, the interior of watertight Ahmed body is very uniform in terms of pressure distribution. We can also find that in the rear 10% and bottom 1% cases the pressure is also uniformm despite the pressure decrease is different in each case. In the other cases, pressure distribution becomes nonuniform. In fact, the pressure

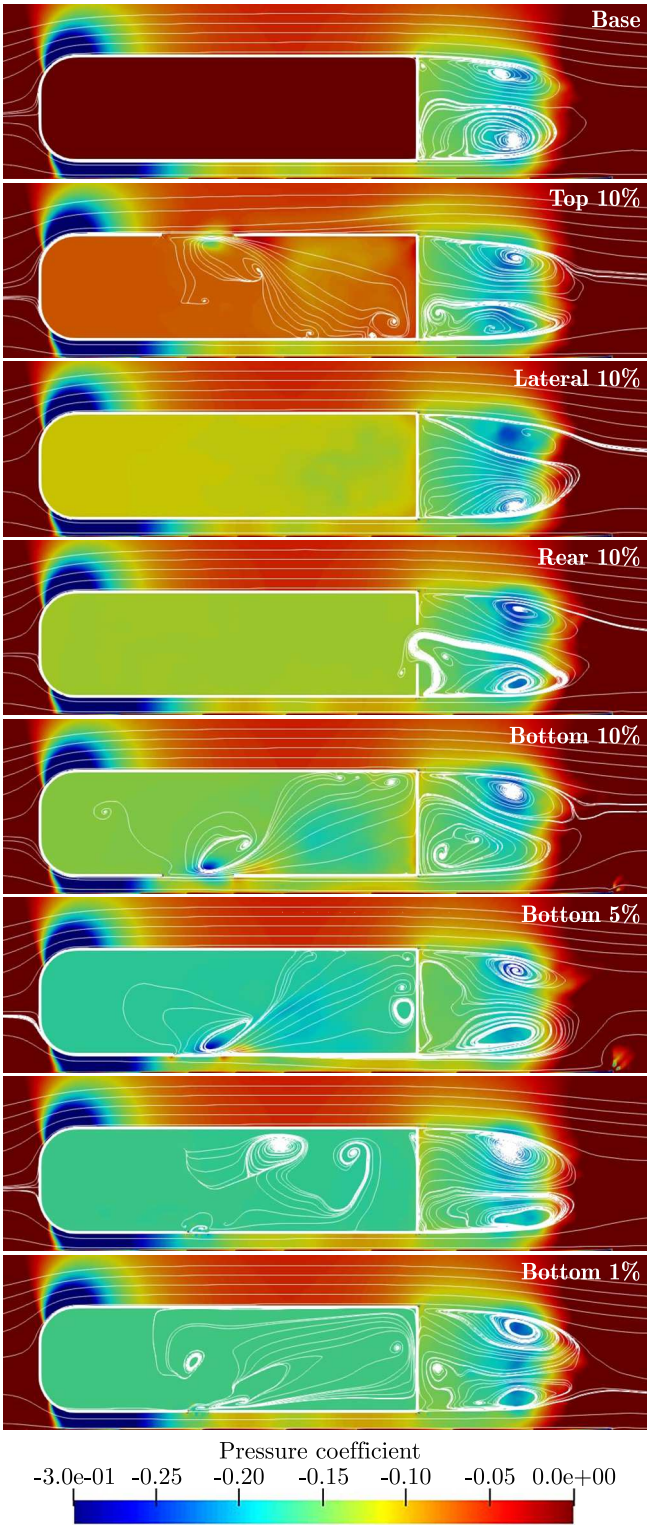


Fig. 9. Pressure coefficient and the streamline in $Y = 0$ slice

distribution is highly dependent on the flow patterns inside the Ahmed body. With large area ratios, the flow passage is large and vortex shedding is strong at the openings. Therefore, the flow patterns become strong and energetic. These phenomena are also reflected in the drag coefficient curves.

4.2.2. Drag coefficient

The pressure distribution inside the Ahmed body will influence the drag coefficient. However, the pressure decrease in the interior is almost irrelevant to the changes of the drag coefficient.

Figure 10 shows the temporal variation curve of the drag coefficient under various geometry configurations (two figures in vertical is a group, the upper one exhibits the complete curve from 0-1.0 s, and the lower one plots the enlarged curve from 0.75 s-1.0 s). To compare the drag coefficients, the range of the vertical axis is set to be of the same scale. The drag coefficient of the watertight Ahmed body fluctuates around the mean value at the beginning of simulation, and after the flow field stabilizes, the variation of the drag coefficient is only ±0.02.

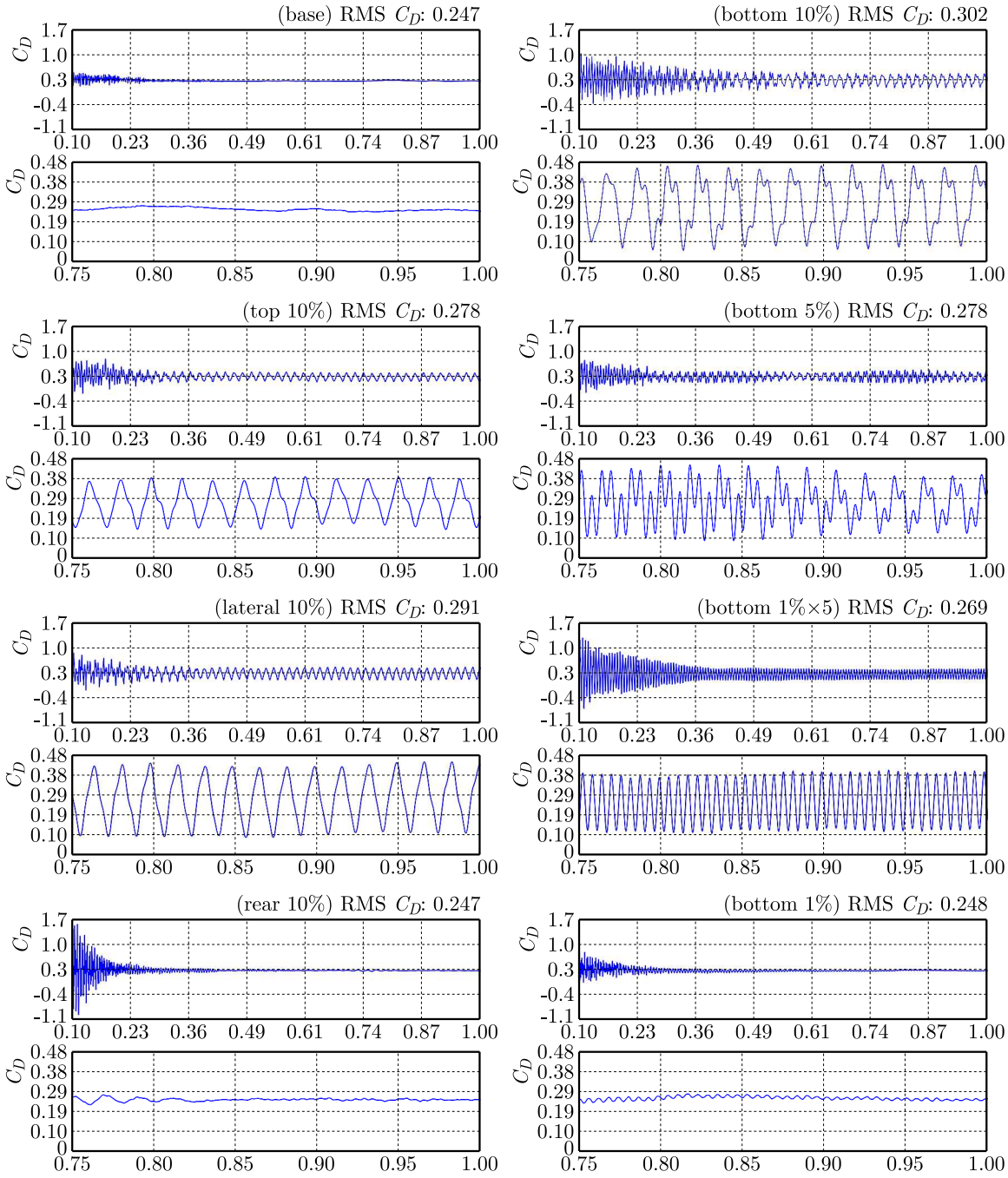


Fig. 10. Evolution of the drag coefficient with time

It worth to note two cases in Fig. 10, rear 10% and bottom 1%, in which the drag coefficient is almost unchanged compared to the base simulation. The pressure decrease in the bottom 1% case is even strongest among the others. In addition, the pressure is uniform inside the Ahmed body for these two cases.

In fact, the uniform pressure decrease inside the geometry will not result in changes in the drag coefficient. Since the pressure is balanced at the opening, and the integral of aerodynamic force in internal surface is identical to the integral in the outer surface of the opening. As illustrated in Fig. 11, the pressure force exerted on the geometry can be calculated as follows

$$F_p = \int\limits_{\Omega_{ex}} -np\, dS + \int\limits_{\Omega_{in}} np\, dS \tag{4.1}$$

Considering the pressure balance at the opening area and Gauss formula, we have

$$\begin{aligned} 0 &= \int\limits_{\Omega_{o,in}} -n_o p\, dS + \int\limits_{\Omega_{o,ex}} n_o p\, dS \\ 0 &= \int\limits_{\Omega_{in}} -np\, dS + \int\limits_{\Omega_{o,in}} -n_o p\, dS \end{aligned} \tag{4.2}$$

Therefore

$$F_p = \int\limits_{\Omega_{ex}} np\, dS + \int\limits_{\Omega_{o,ex}} n_o p\, dS \tag{4.3}$$

It means the pressure force exerted on the non-watertight geometry is identical to the pressure force exerted on the watertight geometry if the external flow is identical. Since the friction force generally is small and the opening area could be small too, we neglect the friction force in this discussion. Accordingly, the discussion for the pressure force is extended to the total aerodynamic force hereafter.

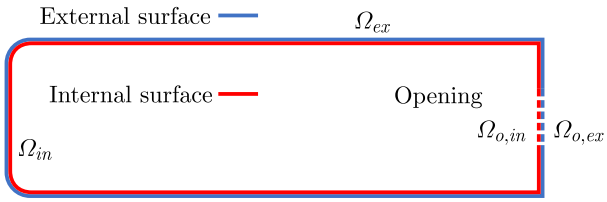


Fig. 11. The surface of the non-watertight model

How small the opening area ratio needs to be is dependent on its specific location in the surface. As four cases with the 10% opening area ratio demonstrate, 10% is small enough when the opening locates at the rear part of the Ahmed body. It is because that the flow at the rear is slow and with moderate pressure variation. Although the opening area seems large, the flow keeps almost quiescent inside the Ahmed body. On the contrary, the other three openings locate at top, bottom and side, where the stream is very strong. Consequently, the massive flow passes through the opening, disturbs the internal pressure distribution and external main flow, and leads to a change in the drag coefficient.

Figure 12 compares the average drag coefficient (0.5 -1.0 s) under various geometrical defects, and the specific values are listed in Table 4. In the case of the same opening area, the bottom opening has the most obvious effect. And when the area at the bottom is smaller, the drag coefficient is lower. With the 5% area ratio at the bottom, if the opening area are scattered

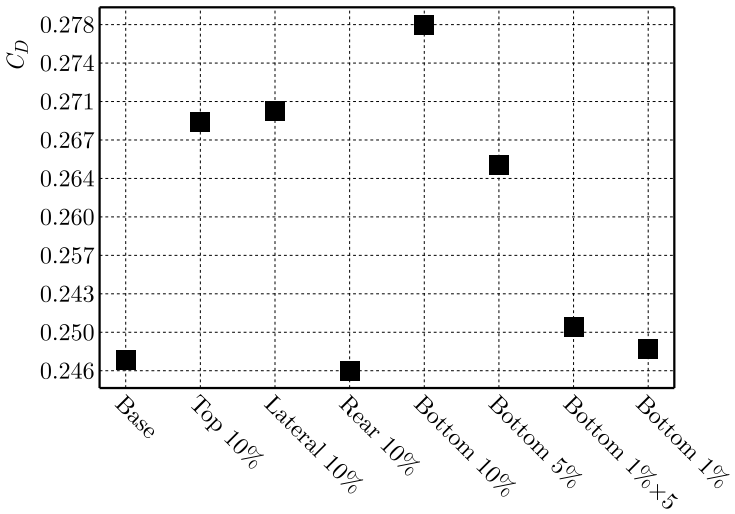


Fig. 12. Drag coefficient of different non-watertight models

Table 4. Drag coefficient of different cases

Case	Base	Top 10%	Lateral 10%	Rear 10%
C_D	0.247	0.278	0.291	0.247

Case	Bottom			
	10%	10%	1% × 5	1%
C_D	0.302	0.278	0.269	0.248

(bottom $1\% \times 5$) at the bottom, the drag coefficient will be lower than that of a single opening (bottom 5%).

Based on the previous discussion, it can be concluded that the increase of the drag coefficient of the Ahmed body is inversely correlated to the uniformity of pressure distribution within it. Specifically, when the pressure distribution inside the Ahmed body is more uniform, the drag coefficient experiences lesser variations. Conversely, as the internal flow disturbance increases and the pressure distribution becomes more uneven, the drag coefficient rises significantly. This phenomenon underscores the importance of optimizing the pressure distribution and minimizing flow disturbances within the Ahmed body to achieve lower drag coefficients.

5. Conclusion

This study first validated the TF Lattice's capability in simulating the drag coefficient of the standard Ahmed body through comparison with the established experimental data. For slant angles of 0°, 12.5°, 25° and 35°, the simulation results of TF Lattice are in good agreement with the literature results. Especially, it is almost identical to the experimental results when the back angle is 0°.

After that, to examine the influence of geometric defects on the drag coefficient, non--watertight Ahmed bodies were created by introducing square holes into the standard Ahmed body with a rear slant angle of 0°, which were then used in further case study simulations.

Based on the simulation results, the geometrical defects have great influence on the drag coefficient. The influence path is identified as follows: Initially, defects or openings in the Ahmed body create a connection between its exterior and interior, allowing for pressure equilibration at these openings. Consequently, this leads to a reduction in the internal pressure. It is important

to note that a uniform decrease in pressure does not alter the drag coefficient. It is only when the opening is significant or the external flow stream is intense so non-uniformities arise in the pressure field, ultimately causing a rise in the drag coefficient.

These phenomena imply that the location of the opening is an important factor that determines how crucial is the opening area effect on the drag coefficient. In general, the larger the opening area is, the greater change of the drag coefficient will be. Meanwhile, provided that the opening area is enough small, the non-watertight geometry can reproduce the drag coefficient derived by watertight geometry.

Non-watertight geometry is natural and ubiquitous in engineering. In the past, most CFD simulations need preprocessing to enclose the geometry which is time consuming and expensive. The results and discussion in this study show another promising way which directly solves flow problems on non-watertight geometry, and which deserves more attention and further investigation.

Appendix

A.1. Flow solver

We employed a newly emerging lattice Boltzmann method (LBM) based software, TF-Lattice V1.0, as the flow solver in this study. LBM is very suitable for solving unsteady problems, such as high Reynolds number flows. Commercial CFD software based on LBM, including PowerFLOW (Dassault) and Ultra Fluid X (Altair), has become an important tool in the automobile industry. The TF-Lattice is another LBM based CFD software developed by Tenfeng Technology in China. It utilizes a multi-layer Cartesian grid to discretize the space, and a cutting cell technique to handle an arbitrary curved surface. Therefore, the TF-Lattice is capable of simulating on a non-watertight geometry with various kinds of geometric defects, such as punctures, small cracks, overlapping surfaces, etc., and can be used directly in simulation. This feature is a perfect match for our researches.

A.2. Lattice-Boltzmann method

For the sake of completeness, the numerical method in TF-Lattice is briefly introduced. LBM is an explicit mesoscopic method that simulates the fluid motion by imitating particle stream and collision. The evolution equation of the lattice Boltzmann method is actually a special discrete form of the kinetic equation in a specific velocity space (He and Luo, 1997)

$$f_i(x + c_i\,dt, t + \Delta t) - f_i(x,t) = -\frac{1}{\tau}[f_i(x,t) - f_i^{eq}(x,t)] \tag{A.1}$$

where f is the velocity distribution function which depends on location x, time t and particle discrete velocity c_i. The subscript i represents a group of discrete velocity directions. Common discrete velocity models include D2Q7, D2Q9, D3Q15, and D3Q19 (Qian *et al.*, 1992). The D3Q19 discrete velocity model is employed in TF-Lattice.

The right-hand side of Eq. (A.1) is solved by a collision model called Hybrid Recursive Regularization (Jacob *et al.*, 2018). This model combines the standard BGK (Bhatnaga *et al.*, 1954) single relaxation, regularization procedure and a distribution function reconstructed from finite difference scheme together. This model can effectively suppress non-hydrodynamic modes generated by LBM in high Reynolds number flows (Jacob *et al.*, 2018; Malaspinas, 2015). In TF-Lattice, the hybrid part of HRR is replaced by a reconstructed distribution function invented by Tenfong Technology. For more information about the LBM, please refer to the special articles (Qian *et al.*, 1992; Chen and Doolen, 1998; Jacob *et al.*, 2018) and references therein.

A.3. Validation

The software, TF-Lattice, was validated by many test cases. Most of the following testing results are excerpted from the internal technical report of TF-Lattice software.

A.3.1. Laminar flow past sphere

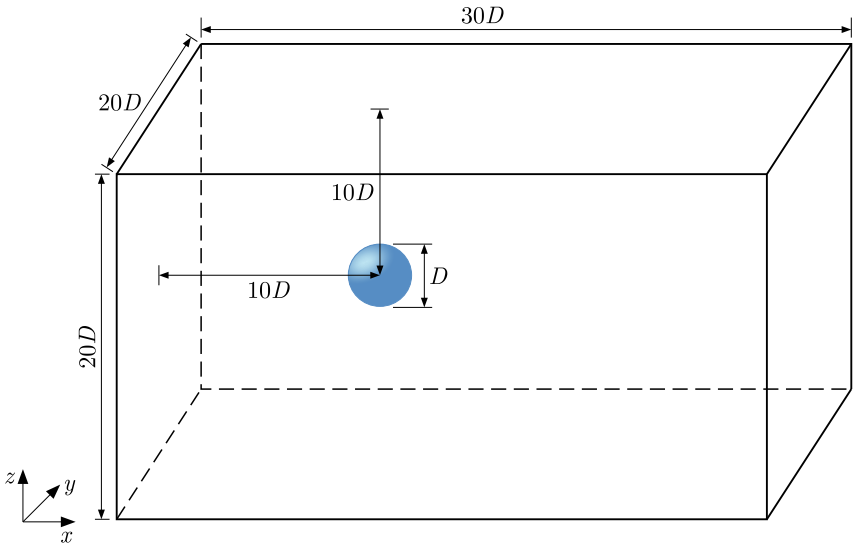


Fig. 13. Configuration of the flow past sphere

Table 5. Drag coefficient of the sphere at different Reynolds numbers

Re	TF--Lattice	Schiller and Naumann (1933)	Johnson and Patel (1999)	Nagata *et al.* (2016)	Sheard *et al.* (2005)	Average	Error
50	1.5774	1.5381	1.58	1.54	1.59	1.5620	0.99%
100	1.0918	1.0917	1.09	1.07	1.09	1.0854	0.58%
200	0.7915	0.8056	0.78	0.76	0.77	0.7789	1.62%

A.3.2. Laminar flow past sphere

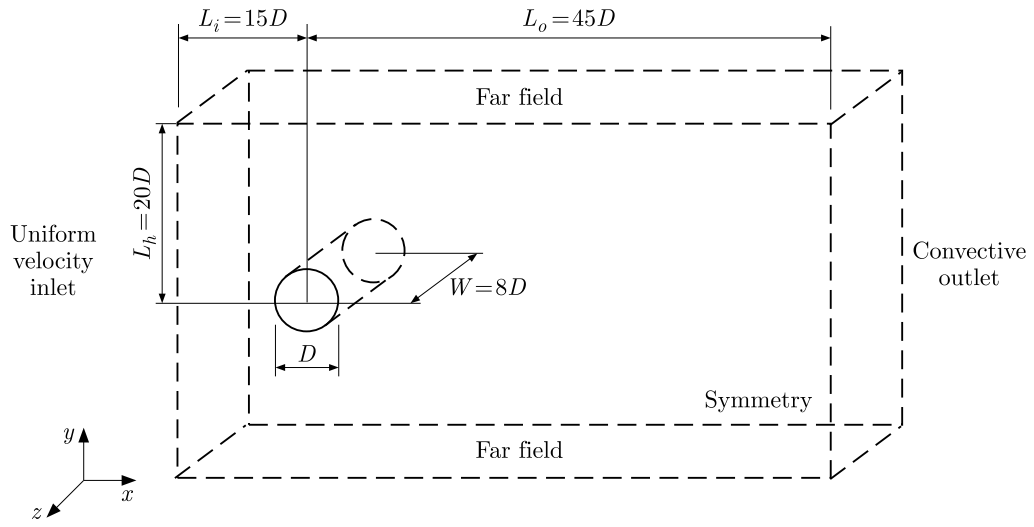


Fig. 14. Configuration of a flow past cylinder

Table 6. Drag coefficient of the cylinder at different Reynolds numbers

Drag coefficient	Reynolds number				
	20	40	60	80	100
Takami & Keller	2.00	1.54	1.32	–	–
S.C.R. Dennis	2.05	1.52	–	–	1.06
Nieuwstadt & Keller	2.05	1.55	–	–	–
M. Nishioka	1.94	1.47	1.31	–	–
R. Franke	–	–	1.35	1.35	–
J. Park	2.01	1.51	1.39	1.35	1.33
TF-Lattice	2.09	1.57	1.36	1.32	1.29

A.3.3. Flow past sphere at Re = 10000

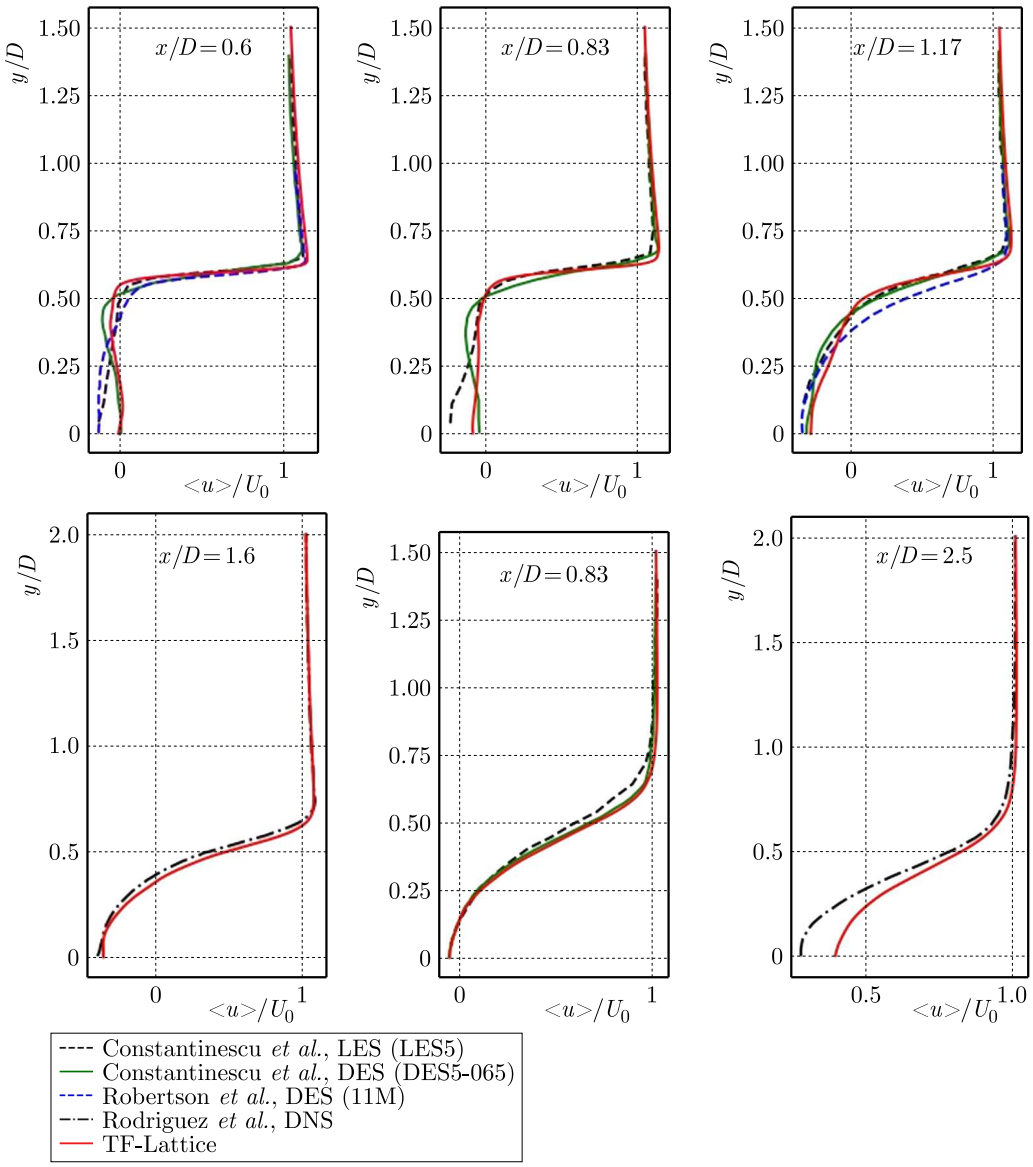


Fig. 15. Average velocity profiles of the fluid flow past a sphere (Re = 10000) at varying distances from the sphere rear: $x = 0.6D$, $0.83D$, $1.17D$, $1.6D$, $2.0D$, $2.5D$

Table 7. Drag coefficient of the cylinder at Reynolds number 10000

Reference example	Method	C_D	Lr/D
Achenbach	Experiment	0.4	–
Ozgoren	Experiment	–	1.4
Rodriguez	DNS	0.402	1.657
Yun	LES	0.393	1.364
Const. & Squires	LES, DES	0.393, 0.397	1.55, 1.48
Robertson	DES	0.393	–
Stiebler	LES (LBM-MRT)	0.443	–
Eitel-Amor	LES (LBM-MRT)	0.392	1.788
Geier	ILES (LBM-Cum.)	0.430	–
TF-Lattice	ILES (LBM-iHRR)	0.397	1.539

A.3.4. DrivAer tests

DrivAer (Heft *et al.*, 2012) is a hybrid model, combining features from both the Audi A4 and BMW 3 Series. This model was designed to advance numerical calculations and experimental methods in vehicle aerodynamics. TF-Lattice uses 8 layers and a total of 22 million grid points to simulate the Notchback and Estateback cases.

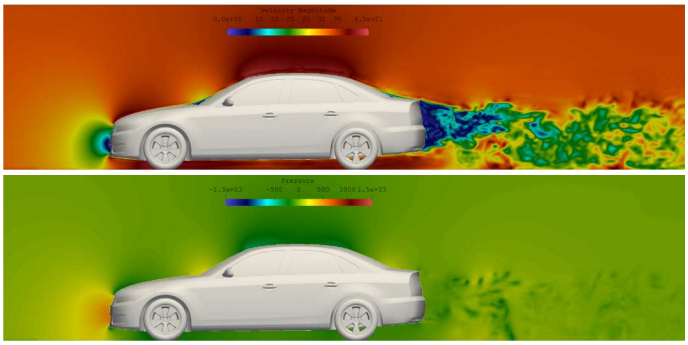


Fig. 16. Instantaneous flow field of the Notchback case (NSwMwWwGS, Heft *et al.*, 2012). The upper figure presents the velocity magnitude field, the lower figure presents the pressure flied

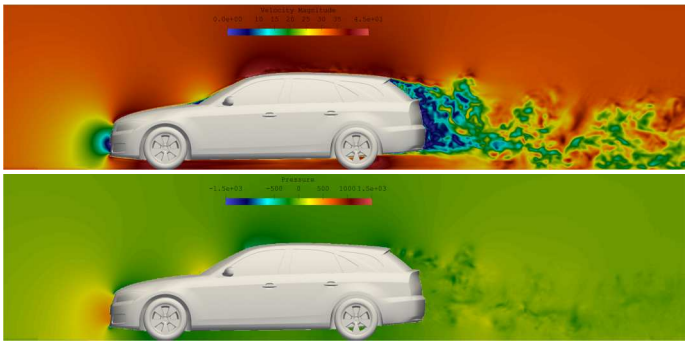


Fig. 17. Instantaneous flow field of the Estateback case (ESwMwWwGS, Heft *et al.*, 2012). The upper figure presents the velocity magnitude field, the lower figure presents the pressure flied

In addition to the exhibited test cases, TF-Lattice has conducted a large number of tests to validate and verify its accuracy and reliability.

Table 8. Drag coefficient of DrivAer

C_D	TF-Lattice	Experiment	Error
NSwMwWwGS	0.2386	0.246	3.0%
ESwMwWwGS	0.2939	0.292	0.7%

Acknowledgement

Special thanks to Tianjin Key Laboratory of Automotive Wind Tunnel Testing and Application and Tenfong Technology for their support and services.

References

1. Ahmed S.R., Ramm G., Faltin G., 1984, Some salient features of the time-averaged ground vehicle wake, *SAE Technical Paper Series*, DOI: 10.4271/840300
2. Bayraktar I., Landman D., Baysal O., 2001, Experimental and computational investigation of Ahmed body for ground vehicle aerodynamics, *SAE Technical Paper Series*, DOI: 10.4271/2001--01-2742
3. Bhatnagar P.L., Gross E.P., Krook M., 1954, A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems, *Physical Review*, **94**, 3, 511-525
4. Chen S., Doolen G.D., 1998, Lattice Boltzmann Method for fluid flows, *Annual Review of Fluid Mechanics*, **30**, 329-364
5. ERCOFTAC Knowledge Base Wiki, page "Ahmed body", 2023, https://kbwiki.ercoftac.org/w/index.php/Abstr:Ahmed_body
6. Fares E., 2006, Unsteady flow simulation of the Ahmed reference body using a lattice Boltzmann approach, *Computers and Fluids*, **35**, 8-9, 940-950
7. Gilliéron P., Chometon F., 1999, Modelling of stationary three-dimensional separated air flows around an Ahmed reference model, *ESAIM: Proceedings*, **7**, 173-182
8. Han T., 1989, Computational analysis of three-dimensional turbulent flow around a bluff body in ground proximity, *AIAA Journal*, **27**, 9, 1213-1219
9. He X., Luo L.-S., 1997, Theory of the lattice Boltzmann method: from the Boltzmann equation to the lattice Boltzmann equation, *Physical Review E*, **56**, 6, 6811
10. Heft A.I., Indinger T., Adams N.A., 2012, Introduction of a new realistic generic car model for aerodynamic investigations, *SAE Technical Paper Series*, DOI: 10.4271/2012-01-0168
11. Hinterberger C., García-Villalba M., Rodi W., 2004, Large eddy simulation of flow around the Ahmed body, [In:] *The Aerodynamics of Heavy Vehicles: Trucks, Buses, and Trains, Conference Proceedings*, McCallen R., Browand F., Ross J. (Eds.), *Lecture Notes in Applied and Computational Mechanics*, **19**, Springer, 77-87
12. Holman D.M., Brionnaud R.M., Abiza Z., 2012, Solution to industry benchmark problems with the Lattice-Boltzmann code Xflow, *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012)*, Eberhardsteiner J. *et al.* (Eds.), Vienna, Austria
13. Jacob J., Malaspinas O., Sagaut P., 2018, A new hybrid recursive regularised Bhatnagar--Gross-Krook collision model for lattice Boltzmann method-based large eddy simulation, *Journal of Turbulence*, **19**, 11-12, 1051-1076
14. Johnson T.A., Patel V.C., 1999, Flow past a sphere up to a Reynolds number of 300, *Journal of Fluid Mechanics*, **1999**, 378, 19-70
15. Keating A., Shock R., Chen H., 2008, Lattice Boltzmann simulations of the unsteady flow behind the Ahmed body, *SAE Technical Paper Series*, DOI: 10.4271/2008-01-0740

16. Keogh J., Barber T., Diasinos S., Doig G., 2016, The aerodynamic effects on a cornering Ahmed body, *Journal of Wind Engineering and Industrial Aerodynamics*, **154**, 34-46

17. Kotapati R., Keating A., Kandasamy S., Duncan B., Shock R., Chen H., 2009, The Lattice-Boltzmann-VLES method for automotive fluid dynamics simulation, *Review, SAE Technical Paper Series*, DOI: 10.4271/2009-26-0057

18. Lienhart H., Becker S., 2003, Flow and turbulence structure in the wake of a simplified car model, *SAE Technical Paper Series*, DOI: 10.4271/2003-01-0656

19. Malaspinas O., 2015, Increasing stability and accuracy of the lattice Boltzmann scheme: recursivity and regularization, *ArXiv preprint*, 1505.06900

20. Nagata T., Nonomura T., Takahashi S., Mizuno Y., Fukuda K., 2016, Investigation on subsonic to supersonic flow around a sphere at low Reynolds number of between 50 and 300 by direct numerical simulation, *Physics of Fluids*, **28**, 5, 056101

21. Qian Y.-H., d'Humières D., Lallemand P., 1992, Lattice BGK models for Navier-Stokes equation, *Europhysics Letters*, **17**, 6, 479-484

22. Schiller L., Naumann A.Z., 1933, Uber die grundlegenden Berechnungen bei der Schwerkraftaufbereitung, *Vereines Deutscher Ingenieure*, **77**, 318-320

23. Sheard G.J., Hourigan K., Thompson M.C., 2005, Computations of the drag coefficients for low-Reynolds-number flow past rings, *Journal of Fluid Mechanics*, **526**, 257-275

24. Strachan R.K., Knowles K., Lawson N.J., 2007, The vortex structure behind an Ahmed reference model in the presence of a moving ground plane, *Experiments in Fluids*, **42**, 5, 659-669

Manuscript received January 5, 2024; accepted for publication October 22, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 787-798, Warsaw 2024
https://doi.org/10.15632/jtam-pl/194401

DYNAMIC CALCULATION OF THE FIRE ZONE FOR ANTI-AIRCRAFT ARTILLERY

Leszek Baranowski

Military University of Technology, Faculty of Mechatronics, Armament and Aeropace, Warsaw, Poland

e-mail: leszek.baranowski@wat.edu.pl

Przemysław Majewski, Jacek Szymonik

PIT-RADWAR S.A., Warsaw, Poland

e-mail: przemek.majewski@gmail.com; szymonik.jacek@gmail.com

The article presents a fast algorithm for target hit probability calculation in real time for the use in modern fire control systems (FCS). The idea behind this algorithm is to calculate the hit probability using the information about the projectile motion and the estimated motion of the tracked target. The first part of the article describes the analytical solution to the problem of hit probability calculation. The results obtained from the analytical method are then compared with a simulation method developed specifically for the analytical method verification. This comparison led to the conclusion that the presented analytical method is suitable for the use in modern FCS.

Keywords: exterior ballistics, equations of motion, projectile trajectory

1. Introduction

The analysis of requirements posed by modern fire control systems showed the need to design and implement an algorithm that enables the target hit probability calculation in real time. This is a big challenge considering the complexity of algorithms required to calculate such probability. Those will include: processing of data coming from sensors (radars, cameras, lidars etc.), estimating motion parameters of the target, calculating the trajectory of the projectile. In the literature, one can find different approaches to hit probability definitions and calculation methods. In (Kang *et al.*, 2016), the hit probability prediction for the anti-aircraft artillery is based on errors of muzzle velocity and the cant error related to the fire power. Liu and Shi (2022) presented a method that extended the notion of hit probability to calculstion of different damage levels by incorporating the Bayesian inference method. Different approaches to artillery effectiveness are discussed in Katsev (2018), where authors defined it as a function of dependent and independent errors related to weapon-target interaction and how much would be artillery shell lethality. They showed an "offline" method to estimate effectiveness based on number of rounds, variations of random errors, dispersion errors, target lethal area. A similar approach was presented in (Obradović *et al.*, 2023), where authors also included errors in meteorological conditions preparation. Our innovative method proposes an analytical solution for real time calculation of hit probabilities based on both the projectile motion model and estimated parameters of the tracked motion. In order to verify the correctness of the algorithm, we also prepared an application that enabled simulation of the projectile flight and target movement, taking into account possible disturbances that might occur, such as: dispersion of the initial velocity of the projectile, deviations of projectile parameters from nominal values, cannon barrel jerk, etc.

2. Analytical method for target hit probability calculation

The analytical method for target hit probability calculations essentially boils down to four main steps (assuming that coordinates of the hit point and time to hit t_{hit} are already calculated):

1. Calculate how the perturbations of initial conditions of differential equations describing projectile motion propagate along its trajectory. In order to do that, one needs to solve variational equation (2.15) which is presented in Subsection 2.1. The system of equations describing projectile motion, Jacobi matrix for the system (which is necessary to solve the variational equation) and a numerical method (Runge-Kutta) for solving the system of differential equations is presented in Subsection 2.2.
2. Having the solution for the variational equation and the covariance matrix with initial variances for the system of equations, one can apply a rule of error propagation in order to find the covariance matrix at a specified time – Eq. (2.23).
3. The last element needed for hit probability calculation is the covariance matrix related to the target motion parameters. The initial covariance matrix is based on the Kalman filter output for the chosen motion model of the target. The value of the covariance matrix at time t_{hit} is calculated, as in the previous point, by applying the rule of error propagation – Eq. (2.24).
4. Having covariance matrices for the projectile and the target, calculate the hit probability using the Monte Carlo method described in Subsection 2.3.

The steps briefly listed above are described in more details in the following subsections.

2.1. Variational equation and error propagation

The variational equation describes how the disturbances of the initial conditions of the system will evolve along its trajectory[1]. There are several mathematical models commonly used to describe the motion of a projectile: the point-mass trajectory model (McCoy, 1999, the rigid body trajectory model (Baranowski, 2013a), the modified point mass trajectory model (Baranowski, 2013a).

Given any system of equations $\dot{\mathbf{y}}(t) = \mathbf{F}(\mathbf{y}(t), t)$ with the initial conditions $\mathbf{y}(t_0)$, one can easily generate a trajectory using numerical methods for iterative solution of ordinary differential equations (e.g. Runge-Kutta method). Te main goal in this part will be to determine how the solution of a system of equations behaves when a small disturbance $\boldsymbol{\delta}$ is introduced into the vector of initial conditions $\mathbf{y}(t_0)$.

Let

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{y}(t) = \mathbf{F}(\mathbf{y}(t), t) \tag{2.1}$$

be a system of n-th order differential equations with an initial condition $\mathbf{y}(t_0) = \mathbf{y}_0$ and a solution $\boldsymbol{\Phi}_t(\mathbf{y}_0, t_0)$, i.e. (Parker and Chua, 1989)

$$\dot{\boldsymbol{\Phi}}_t(t_0, \mathbf{y}_0) = \mathbf{F}(\boldsymbol{\Phi}_t(t_0, \mathbf{y}_0), t) \qquad \boldsymbol{\Phi}_{t_0}(t_0, \mathbf{y}_0) = \mathbf{y}_0 \tag{2.2}$$

Moreover, let

$$\boldsymbol{\delta} = \boldsymbol{\Phi}_{t_0}(t_0, \mathbf{y}_0) - \boldsymbol{\Phi}_{t_0}(t_0, \mathbf{y}_0 + \boldsymbol{\delta}) \tag{2.3}$$

be a disturbance of the initial conditions of the system and

$$\Delta\boldsymbol{\Phi}_t(t_0, \mathbf{y}_0) = \boldsymbol{\Phi}_t(t_0, \mathbf{y}_0) - \boldsymbol{\Phi}_t(t_0, \mathbf{y}_0 + \boldsymbol{\delta}) \tag{2.4}$$

[1] A problem often used when studying the stability of dynamical systems (Jordan and Smith, 2007)

be a result of the disturbance propagation after time t. Using the Taylor series expansion with the simultaneous omission of terms of higher orders (thus only the linear terms of the expansion are taken into account), one can write (2.4) as a function of the original disturbance

$$\mathbf{\Phi}_t(t_0, \mathbf{y}_0 + \boldsymbol{\delta}) = \mathbf{\Phi}_t(t_0, \mathbf{y}_0) + \frac{\partial \mathbf{\Phi}_t(t_0, \mathbf{y}_0)}{\partial \mathbf{y}_0}\boldsymbol{\delta} \tag{2.5}$$

and thus

$$\mathbf{\Phi}_t(t_0, \mathbf{y}_0 + \boldsymbol{\delta}) - \mathbf{\Phi}_t(t_0, \mathbf{y}_0) = \frac{\partial \mathbf{\Phi}_t(t_0, \mathbf{y}_0)}{\partial \mathbf{y}_0}\,\boldsymbol{\delta}. \tag{2.6}$$

Using definitions (2.3) and (2.4)

$$\Delta\mathbf{\Phi}_t(t_0, \mathbf{y}_0) = \frac{\partial \mathbf{\Phi}_t(t_0, \mathbf{y}_0)}{\partial \mathbf{y}_0}\Delta\mathbf{\Phi}_{t_0}(t_0, \mathbf{y}_0) \tag{2.7}$$

Resulting equation (2.7) is called the variational equation. In order to investigate the propagation of disturbances, it is necessary to find a part of the equation

$$\frac{\partial \mathbf{\Phi}_t(t_0, \mathbf{y}_0)}{\partial \mathbf{y}_0} \equiv \nabla_{\mathbf{y}_0}\mathbf{\Phi}_t(t_0, \mathbf{y}_0) \tag{2.8}$$

By integrating equation (**??**) an integral equation will be obtained

$$\mathbf{\Phi}_t(t_0, \mathbf{y}_0) = \mathbf{\Phi}_{t_0}(t_0, \mathbf{y}_0) + \int_{t_0}^{t} \mathbf{F}(\mathbf{\Phi}_\tau(\tau_0, \mathbf{y}_0), \tau)\, d\tau \tag{2.9}$$

Differentiating with respect to initial values $\mathbf{y}_0$, we obtained

$$\nabla_{\mathbf{y}_0}\mathbf{\Phi}_t(t_0, \mathbf{y}_0) = \frac{\partial \mathbf{\Phi}_{t_0}(t_0, \mathbf{y}_0)}{\partial\, \mathbf{y}_0} + \int_{t_0}^{t} \frac{\partial \mathbf{F}(\mathbf{\Phi}_\tau(\tau_0, \mathbf{y}_0), \tau)}{\partial\, \mathbf{y}_0}\, d\tau \tag{2.10}$$

Using the chain rule, and bearing in mind that

$$\frac{\partial \mathbf{\Phi}_{t_0}(t_0, \mathbf{y}_0)}{\partial \mathbf{y}_0} = \mathbb{I}_n \tag{2.11}$$

we obtained

$$\nabla_{\mathbf{y}_0}\, \mathbf{\Phi}_t(t_0, \mathbf{y}_0) = \mathbb{I}_n + \int_{t_0}^{t} \frac{\partial \mathbf{F}(\mathbf{\Phi}_\tau(\tau_0, \mathbf{y}_0), \tau)}{\partial \mathbf{\Phi}_\tau(t_0, \mathbf{y}_0)} \frac{\partial \mathbf{\Phi}_\tau(t_0, \mathbf{y}_0)}{\partial\, \mathbf{y}_0}\, d\tau \tag{2.12}$$

By introducing the following notation

$$\mathbf{K}(t) = \nabla_{\mathbf{y}_0}\, \mathbf{\Phi}_t(t_0, \mathbf{y}_0) \qquad \mathbf{J}(t) = \frac{\partial \mathbf{F}(\mathbf{\Phi}_t(t_0, \mathbf{y}_0), t)}{\partial \mathbf{\Phi}_t(t_0, \mathbf{y}_0)} \tag{2.13}$$

and then returning to the differential form of Eq. (2.12), we finally got

$$\frac{\mathrm{d}}{\mathrm{dt}} \left(\nabla_{\mathbf{y}_0}\mathbf{\Phi}_t(t_0, \mathbf{y}_0)\right) = \frac{\partial \mathbf{F}(\mathbf{\Phi}_t(t_0, \mathbf{y}_0), t)}{\partial \mathbf{\Phi}_t(t_0, \mathbf{y}_0)} \frac{\partial \mathbf{\Phi}_t(t_0, \mathbf{y}_0)}{\partial\, \mathbf{y}_0} \tag{2.14}$$

then

$$\dot{\mathbf{K}}(t) = \mathbf{J}(t)\mathbf{K}(t) \qquad \mathbf{K}(t_0) = \mathbb{I}_n \tag{2.15}$$

here $\mathbf{J}(t)$ is a Jacobi matrix for the system of differential equations (2.3) at time t.

2.2. Error propagation for the M-model

The considerations presented in the previous subsection will be used to investigate the propagation of disturbances in the initial conditions for the explicit form of the modified point mass trajectory model, in (Baranowski, 2016b) called the M-model. This model is equivalent to the MPMTM model presented in (McCoy, 1999) and (STANG 4355, 2009) as demonstrated in (Baranowski *et al.*, 2016b)

$$\begin{aligned}
&\dot{x} = v_x + w_x \qquad \dot{y} = v_y + w_y \qquad \dot{h} = v_h + w_h \\
&\dot{p} = \frac{\rho v^2}{2I_x} S d C_{spin} \widehat{p} \qquad S = \frac{\pi d^2}{4} \\
&\dot{v} = -\frac{\rho v^2}{2m} S \left(C_{D_0} + \widehat{C}_{D_\alpha^2} \Big(\frac{2mg}{\rho v^2 S} \Big)^2 \frac{\widehat{I}_x^2 \widehat{p}^2 \cos^2 \gamma_a}{(1 - \widehat{I}_x \widehat{p}^2 \widehat{C}_{mag-f})^2 + (\widehat{I}_x \widehat{p} \widehat{C}_{L\alpha})^2} \right) - g \sin \gamma_a \\
&\begin{bmatrix} \dot{\gamma}_a \\ \dot{\chi}_a \cos \gamma_a \end{bmatrix} = -\frac{g}{v} \frac{\cos \gamma_a}{(1 - \widehat{I}_x \widehat{p}^2 \widehat{C}_{mag-f})^2 + (\widehat{I}_x \widehat{p} \widehat{C}_{L\alpha})^2} \begin{bmatrix} 1 - \widehat{I}_x \widehat{p}^2 \widehat{C}_{mag-f} \\ \widehat{I}_x \widehat{p} \widehat{C}_{L\alpha} \end{bmatrix}
\end{aligned} \tag{2.16}$$

where: $\mathbf{x} = [x, y, h]$ – 3-dimensional position vector, $\mathbf{v} = [v_x, v_y, v_h]$ – velocity of the projectile with respect to the air, $\mathbf{w} = [w_x, w_y, w_h]$ – wind velocity vector, p – angular velocity of the spinning motion, ρ – air density, m – mass of the projectile, $\widehat{p} = pd/v$ - dimensionless coefficient, S – cross-section area of the projectile, C_{spin} – spinning drag coefficient, I_x – moment of inertia along the x axis, d – caliber, C_{D_0} – zero-yaw drag coefficient, g – gravitational acceleration, γ_a – the elevation angle of $\mathbf{v}$ measured from the horizontal direction, i.e. the air-path inclination angle, χ_a – the azimuth angle of $\mathbf{v}$, i.e. the air-path azimuth angle, $\widehat{C}_{mag-f}$ – dimensionless Magnus force coefficient, $\widehat{C}_{L\alpha}$ – dimensionless linear lift force coefficient, $\widehat{C}_{D_\alpha^2}$ – dimensionless yaw drag coefficient. The dimensionless coefficients are defined as

$$\widehat{C}_{D_\alpha^2} = \frac{C_{D_\alpha^2}}{C_{M_\alpha}} \qquad \widehat{C}_{L\alpha} = \frac{C_{L\alpha}}{C_{M_\alpha}} \qquad \widehat{C}_{mag-f} = \frac{C_{mag-f}}{C_{M_\alpha}} \tag{2.17}$$

where C_{M_α} – overturning moment coefficient, $C_{D_\alpha^2}$ – yaw drag coefficient, $C_{L\alpha}$ – lift force coefficient, C_{mag-f} – Magnus force coefficient.

In order to solve variational equation (2.15), it will be necessary to compute the Jacobi matrix for the explicit form of the modified point mass trajectory model presented above

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial F_1}{\partial \mathbf{y}_1} & \cdots & \dfrac{\partial F_1}{\partial \mathbf{y}_7} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_7}{\partial \mathbf{y}_1} & \cdots & \dfrac{\partial F_7}{\partial \mathbf{y}_7} \end{bmatrix}. \tag{2.18}$$

Using the Runge-Kutta method of the 4th order, one can find a solution $\mathbf{\Phi}_t(\mathbf{y}_0, t_0)$ of the system of equations (2.16), which will then enable us to calculate the value of Jacobi matrix (2.18) at each of the nodal points. The Runge-Kutta method is applied again to linear non-stationary system (2.15) in order to propagate the Jacobi matrix along the solution $\mathbf{\Phi}_t(\mathbf{y}_0, t_0)$ (Randez, 1992)

$$\begin{aligned}
&k_1^n = \mathbf{J}^{n-1} \qquad k_2^n = \mathbf{J}_H \Big(\mathbb{I} + \frac{h}{2} k_1^n \Big) \qquad k_3^n = \mathbf{J}_H \Big(\mathbb{I} + \frac{h}{2} k_2^n \Big) \qquad k_4^n = \mathbf{J}^n (\mathbb{I} + h k_3^n) \\
&\mathbf{K}_n = \Big(\mathbb{I} + \frac{h}{6} [k_1^n + 2(k_2^n + k_3^3) + k_4^n] \Big) \mathbf{K}_{n-1}
\end{aligned} \tag{2.19}$$

where $\mathbf{J}_H$ is the Jacobi matrix obtained as a result of interpolation between the nodal points $n-1$ and n, while the letter h is the integration step in the Runge-Kutta method.

2.3. Probability of hitting the target with an artillery projectile

Let the initial condition of equations (2.16) be described by a normal distribution

$$\mathbf{y}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{y_0}, \boldsymbol{\Sigma}_{y_0}) \tag{2.20}$$

where

$$\boldsymbol{\mu}_{y_0} = [\mu_{x_0}, \mu_{y_0}, \mu_{h_0}, \mu_{p_0}, \mu_{v_0}, \mu_{\gamma_{a,0}}, \mu_{\chi_{a,0}}]^{\mathrm{T}} \tag{2.21}$$

is a vector of averages of the initial values of the system, and

$$\boldsymbol{\Sigma}_{y_0} = \begin{bmatrix} \sigma_x^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_h^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_v^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\chi_a}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\gamma_a}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_p^2 \end{bmatrix} \tag{2.22}$$

is a matrix containing the variances of the initial position of the system. By using the rule of error propagation (Ochoa and Belongie, 2006), covariance matrix for the position of the considered system after time t will be

$$\boldsymbol{\Sigma}_{y(t_{hit})} = \mathbf{K}(t_{hit})\boldsymbol{\Sigma}_{y_{t_0}}\mathbf{K}^{\mathrm{T}}(t_{hit}) \tag{2.23}$$

where t_{hit} is the time in which the projectile travels to the point where it meets the target.

The respective variances in (2.22) refer to: projectile position in three dimensional space, its velocity, elevation and azimuth angles, rotational speed. At this point, it is necessary to discuss some assumptions that were adopted when solving the problem for the selected model of the projectile flight:

1. The physical parameters of the projectile, such as mass and caliber, are treated as constant values.

2. The quantities used in Eq. (2.22) are described with the normal distribution where:
   - the initial position will depend on the measuring device used to establish the initial position;
   - parameters of speed distribution and dispersion of the gun in the elevation and azimuth angle planes are provided by the manufacturer;
   - the parameters of the distribution of the rotational speed of the projectile result directly from the parameters of the distribution of its muzzle velocity.

The covariance matrix related to the position of the tracked target at the meeting point has yet to be found. The value of this matrix will be calculated using the error propagation law

$$\boldsymbol{\Sigma}_{c(t_{hit})} = \mathbf{A}(t_{hit})\boldsymbol{\Sigma}_{c(t)}\mathbf{A}^{\mathrm{T}}(t_{hit}) \tag{2.24}$$

where $\mathbf{A}$ is the matrix of motion dynamics of the tracked target depending on the assumed target motion hypothesis (constant velocity motion, constant acceleration motion, constant turn motion etc.), $\boldsymbol{\Sigma}_{c(t)}$ is the covariance matrix related to the estimated (e.g. by the use of Kalman filtering) target motion parameters at the current moment t for the selected motion hypothesis, $\boldsymbol{\Sigma}_{c(t_{hit})}$ is the covariance matrix after the time t_{hit} for the selected motion hypothesis. Having the information about the target and projectile position in the area around the meeting point, the next step will be to calculate the probability of projectile and target collision. To solve this problem, the Monte Carlo method can be used, and the calculations are as follows:

1. Finding vectors and eigenvalues for the covariance matrices $\mathbf{\Sigma}_{y(t_{hit})}$ and $\mathbf{\Sigma}_{c(t_{hit})}$ such that

$$\mathbf{\Sigma}_{y(t_{hit})} = \mathbf{V}_p \mathbf{D}_p \mathbf{V}_p^{\mathrm{T}} \qquad \mathbf{\Sigma}_{c(t_{hit})} = \mathbf{V}_c \mathbf{D}_c \mathbf{V}_c^{\mathrm{T}} \tag{2.25}$$

   where $\mathbf{D}_p$ and $\mathbf{D}_c$ are diagonal matrices containing eigenvalues of the covariance matrix for the calculated projectile and target position respectively, $\mathbf{V}_p$ and $\mathbf{V}_c$ – matrices whose columns contain the matrix eigenvectors for the calculated projectile position and target, respectively.

2. Drawing two sets of points in the three-dimensional Cartesian space

$$\mathbf{X}_c = [\mathbf{x}_c, \mathbf{y}_c, \mathbf{h}_c]\mathbf{D}_c \qquad \mathbf{X}_p = [\mathbf{x}_p, \mathbf{y}_p, \mathbf{h}_p]\mathbf{D}_p\mathbf{R} \qquad \mathbf{R} = \mathbf{V}_c\mathbf{V}_p \tag{2.26}$$

   where $\mathbf{x}_c, \mathbf{y}_c, \mathbf{h}_c, \mathbf{x}_p, \mathbf{y}_p, \mathbf{h}_p$ are N-element column vectors drawn from a standardized normal distribution.

3. Calculate probability by summing the number of elements in both sets ($\mathbf{X}_c$ and $\mathbf{X}_p$) that are simultaneously in the target surrogate area, and then divide the result by N.

The probability of hitting the target, calculated according to the above algorithm, is determined with a certain accuracy and strictly depends on the selection of the surrogate area referred to in point 3 of the algorithm description.

3. Simulation method for finding the probability of hitting the target

The method of calculating the probability that will be presented in this s Section has been prepared as a reference tool. The main reasons why this method cannot be used in a real time calculations of the hit probability are:

- basing the algorithm on the assumption that the ideal trajectory of the tracked object is known;
- very long simulation time, the performance of which is necessary to determine the target hit probability.

This algorithm can be divided into two main parts, the first of which will be responsible for simulating the projectile flight in the atmosphere, while the second will focus on simulating the movement of the tracked target.

Projectile motion in the atmosphere will be described by a system of differential equations (2.16)

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{y}(t) = F_p(t, \mathbf{y}(t)) \tag{3.1}$$

with rhe initial condition

$$\mathbf{y}(t_0) = \mathbf{y}_0 \tag{3.2}$$

In real conditions, each quantity in vector (3.2) is burdened with a certain error, which will be described for the purposes of simulation tests with a normal distribution. In the conducted simulations, the vector of initial conditions will have the form

$$\mathbf{y}(t_0) = \mathbf{y}_0 + \Delta\mathbf{y}_0 \tag{3.3}$$

where the respective elements of the vector $\Delta\mathbf{y}_0$ are described with the use of normal distribution

$$\Delta {y_0}^i \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{3.4}$$

It should also be remembered that not only the initial conditions affect the projectile trajectory. The equations also use the projectile physical parameters (mass, diameter) denoted as $\mathbf{q}$, which are also burdened with a normally distributed error

$$\mathbf{q}_n = \mathbf{q} + \Delta\mathbf{q} \qquad \Delta q_j \sim \mathcal{N}(\mu_j, \sigma_j^2) \tag{3.5}$$

The values of distribution parameters for each quantity depend on the type of ammunition and are specified by the manufacturer. The flight trajectory simulations will be carried out in a standard atmosphere – uncertainties related to determination of the meteorological situation will be ignored. The reason for this approach is primarily the problem of obtaining appropriate data that can be used to define the aforementioned uncertainties. However, it should be emphasized that the prepared application allows one to take them into account as much as possible. The second essential part of the probability calculation algorithm is related to the simulation of the target movement being fought by the anti-aircraft system. In order to carry out simulation tests, sets of test trajectories were prepared with the use of the route generator.

Each of the sets included two types of routes: **ideal trajectory** with successive target positions generated with a selected frequency, treated at a later stage as a reference trajectory, **noisy trajectory** with subsequent measurements of angular coordinates and distance to the target – measurement errors and the frequency of measurements depends on the sensors used in the anti-aircraft set (radar, optoelectronic head, etc.). After each simulation, it is checked whether the projectile hits the tracked object with the given perturbations for the parameters and the initial conditions of the differential equations describing the projectile motion. The so-called standard NATO objectives being cuboids with dimensions:

- $12 \times 2.3 \times 2.3$ m represents fighter-type targets;
- $2 \times 0.5 \times 0.5$ m represents drone targets.

The application developed in the Matlab environment has the ability to simulate the described situation for various types of targets and ammunition together with visualization of the entire situation. This approach to calculating the probability is a very good verification tool, but requires a lot of time and computing power.

4. Comparison of the results for algorithms calculating the target hit probability

In order to verify the correctness of the algorithm of the analytical method for calculating the probability of hitting a tracked object with a single bullet, appropriate simulation tests were carried out. The research included calculating the hit probability using the analytical and simulation method (considered as the reference method) for the following variants of air target movement: stationary target, target approaching the artillery, moving with constant velocity motion, target moving away from the artillery with constant velocity motion. During the simulation tests, only uncomplicated types of target dynamics were considered – which are sufficient to evaluate the probability calculation method, which does not depend on the type of target movement. For all the cases described above, the centre of the coordinate system was assumed at the point where the artillery is standing. The generated ideal trajectory of the target was disturbed, for the purposes of research, with the following measurement errors:

- white noise has been added to the azimuth and elevation angles $\mathcal{N} \sim (0, \sigma_{a,e})$, where $\sigma_{a,e} = 0.3$ mrad;
- quantization error was added to the distance measurement $\Delta_r = 2$ m.

The trajectory of the projectile fired towards the meeting point will be calculated using the equations of the modified point mass trajectory model in its explicit form (see equations

(2.16)). Propagation of disturbances of the initial conditions of differential equations around the solution is performed according to the algorithm presented in Subsection 2.2, using Jacobian for the explicit form of the modified point mass trajectory model. The initial conditions disturbances used in the simulations have the following values:

- $\Delta x_0, \Delta y_0, \Delta h_0 \sim \mathcal{N}(0, \sigma_x)$, where $\sigma_x = 0.5\,\mathrm{m}$;
- $\Delta v_0 \sim \mathcal{N}(0, \sigma_v)$, where $\sigma_v = 1.5\,\mathrm{m/s}$;
- $\Delta \chi_{a,0} \sim \mathcal{N}(0, \sigma_{\chi_a})$, where $\sigma_{\chi_a} = 0.32\,\mathrm{mrad}$;
- $\Delta \gamma_{a,0} \sim \mathcal{N}(0, \sigma_{\gamma_a})$, where $\sigma_{gamma_a} = 0.32\,\mathrm{mrad}$;
- $\Delta p_0 \sim \mathcal{N}(0, \sigma_p)$, where $\sigma_p = 0\,\mathrm{rad/s}$.

The above values of the initial conditions disturbance are used both in the analytical and reference methods for calculating the target hit probability.

4.1. Stationary target

The tests were carried out for different distances of the target from the firing position. The paper presents exemplary calculation results for the target located on the border of the effective range. In this case, it was assumed that the target is at point $C = (x_c, y_c, h_c) = (-4000, 300, 700)$ relative to the starting point of the projectile, which is also the center of the coordinate system. The target tracking process took 25 s, and the hit probability was calculated at $t_p = 1, 3, 5, \ldots, 19, 21\,\mathrm{s}$. The results of simulation tests for the discussed case are included in Table 1.

Table 1. The probability of hitting a stationary target calculated in subsequent moments of simulation. P_{H1} – probability calculated using the reference method, P_{H2} – probability calculated using the analytical method

	t_1 [s]										
	1	3	5	7	9	11	13	15	17	19	21
P_{H1}	0.570	0.568	0.550	0.564	0.556	0.574	0.534	0.594	0.570	0.550	0.572
P_{H2}	0.002	0.030	0.481	0.481	0.408	0.477	0.481	0.481	0.483	0.480	0.482

At this point, it is necessary to discuss the differences that occurred when calculating the probability of an artillery projectile hitting the tracked target. The most distinct difference occurred in the first second of the tracking process (probability values are highlighted in orange). In Fig. 1a, one can see how the trace of the covariance matrix resulting from the operation of the Ballistic Computing Module and related to the calculated future target position has changed. Only after 3 s, the filter went into a steady state – the appropriate target dynamics was adjusted, which increased the accuracy with which the target position was determined.

A stationary target close to the limit of the effective range of the tested type of artillery ammunition is, contrary to intuition, a case for which the determination of the motion parameters (and thus the prediction of the future location) is relatively difficult. From the point of view of the FCS, the target moves completely randomly depending on the measurement errors of sensors (in real conditions, often also on atmospheric conditions) - an examplary trajectory is shown in Fig. 1.

4.2. Constant velocity motion – incoming target

The starting point of the trajectory of the tracked target is $C_0 = (x_0, y_0, h_0) = (-4500, 300, 700)\,\mathrm{m}$; the target is moving with a constant velocity $\mathbf{V}_c = [v_x, v_y, v_h] = [100, 0, 0]\,\mathrm{m/s}$. The movement of the target is simulated over the time $t_s = 22\,\mathrm{s}$ (the target

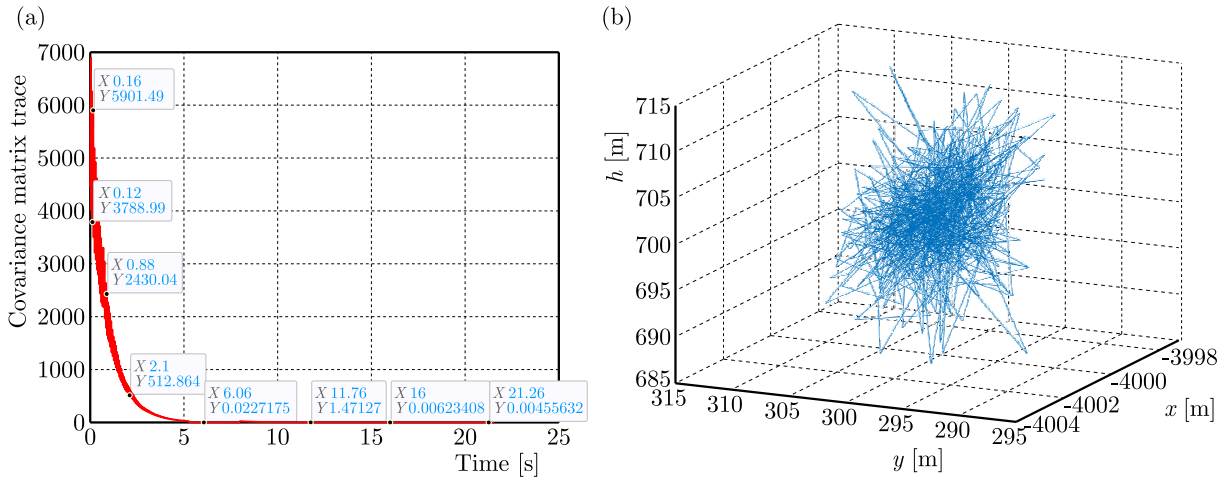


Fig. 1. (a) Trace of the covariance matrix resulting from the operation of BCM and related to the location of the tracked target (point $C = (-4000, 300, 700)$). The figure shows trace values for the covariance matrix at selected moments in the tracking process. (b) The trajectory of a stationary point recorded by the system after taking into account sensor measurement errors

trajectory is shown in Fig. 2b). The probabilities of hitting the target, calculated in the moments $t_p = 1, 3, 5, \ldots, 19, 21$ s are presented in Table 2. In the case of the analytical method, the values of target hit probability increase over time. This is understandable given the trace of the covariance matrix (Fig. 2a) related to the determination of the target location by the chosen filtering method. Compared to the stationary target, the target dynamics was selected by the system very quickly (less than 1 s), and because it was selected in accordance with the actual target movement, errors in determining target motion parameters decreased shortly after starting the tracking process.

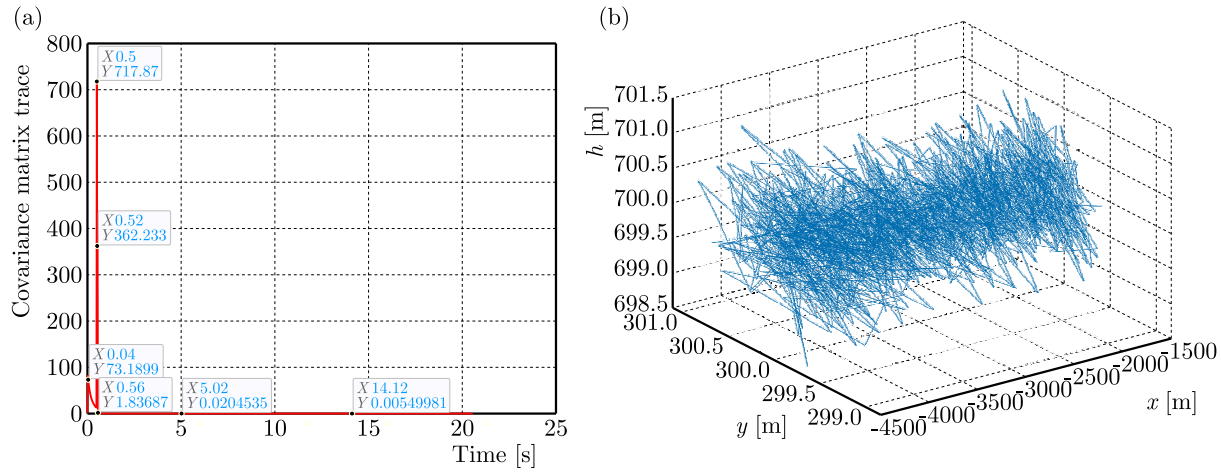


Fig. 2. (a) Trace of the covariance matrix resulting from the operation of FCS and related to the location of the tracked target (the target approaching the artillery gun with constant velocity motion). The figure shows the trace values of the covariance matrix at selected moments in the tracking process. (b) The trajectory of the target moving with the constant velocity motion model (the target is approaching) recorded by the system after taking into account sensor measurement errors

The analytical method is sensitive to the information coming from the FCS, and more precisely, to the values of the covariance matrix related to the parameters of the target movement. The stabilization of the tracking process with the Kalman filter causes that the set of random points representing the position of the target around the meeting point, Eq. $(2.25)_2$, is very

Table 2. The probability of hitting a target moving with a constant velocity, calculated in successive moments of the simulation. P_{H1} – probability calculated using the reference method, P_{H2} – probability calculated using the analytical method

	t_1 [s]										
	1	3	5	7	9	11	13	15	17	19	21
P_{H1}	0.418	0.628	0.670	0.686	0.706	0.768	0.798	0.862	0.912	0.910	0.894
P_{H2}	0.638	0.682	0.737	0.710	0.775	0.842	0.881	0.914	0.948	0.989	0.996

concentrated around that point. Moreover, in the case of an approaching target, the projectile trajectory to the meeting point, calculated in each subsequent iteration of the FCS, is shorter than the previous one. Therefore, it can be expected that as a result of the propagation of disturbances around the solution of the projectile motion equations, the obtained values of the variance of the projectile motion parameters decrease with each subsequent iteration of the FCS. This naturally translates into an increase in the value of the calculated probability.

Also in the case of the reference method, the values of the probability of hitting the target increase in the following moments. The reasons are similar to those given for the analytical method – approaching the target to the center of fire reduces the dispersion of projectiles around the meeting point and, at the same time, the gun settings used in the simulation of the projectile flight are determined by the FCS with a smaller error (resulting directly from the accuracy of determining the movement parameters of the tracked target).

In the case under consideration, the discrepancies between both methods of calculating the probabilities are greater. This indicates a slightly greater sensitivity of the analytical method to the information on accuracy of determining the target location coming from the FCS. One should also remember about a certain assumption that affects the probabilities between the two methods - the reference method is based on the assumption that the real trajectory of the target is known. On the other hand, in the analytical method, the location of the target is determined only on the basis of the estimated parameters of the target movement from the FCS.

4.3. Constant velocity motion – receding target

The initial position of the trajectory of the tracked target is $C_0 = (x_0, y_0, h_0) = (500, 300, 700)$ m; the target moves with a constant velocity $\mathbf{V}_c = [v_x, v_y, v_h] = [100, 0, 0]$ m/s (Fig. 3b). The target movement is simulated for $t_s = 22$ s. The probability of hitting the target, calculated at times $t_p = 1, 3, 5, \ldots, 19, 21$ s, is presented in Table 3.

Table 3. The probability of hitting a target moving with a constant velocity, calculated in successive moments of the simulation. P_{H1} – probability calculated using the reference method, P_{H2} – probability calculated using the analytical method

	t_1 [s]										
	1	3	5	7	9	11	13	15	17	19	21
P_{H1}	0.851	0.754	0.946	0.878	0.858	0.890	0.8680	0.830	0.798	0.816	0.824
P_{H2}	1.000	1.000	1.000	0.998	0.982	0.966	0.9170	0.852	0.795	0.731	0.679

In this case, the probability calculated by the analytical method decreases as the target moves away. As onr can see in Fig. 3a, the errors in determining the target motion parameters in the first moments of the tracking process are very small (especially compared to the errors occurring in the case of an approaching or stationary target). It is understandable, taking into account the fact that measurement errors of the azimuth and elevation angles propagate with the distance from the sensor. Fluctuations in the values of the probability determined with the use

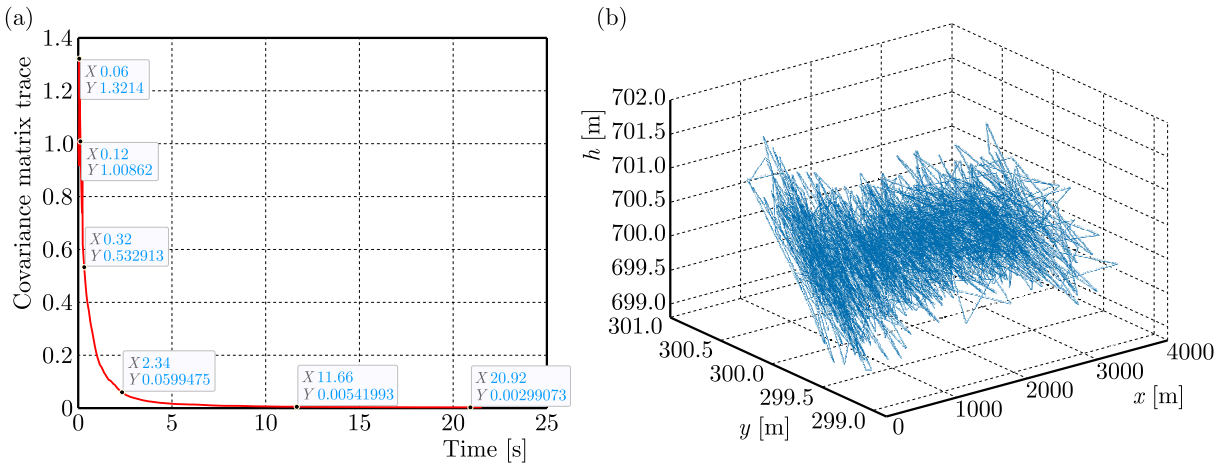


Fig. 3. (a) Trace of the covariance matrix resulting from the operation of FCS and related to the location of the tracked target (the target moving away from the fire center with a constant velocity). The figure shows the trace values of the covariance matrix at selected moments in the tracking process. (b) The trajectory of the target moving with a constant velocity (the target moves away) recorded by the system after taking into account sensor measurement error

of the reference method are caused by the sensitivity of the method to incorrectly determined gun settings at a given moment.

5. Conclusions

The algorithm for analytical calculation of the probability was verified using a simulation tool prepared in the Matlab environment based on the assumption that the real target trajectory is known. Significant discrepancies between the results obtained from both methods are a direct outcome of the filter bank implemented in the Fire Control System. As has been shown, the algorithms produce very similar results when the Kalman filter matched to a given target dynamics enters a steady state. Until this moment, the uncertainties in determining the target location around the meeting point are too high. It should be emphasized, however, that the time needed by the filter to reach the steady state is very short, up to 2 s.

The time of the filter transition to the steady state, unfortunately, cannot be predicted with an efficient method – it depends on many factors: measurement errors of sensors used for detecting objects, dynamics and motion parameters of the tracked target. An important factor in the tracking process are maneuvers performed by the tracked target, weather conditions that may have a negative impact on the correct operation of the sensors, e.g. the influence of rain on the operation of the laser rangefinder, the influence of clouds on the operation of the videotracker. The main advantage of the developed analytical method (apart from the aforementioned compatibility with the simulation tool) is its runtime. The time of one FCS iteration is 20 ms. The time allocated to the Ballistic Calculation Module is 14 ms. In the case of using 6 anti-aircraft guns, the maximum execution time of the probability calculation algorithm is 1 ms.

The algorithm operation time was tested on a ballistic computer used in the Fire Control System. Runtime of the algorithm is mainly influenced by three factors: chosen mathematical model for projectile motion, number of steps in the integration method of differential equations describing motion of a projectile, number of samples in the Monte Carlo simulation. It is obvious that reducing the number of steps and samples will shorten the algorithm runtime, but will negatively affect its accuracy. These values should be selected depending on the specification of the hardware by which the calculations will be performed. The analytical method for calculating

the hit probability can be used as a tool to support the operator decision making. Firing a shot after reaching the appropriate hit probability level should reduce the amount of ammunition needed to neutralize the target being tracked. Fewer projectiles will translate into increased service life of both the barrel and the gas chamber of the anti-aircraft gun.

Acknowledgements
This work was financed by Military University of Technology under research project UGB 824/2023.

References

1. Baranowski L., 2013a, Effect of the mathematical model and integration step on the accuracy of the results of computation of artillery projectile flight parameters, *Bulletin of the Polish Academy of Sciences – Technical Sciences*, **61**, 2, 475-484
2. Baranowski L., 2013b, Feasibility analysis of the modified point mass trajectory model for the need of ground artillery fire control systems, *Journal of Theoretical and Applied Mechanics*, **51**, 3, 511-522
3. Baranowski L., Gadomski B., Majewski P., Szymonik J., 2016a, Explicit "ballistic M-model": a refinement of the implicit "modified point mass trajectory model", *Bulletin of the Polish Academy of Sciences – Technical Sciences*, **64**, 1, 81-89
4. Baranowski L., Gadomski B., Szymonik J., Majewski P., 2016b, Comparison of explicit and implicit forms of the modified point mass trajectory model, *Journal of Theoretical and Applied Mechanics*, **54**, 4, 1183-1195
5. Jordan D.W., Smith P., 2007, *Non-Linear Ordinary Differential Equations. An Introduction for Scientists and Engineers*, Oxford University Press, New York
6. Kang H., Kim H., Kim G., 2016, A study on prediction of probability of hit for an anti-aircraft artillery, *Indian Journal of Science and Technology*, **9**, Special Issue 1, 1-5
7. Katsev I., 2018, Evaluation method of the artillery's effectiveness against unitary target, *International Scientific Journal Security and Future*, **2**, 4, 196-198
8. Liu H., Shi X., 2022, Damage effectiveness calculation of hitting targets with ammunition based on Bayesian multinomial distribution, *Symmetry*, **14**, 5, 892
9. McCoy R.L., 1999, *Modern Exterior Ballistics. The Launch and Flight Dynamics of Symmetric Projectiles*, Schiffer Publishing
10. Obradović Z.B., Nikolić B.B., Popović G.P., Gadzhalov N.D., Krastev K.I., 2023, Application of the modeling method to the calculation of the probability of hitting a stationary target during the fire action of a tank squad in defence, *Military Technical Courier*, **71**, 3, 722-739
11. Ochoa B., Belongie S., 2006, Covariance Propagation for Guided Matching, *Proceedings of the Workshop on Statistical Methods in Multi-Image and Video Processing*, 1-12
12. Parker T., Chua L., 1989, *Practical Numerical Algorithms for Chaotic Systems*, Springer-Verlag, New York
13. Rández L., 1992, On the simultaneous numerical integration of an IVP and its associated variational equation, *Journal of Computational and Applied Mathematics*, **39**, 1, 103-108
14. STANAG 4355, 2009, *The Modified Point Mass and Five Degrees of Freedom Trajectory Models*, Edition 3

Manuscript received September 25, 2023; accepted for publication October 9, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 799-826, Warsaw 2024
https://doi.org/10.15632/jtam-pl/194319

TWO TOOLS FOR THE SCIENCE OF MATERIAL EFFORT – *a review paper*[1]

Waldemar Dudda
University of Warmia and Mazury, Faculty of Technical Sciences, Olsztyn, Poland
e-mail: dudda@uwm.edu.pl

Zdzisław Nowak
Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
corresponding author, e-mail: znowak@ippt.pan.pl

Tomasz Ochrymiuk, Janusz Badur
Institute of Fluid-Flow Machinery, Energy Conversion Department, Polish Academy of Sciences, Gdańsk, Poland
e-mail: tomasz.ochrymiuk@imp.gda.pl; janusz.badur@imp.gda.pl

This paper discusses the science of material effort from the historical viewpoint. Two general scientific tools: the geometrical descriptive method of Mohr, and the energetic method of Huber are compared and evaluated from the very beginning. Three appropriate stress invariants are taken into account: stress intensity, stress triaxiality and stress shearness. Especially, much attention is devoted to explanation of the stress shearness invariant, which aims at describing the Lode parameter in a more analytical manner. Two different tools of finding a proper yield surface which contains the above mentioned three stress invariants are discussed in the literature perspective. In particular, the three-parameter yield surface, called the Burzyński-Pęcherski hypothesis is researched and explained from this new point of view.

Keywords: Tresca-Mohr hypothesis, Beltrami-Huber, Mises, Burzyński Hypothesis, Lode parameter, material effort

1. Introduction and motivation

1.1. Huber's science of material effort

In 1904, Tytus Huber established the science of material effort as a field of knowledge between three-dimensional theory of elasticity and zero-dimensional (engineering) science of the strength of materials. The science of effort was applied then to accomplish the most difficult thing – it was to use the language and mathematical methods of the three-dimensional theory of elasticity to determine the state of material effort by means of a single characteristic quantity called "effort". This quantity, measured as volumetric energy density [J/m^3], was denoted by Huber with the letter W (from the Polish word "wytężenie").

Why did Huber decide to determine the effort state of a material by a single characteristic scalar $W(t,\mathbf{x})$? Why did he choose such a simple measure of effort? Why did he not take, for instance, three invariants of the stress tensor $I_1=\sigma_{ii}$, $J_2=(1/2)s_{ij}s_{ij}$, $J_3=(1/3)s_{ij}s_{jk}s_{ji}$ as three measures of effort?

Huber assumed that the effort state of the material must be expressed by a single scalar quantity, since the science of effort had been supposed to combine mathematical theory with

[1]The paper was presented at the XIIIth conference PLASTMET'2023, Łańcut Zamek, 7-10 October, 2023, in a session in honour of Prof. Ryszard B. Pęcherski organized by prof. Romana Śliwa and prof. Katarzyna Kowalczyk-Gajewska

engineering practice, and engineering practice at that time used a single critical quantity, which was the "maximum strength". In order to compare that integral, 0-dimensional quantity, measured at the level of the entire device, with the effort $W(t, \mathbf{x})$ determined locally at the continuum point level. Huber (1904) introduced the concept of critical energy density and denoted it with the letter K [J/m^3]. It no longer depended on a point in the material, as it had been defined from measurements on a large sample. This was when the first mathematical expression of the science of material effort arose

$$W(t, \mathbf{x}) \leqslant K \tag{1.1}$$

It was spoken as a rule: the material effort must not exceed the strength of the material. The engineers adopted this principle as a design principle to "effort yourself to endure" (in Polish: wytężać tak, aby wytrzymywać). In his work, Huber develops three different definitions of the measure of the material effort. To prove which ones are better, he compares the results with experimental results and with the predictions resulting from Otto Mohr's geometrical approach (Mohr, 1882). At that time, geometrical methods were considered superior to analytical methods. It must be remembered what was the dominant status of "methods of descriptive geometry" in science and mechanics at that time. An analytical model that did not agree with Mohr's geometrical results was considered inferior or even conceptually wrong by early 20th-century scholars. A dispute arose among theoreticians as to which of the approaches was the right one: Huber's or Mohr's? In other words, which of the two research tools – geometrical or energetic – is correct? A challenge arose among experimenters to prove experimentally which of the approaches was valid? Stepan Timoshenko (1953) mentions and discusses the key experiments of the period 1900-1930, among them the Lode experiment (1926).

1.2. Mohr's science of material strength

Mathematically, Mohr's geometric approach was also innovative. It referred to the method proposed by Clapeyron and Lame in 1832 and Stokes in 1845 to construct constitutive equations only through principal values of the stress tensor and principal values of the strain tensor, respectively. The most important thing in that method was to find the principal stresses σ_1, σ_2, σ_3 using the descriptive method (the so-called Mohr circle; Mohr, 1882). And then, in the corresponding diagram, τ-σ making the so-called maximum envelopes of the circles (Fig. 1).

Let us emphasize, the geometrical method led Mohr to the hypothesis of marginal envelopes, which does not belong to any of the three types of hypotheses – it is not the hypothesis of extreme stresses, extreme deformations, or the hypothesis of extreme energies. It is a geometrical hypothesis developed for the needs of bulk materials such as sand or grain. Perhaps, the envelope hypothesis, so strongly related to the construction of Mohr's circle, is some kind of "circling" around a new analytic expression for effort. Therefore, when we begin to study Mohr's hypothesis, we should keep in mind Burzyński's question "Are Mohr's envelopes the complete content of his hypothesis of material effort, or are they an illustration of some unknown hypothesis, being a mere means of drawing, more or less successful?"

In 1900, Otto Mohr decided to introduce a new concept: the limit shear strength k_{tr}, which had a weak experimental reference (Mohr, 1900). This means that among the conceivable deformation states, in addition to the torsional state, there is a condition that shows similarity to technological cutting. Mohr explains this limit state as an independent state, always occurring in addition to tension, compression and torsion, being the fourth limit state – Mohr hypothesizes that this state can be determined by shear stresses and that the maximum limit value of k_{tr} occurs in a cross-section in which the normal stress is equal to zero. Hence Mohr's magnitude: k_{tr} he calls "shear strength" – but this is a concept difficult to grasp experimentally. Hence, this notion is criticized by, for instance, A. Föppl in his monograph: *Exercises in Technical Mechan-*

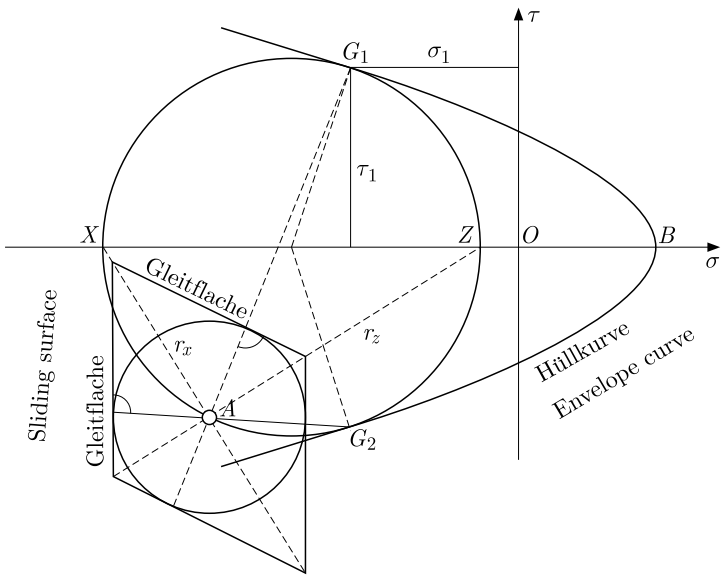


Fig. 1. Mohr's envelope plot in the plane σ-τ (Mohr, 1906, Fig. 13)

ics, from 1907 (Föppl, 1907). The point is how to express it: k_{tr} through the measurement-known limits of k_t, k_c.

Undoubtedly, the starting point of Mohr's reasoning was the Tresca hypothesis, which referred to the maximum shear stress and the parameter k_s that determines it. It can also be called a constant marginal shear stress hypothesis (Pełczyński, 1958). It is defined in the principal axes as $\sigma_{eq,t}^{Tresca} = (\sigma_1 - \sigma_3)/2 \leqslant k_s$ where the stress σ_{eq}^{Tresca} is the stress equivalent to the shear test from which we obtain the value of k_s. If we assume, as it is "done" by many researchers, that $k_s = k_t/2$, then the Tresca stress equivalent to the uniaxial tensile test is the equation $\sigma_{eq,r}^{Tresca} = \sigma_1 - \sigma_3 \leqslant k_r$.

Then, in the Mohr plane τ-σ, one can draw the boundaries of the Mohr circles that have radius $(\sigma_1 - \sigma_3)/2$, and this boundary is two lines parallel to σ. On the other hand, assuming that $k_t = k_c$, then in the space of principal stresses, we can assign a prism with a symmetrical hexagonal cross-section to the above condition.

As Huber mentions in his historical outline of the development of the effort hypotheses (*Technical Stereomechanics*, p. 88, Huber, 1948), Mohr's motivation was to consider a combined shear in the slip plane τ and the internal friction forces proportional to the pressure force σ in this cross-section. With this motivation in front of him, Mohr writes the effort condition as $\sigma_{eq}^{CM} = \tau_{max} = \tau + f\sigma \leqslant k_{tr}$.

It is therefore an adaptation of the hypothesis of extreme internal friction developed by Coulomb in 1776 known as the hypothesis *slippage with friction* $\pm\tau + f\sigma \leqslant k_{tr}$. It is a hypothesis belonging to the group of hypotheses based on two constants: the coefficient of friction f and the cohesive force k_{tr}. It was developed by L. Navier (1837) and Ch. Duguet (1885); especially for loose and porous media. Therefore, the critical state is reached as a result of the formation of slip planes, which, thanks to the forces of internal friction and cohesion, are inclined in relation to the main axes at angles slightly different from $\pi/4$ and for metals, according to Duguet, it is about $\beta = 10^\circ$, which gives the coefficient of friction $f = \tan\beta = 0.176$ (Timoshenko, 1953).

Let us remember that other concepts of the material effort, such as Mohr's or Burzyński's, have given up on the constancy of the f coefficient. In Coulomb's hypothesis, the envelopes of Mohr's circles are instantaneous. Moving now from the constants f and k_{tr} to the available experimental data, we can see that at least two types of experiments are needed to determine

them – for example, a uniaxial tensile test and a uniaxial compression test. Then we first write the above condition as (Pełczyński, 1959)

$$\frac{\sigma_1 - \sigma_3}{2} + \frac{\sigma_1 + \sigma_3}{2} n \leqslant k_{tr} \tag{1.2}$$

We have gone here permanently, $n = \sin\beta$ and $k_{tr} = k_t \cos\beta$. In turn, these constants are easy to express in terms of tensile and compressive limit constants k_t and k_c as

$$n = \frac{k_c - k_t}{k_c + k_t} \qquad k_{tr} = \frac{k_c k_t}{k_c + k_t} \tag{1.3}$$

This means that we do not have to measure the angle of internal friction and the constant of cohesion for metals, but it is enough to perform two uniaxial tests, compression and tensile. When the limits of $k_t = k_c$ are equal, the angle of internal friction disappears, and Coulomb's hypothesis comes down to Tresca's hypothesis.

A fundamental feature of Mohr's exertion model has gone down in the history of the problem, namely that it does not take into account the mean (second) principal stress σ_2 while Huber's hypothesis does. Further reconstructions of Mohr's approach are provided in Dudda's (2021) paper.

1.3. Lode's crucial experiment

It was Walther Lode who undertaken the question which hypothesis – Huber's or Mohr's – is correct (Lode, 1926). Lode having only 11 probes (5 cast iron, 5 cooper and 1 nickel) in a form of long thin-walled pipes, performed tests loaded simultaneously by internal pressure and uni-axial tension or compression. Since, in that case, all three invariants σ_1, σ_2, σ_3 were easy to be determined, Lode was able to proof the influence of the second main stress σ_2 on strength limits. Denoting by μ (now the Lode parameter), some stress shearness losses he expressed second principal stress as $\sigma_2 = 0.5(\sigma_1 + \sigma_3) + 0.5\mu(\sigma_1 - \sigma_3)$. Knowing σ_1, σ_2, σ_3 from the experimental data, he defined the value of μ from (Lode, 1926, his Eq. 23)

$$\mu = \frac{2\sigma_2 - \sigma_1 - \sigma_3}{\sigma_1 - \sigma_3} \qquad \text{with} \qquad \sigma_1 \geqslant \sigma_2 \geqslant \sigma_3 \tag{1.4}$$

Therefore, when the state of pure shear is realized, the Lode parameter is to be zero, what means that in that state the Tresca and Coulomb-Mohr criterions are best. But in a real state of stresses, the prefect shearness is a lit bit loosed, therefore $\mu \neq 0$. By a simple inspection Lode finds that $\mu = -1$ in the state of uniaxial tension or two-axial compression, and $\mu = 1$ for uniaxial compression and double-tension. Next, hopefully, Lode having a possibility of performing experiments with a fixed value of μ parameter ($\mu = -1, -0.5, 0, 0.97$) was able to made the draw a curve $f_{exp} = (\sigma_1 - \sigma_3)/\sigma_z$ where σ_z measured normal stress which has a meaning the equivalent stress (Fig. 2).

The Lode diagram has a great generality – it is enough to replace σ_2 by μ and to reorganize any criterion to a requested form $(\sigma_1 - \sigma_3)/\sigma_{eq} = f(\mu)$ (see Fig. 2).

Lode, having experimental curve, was able to compare it with some theoretical curves based on Tresca, Beltrami, Huber hypothesis as (Fig. 3)

$$\begin{aligned} f_{Tresca} &= \frac{\sigma_1 - \sigma_3}{\sigma_{Tresca}} = 1 = \text{ const} \\ f_{Beltrami} &= \frac{\sigma_1 - \sigma_3}{\sigma_{Beltrami}} = \sqrt{\frac{1}{1 + \frac{1}{4}(1+\mu)^2 - \nu(1+\mu)}} \\ f_{HMH} &= \frac{\sigma_1 - \sigma_3}{\sigma_{HMH}} = \frac{\sigma_1 - \sigma_3}{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 - \sigma_1\sigma_2 - \sigma_2\sigma_3 - \sigma_3\sigma_1}} = \frac{2}{\sqrt{3 + \mu^2}} \end{aligned} \tag{1.5}$$

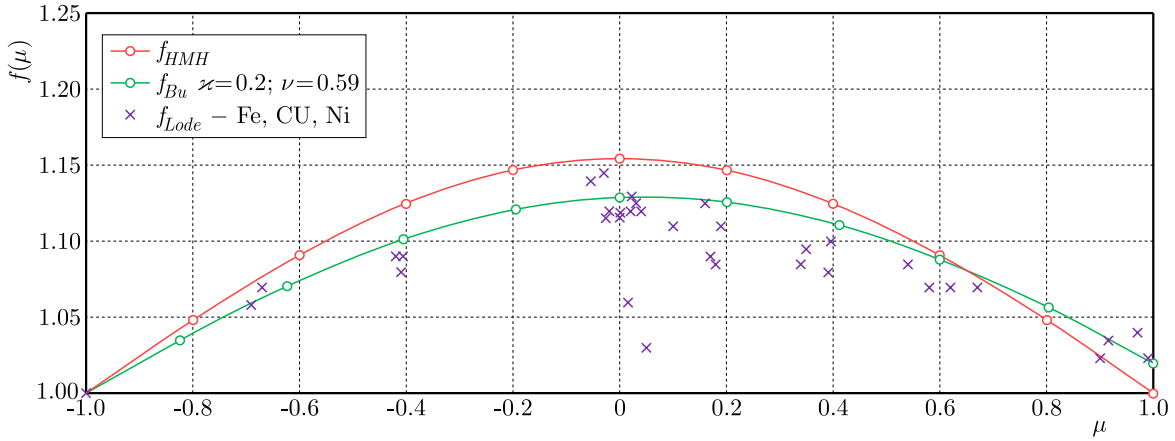


Fig. 2. The Lode diagram showing the influence of stress shearness into a state of material effort. A dimensionless pure shear factor μ, Eq. (1.4), gives a possibility to compare different materials like iron, cooper, nickel. In above, $f_{HMH} = (\sigma_1 - \sigma_3)/\sigma_{HMH} = 2/\sqrt{3+\mu^2}$ and $f_{Bu} = (\sigma_1 - \sigma_3)/\sigma_{Bu}$, where σ_{Bu} is determined by Eq. (1.7)

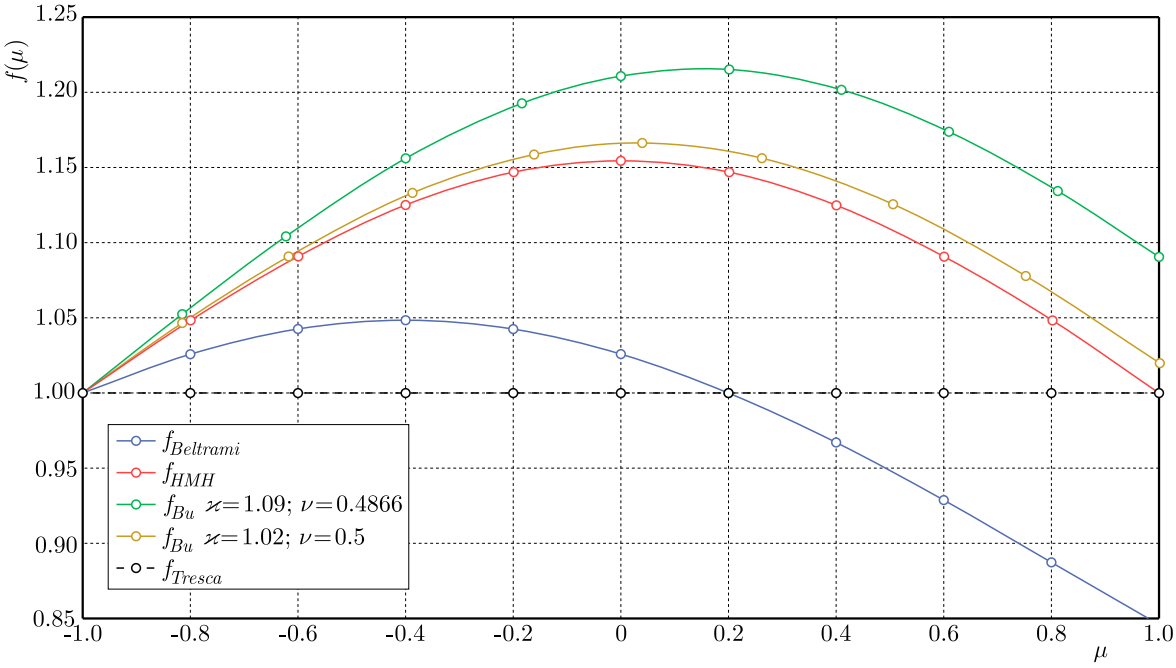


Fig. 3. The Lode diagram showing comparison of theoretical curves: Tresca's, Beltrami's (with Poisson's coefficient $\nu = 0.3$), Huber's, and two Burzyński's

Nowadays, one can add the Burzyński curve

$$f_{Bu} = \frac{\sigma_1 - \sigma_3}{\sigma_{Bu}} \tag{1.6}$$

With the equivalent stress (Dudda, 2021)

$$\sigma_{Bu} = \frac{\sigma_1 + \sigma_2 + \sigma_3}{2\varkappa}\left\{(\varkappa - 1) + \sqrt{(\varkappa-1)^2 + 4\varkappa\left[(1+\overline{\nu})\frac{\sigma_1^2+\sigma_2^2+\sigma_3^2}{(\sigma_1+\sigma_2+\sigma_3)^2} - \overline{\nu}\right]}\right\} \tag{1.7}$$

where the Burzyński coefficients are $\varkappa = k_c/k_t$, $\overline{\nu} = k_c k_t/2k_s^2 - 1$.

From Fig. 2, it follows that the Huber hypothesis is better fitting with the experimental data then the Tresca curve (compare Figs. 3 and 2). The Beltrami hypothesis is always wrong. The

Burzyński curves can be calibrated using appropriate values of $\varkappa$ and $\overline{\nu}$. In Burzyński's curve we have the boundary condition $f_{Bu} = \varkappa$ for $\mu = 1$.

Now, it is interesting to compare the Huber and Mohr hypotheses. The task is to demonstrate that the equivalent Huber stress calculated according to the deformation energy is always smaller than the equivalent stress calculated according to the Tresca hypothesis of the greatest shear stresses. To do this, it is necessary to separate from the definition of σ_{HMH}, before the root, the term $\sigma_1 - \sigma_3 \equiv \sigma_{Tresca}$

$$\sigma_{HMH} = \frac{1}{\sqrt{2}}(\sigma_1 - \sigma_3)\sqrt{\left(\frac{\sigma_1 - \sigma_2}{\sigma_1 - \sigma_3}\right)^2 + \left(\frac{\sigma_2 - \sigma_3}{\sigma_1 - \sigma_3}\right)^2 + 1} \tag{1.8}$$

And further, eliminating, as Burzyński does, the mean stress with the help of the Lode parameter μ

$$\sigma_2 = \frac{1+\mu}{2}\sigma_1 + \frac{1-\mu}{2}\sigma_3 \tag{1.9}$$

We obtain

$$\sigma_{HMH} = \sigma_{Tresca}\sqrt{\frac{3+\mu^2}{4}} \tag{1.10}$$

from which one can see that since $-1 < \mu < 1$ is even for $\mu = 0$, $\sigma_{HMH}/\sigma_{Tresca} = 0.866$. It means, in practice that the HMH hypothesis is safer. But for materials with $\varkappa \neq 1$ the Burzyński is safer then HMH. For instance, for pure compression $\mu = 1$, we obtain $f_{Bu} = \varkappa$, which means that $\varkappa$ changes the shape of limiting curve.

The above operation one can found in the monograph of Krzyś and Życzkowski (1962). This result does not surprise Krzyś and Życzkowski, as they add that "the correctness of the statement is immediately visible". Therefore, safer hypotheses should be used. What is a valuable didactic element of Krzyś and Życzkowski's book is task number 14.7, in which the equivalent stresses calculated from various hypotheses are compared. Let us also recall the outstanding students of the Cracow School of the Material Effort. These are, among others: Jacek Skrzypek, Artur Ganaczarski, Halina Egner, Błażej Skoczeń, Kinga Nalepka. It is worth remembering that the geometric method of Mohr's envelope is a part of the very classical approach, yet developed in Hellenic and Medieval Mechanics, which mechanics, due to their tools, were called as "geometric mechanics". Also, the whole of Newtonian mechanics was expressed using geometric tools and descriptive constructions. It was not until Euler and Lagrange that analytical mechanics were introduced. Therefore, the dispute between Mohr and Huber was de facto a dispute over the superiority of "geometric mechanics" over "analytical mechanics". One of the formal obstacles to the construction of Mohr's circle is the lack of algebraic theorems corresponding to its graphic constructions.

We should remember Burzyński's question "Are Mohr's envelopes the full content of his effort hypothesis, or are they illustrations of some unknown hypothesis, which is a mere means of drawing, more or less successful?" (Burzyński, 1929a). Or speaking in a more contemporary language: which scientific tool energetic or geometrical is more correct.

2. The Tresca-Mohr geometrical approach – a critical review

2.1. Ultimate stresses approach

Let us note that the geometrical approach of Otto Mohr was based on "descriptive geometry techniques". Mohr used his representation of stresses σ_1, σ_2, σ_3 by circles to devise the material

effort hypothesis which can be adopted to various stress conditions. He assumed that it was the maximum shearing stress (Mohr, 1906). Under such circumstances, it is necessary to consider only the largest circle. Mohr called it the principal circle and suggested that such circles should be constructed when experimenting for each stress condition in which the failure occurred. In Fig. 4, the cast iron is tested to fracture in tension I, compression II and pure shear III, and three principal circles I, II, III are depicted. If there is a sufficient number of such principal circles, an envelope of these circles can be drawn. It can be assumed that for any stress condition for which there is experimental data, the corresponding limiting principal circle will also touch the envelope.

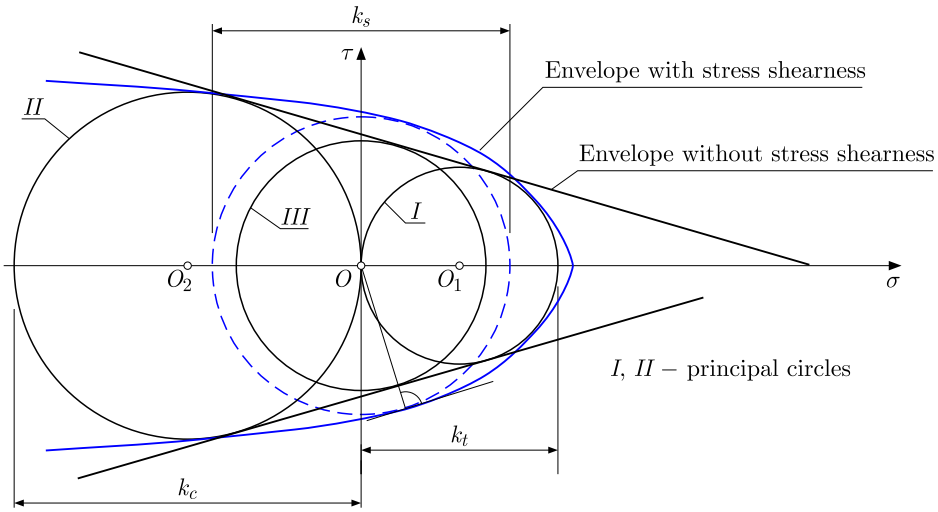


Fig. 4. Mohr limiting the yield surface as an envelope of principal circles: circle I with the radius k_t corresponds to uniaxial tension; circle II with the radius k_c corresponds to uniaxial compression; circle III with the radius k_s corresponds to pure shear. For determination of Mohr's plane, any unit vector $\mathbf{n} \cdot \mathbf{n} = 1$ must be taken, and two Mohr invariants are determined as follows: $\sigma = \sigma_{ij} n^i n^i$, $\tau^2 = \sigma_{ik}\sigma_{kj} n^i n^j - \sigma^2$

Note that one needs three cycles for a whole envelope. For example, when considering the cast iron, Mohr suggested that an envelope be taken as the two outer tangents to circles I and II. The limit strength in shear k_s is then found by drawing circle III, which has its center at O and is tangential to the envelope. It means, if k_t and k_c are values of ultimate strength in tension and compression, one can find from Fig. 4. that the ultimate strength in shear is $k_s = k_t k_c/(k_t + k_c)$, which agrees satisfactorily with the conical Drucker-Prager hypothesis (Drucker and Prager, 1948). In other case (broken line), when k_s is much greater than the arithmetic mean, the envelope must be constructed as an elliptical or paraboloidal section of a yield surface. For instance, in a case of St12T steel measured at temperature $T = 20°\mathrm{C}$, $k_s = 340\,\mathrm{MPa}$ (Dudda, 2020) but from $k_s = k_t k_c/(k_t + k_c)$ it follows that $k_s = 290\,\mathrm{MPa}$ – which means that the envelope cannot be a straight line. It is known from the literature (Kolupaev *et al.*, 2016; Nowak *et al.*, 2014; Olifieruk *et al.*, 2004; Skrzypek and Ganczarski, 2016; Dubey *et al.*, 2023) that measurement of k_s with appropriate accuracy is a difficult task. In needs realization of the state of pure shear where the principal stresses are $\sigma_3 = -\sigma_1$ and $\sigma_2 = 0$. Especially, making $\sigma_2 = 0$ is difficult. It means that a state of "stress shearness" is responsible for partial contribution of σ_2 into the material effort concept.

2.2. Geometrization by von Mises

The first continuation of geometrical approach was October 1913, when Richard von Mises (1913) wrote his paper on the foundations of modelling of plastic bodies. In that time, three-

-dimensional continuum mechanics was already a well-established field of knowledge, but Mises nevertheless decided to derive mathematical models "from the very beginning" in it, resulting in a fairly complete framing of the mechanics of plasticity within the framework of a new "plastic flow model" – today called the Mises flow model.

The question of "material effort" (German: die Anstrengungshypothesen) does not appear in von Mises's work, and only the "condition of plasticity" is discussed. At the beginning, Mises showed how to construct octahedral invariants of the stress tensor, which, following Mohr, are most often built in the principal axes, looking for the smallest and largest shear stress – Mises denotes them by τ_1, τ_2, τ_3, and defines accordingly

$$2\tau_1 = \sigma_2 - \sigma_3 \qquad 2\tau_2 = \sigma_3 - \sigma_1 \qquad 2\tau_3 = \sigma_1 - \sigma_2 \tag{2.1}$$

Further, von Mises emphasizes that an important invariant is the sum of squares of the octahedral stresses $\tau_1^2 + \tau_2^2 + \tau_3^2$, which today is nothing more than $0.5\sigma_{HMH}^2$. A key to Mises's work, the equation is expressed as

$$\tau_1^2 + \tau_2^2 + \tau_3^2 = \frac{1}{2}(\sigma_1^2 + \sigma_2^2 + \sigma_3^2) - \frac{1}{2}(\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_1) \tag{2.2}$$

which nowadays can be written as

$$\tau_1^2 + \tau_2^2 + \tau_3^2 = \frac{1}{2}\sigma_{HMH}^2 \tag{2.3}$$

It is an accidental result when both geometrical and energetic approaches lead to the same formulae. Next, von Mises raises the question of experimental motivations. He found an assumption that in the system of principal stresses, the sum of octahedral stresses at the point at which the elastic limit is reached is equal to zero

$$\tau_1 + \tau_2 + \tau_3 = 0 \tag{2.4}$$

Commenting on this condition, von Mises thanks Mohr, who was the first to analyse this condition at the time of reaching the elastic limit k_s

$$|\tau_1| \leqslant k_s \qquad |\tau_2| \leqslant k_s \qquad |\tau_3| \leqslant k_s \tag{2.5}$$

Von Mises showed it in the figure – in the axes of octahedral stress (Fig. 5), as a cube – when one cuts this cube with a plane, one gets a parallel hexagon in the cross-section.

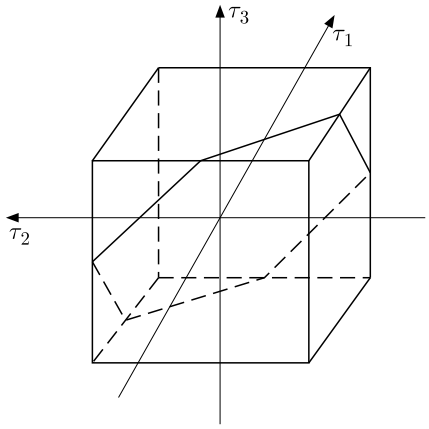


Fig. 5. Mohr's boundary surface in the form of an octahedral cube – interpretation of von Mises (1913, Fig. 4)

Then von Mises, in complete isolation from the energy methods and under the influence of Mohr's geometricism, proposed to modify the surfaces of elastic states in such a way as to break away from Mohr's previous condition and allow only the vertices of this diagram to come into play. In this way, it will be possible to bypass the assumption of invalidity of the central principal stress and, in addition, it will allow one to replace the diagram with a single simple solid without edges, e.g. of the "circumscribed circle" type. Thus, in place of the cube (Eq. (2.5)), von Mises proposed a cylinder as

$$\tau_1^2+\tau_2^2+\tau_3^2=2k_s^2 \qquad \text{or} \qquad \frac{1}{2}\sigma_{HMH}^2=2k_s^2 \qquad \text{or} \qquad \sigma_{HMH}^2=4k_s^2 \tag{2.6}$$

That is, taking into account that $k_s=k_r/2$, we have contemporary

$$\sigma_{HMH}^2-k_t^2=0 \tag{2.7}$$

And he added, "it is obvious that this condition is much simpler to describe analytically, because there are no ambiguities in the corners". In that time corners led to mathematical problems with which Mises's good friend [later wife] Hilda Geiringer had so much trouble (Geiringer and Prager, 1934).

Von Mises ignored the fact that condition (2.6) is physically "wider" (more capacious) than Mohr's condition (2.5). Let us emphasize that the spirit of this physics-free geometrization of Mohr's condition is still prevalent in the literature today and is even highly respected. Despite its apparent attractiveness, this approach should be considered a "wrong direction of research", leading astray determined by modern geometricism, which has dominated mechanics for years.

A generalization of Mises's geometrical reasoning assumes that one can raise the corner condition to any power, and then one has

$$|\sigma_1-\sigma_2|^m+|\sigma_2-\sigma_3|^m+|\sigma_3-\sigma_1|^m=mk_s^m \tag{2.8}$$

when $m=1$ and $m\to\infty$ we have the Tresca condition, $m=2$ the von Mises condition, at $m<1$ we have a concave boundary surface. The cross-sections in the deviator plane are shown in Fig. 6.

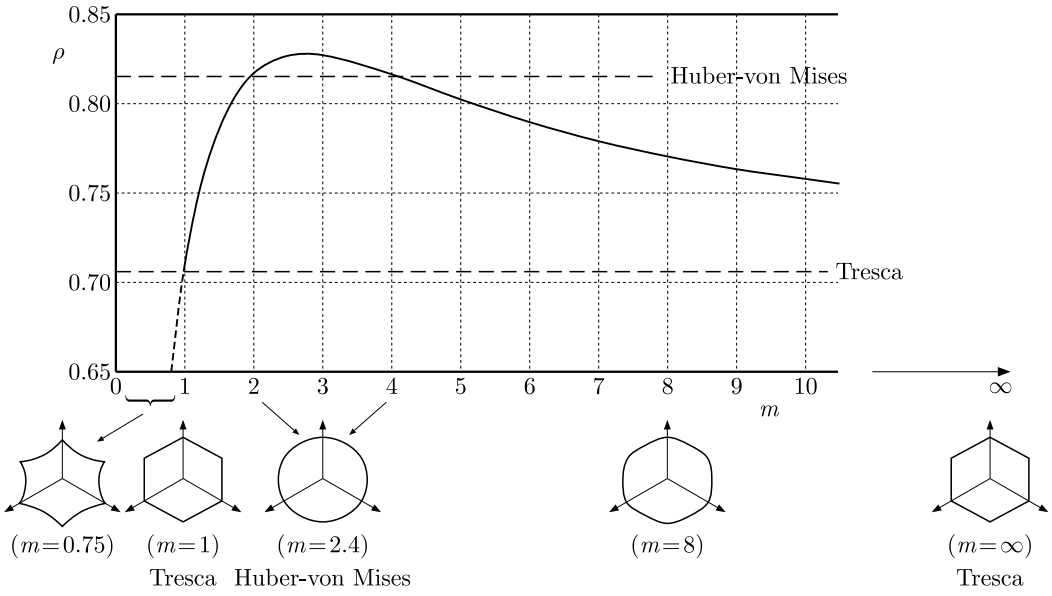


Fig. 6. Drawings of a generalized von Mises surface (Skrzypek, Ganczarski, 2015, p. 165), here $\rho=\sqrt{2J_2}$ is a cylindrical coordinate in the Haigh–Westergaard diagram that differs from the intensity $\sigma_i=\sqrt{3J_2}$

It is worth mentioning that the effect of partial appearance σ_2 within a yield function was experimentally discovered by Lode (1926) in combined tension/compression and internal pressure tests, which was a modernization that consequently omitting the role of σ_2 and was still consistent with the Mohr envelope concept. The second example was the Drucker two-parameter modernization of Mohr's condition, Drucker (1949)

$$\sqrt{J_2^3} - cJ_3 - 3k_s^3 = 0 \tag{2.9}$$

where J_3 changes the role of equal weight for σ_1, σ_2, σ_3 condition and provides a partial contribution of σ_2. Both models Eq. (2.8) and Eq. (2.9) have two unknown constants m, k_s and c, k_s, respectively. They can be expressed on the base of the same experimental data k_c and k_t. For instance, Hosford and Allen (1973) found that for bbc crystals the best fitted calibration was $m = 1.6$ and Cazacu and Barlat (2004) for bbc polycrystals calculated that for $k_c/k_t = 1/(1.28)$ the coefficient c in Eq. (2.9) was equal to 0.92.

2.3. The geometrical stress shearness

Having defined the stress triaxiality (normalized first invariant I_1), the stress intensity (von Mises norm of the second invariant J_2) is (Mises,1913)

$$\eta = \frac{\sigma_m}{\sigma_{HMH}} = \frac{I_1}{3\sqrt{3J_2}} \qquad \sigma_i = \sqrt{2J_2} = |\boldsymbol{\sigma}| \tag{2.10}$$

Now let us discuss the stress shearness concept treated to be a normalized third invariant of the deviatoric stress tensor J_3. The stress shearness is related with the old question discussed between German scientist concerning the appearance (or not) of the second principal stress σ_2 within some material effort hypothesis based on Tresca, Coulomb and, first of all, Mohr concepts of the "maximum shearness". It was shortly after The First World War, and after Guest work (Gao *et al.* 2011), when in many experiments, the question of priority between the two concepts: the maximum shearness and energy-based hypothesis, were researched and determined.

The general conclusion, achieved from Lode's experiments, was that the definition of the state of material effort as well as definition of yield surface are σ_2 dependent. It means, that σ_2 cannot be omitted, like in the Mohr-Coulomb criterion, or cannot be completely taken as in the energy-based Huber-Mises, but contribution of σ_2 should be made in an intermediate manner. It can be geometrically represented within the stress space when it is parametrized by Heigh-Westergaard cylindrical coordinates h, r, θ, where $h = I_1/\sqrt{3}$, $r = \sqrt{2J_2}$, and θ is the Lode angle within the range $0 \leqslant \theta \leqslant \pi/3$. The Lode angle is related with the normalized third invariant ξ (Bai and Wierzbicki, 2010) as

$$\xi = \cos(3\theta) = \cos\frac{\pi(1-(\theta)}{2} = \frac{3\sqrt{3}}{2}\frac{J_3}{\sqrt{J_2^3}} \tag{2.11}$$

determined in the range $-1 \leqslant \xi \leqslant 1$ with values $\xi = 0$ at pure shear $\theta = 30^\circ$, $\xi = 1$ at uniaxial tension $\theta = 0^\circ$, and $\xi = -1$ at uniaxial tension $\theta = 60^\circ$. It works with the Lode parameter (Lou *et al.*, 2014)

$$\mu = \frac{3\tan(\theta) - \sqrt{3}}{\tan(\theta) + \sqrt{3}} \tag{2.12}$$

and with the normalized Lode angle

$$\overline{\theta} = 1 - \frac{6}{\pi}\theta = 1 - \frac{2}{\pi}\arccos\xi \tag{2.13}$$

within the range $-1 \leqslant \overline{\theta} \leqslant 1$.

In general, a state of material effort should be a function of three parameters: stress triaxiality, stress intensity and stress shearness. The same concerns the limit surface – being for instance a yield surface, rupture surface, strength surface, etc. Any constructed yield surface can be defined in the Bai-Wierzbicki space spanned on (η, σ_i, ξ) or $(\eta, \sigma_{HMH}, \xi)$ parameters – such an approach is developed in the papers by Yoon *et al.* (2014), Lou *et al.* (2020), Wierzbicki *et al.* (2005).

2.4. The Lode angle corrections

Generally, the limit surface, especially for pressure sensitive materials, are developed in the three-dimensional space of main stresses σ_1, σ_2, σ_3 in the co-called Heigh-Westergaard cylindrical coordinates (h, r, θ). In this approach, as experimental results show, the important is shape of the deviatoric cross-section, which is perpendicular to the hydrostatic axis (Prager and Hodge, 1954).

This fact led to the proposal of many different shape functions $\mathcal{P}(\theta)$ that are only dependent upon the Lode angle $r = r_0\mathcal{P}(\theta)$ (Podgórski, 1984, 1985) where the value of $r_0 = r(\theta = 0)$. For instance, the most known Gudehus surface for pressure-insensitive materials (Gudehus, 1973), the shape function $\mathcal{P}(\theta)$ takes a very simple form $r^2 = r_0^2(1-\xi)$, where $\xi = \cos(3\theta)$. In the Lade and Duncan (1973) paper, the yield surface is based on one parameter shape function

$$\mathcal{P}(\theta) = \Big[\cos\Big(\frac{1}{3}\arccos\alpha\xi\Big)\Big]^{-1} \tag{2.14}$$

where $\alpha = \text{const}$, satisfying the condition $0 \leqslant \alpha \leqslant 1$. In such an approach, the well-known internal friction angle ϕ, appearing in the Coulomb-Mohr model $|\tau_n| + \sigma_n \tan\phi = c$, can be easily incorporated

$$\mathcal{P}(\theta) = \Big[\cos\Big(\frac{1}{3}\arccos\xi - \phi\Big)\Big]^{-1} \tag{2.15}$$

It means that, in general, the shape function can be dependent on two arbitrary constantans

$$\mathcal{P}(\theta) = \cos(30-\phi)\Big[\cos\Big(\frac{1}{3}\arccos\alpha\xi - \phi\Big)\Big]^{-1} \tag{2.16}$$

which is a generalization of research for large class of materials including metals, rocks, concrete and soils (Lade and Duncan, 1973; Willam and Warnke, 1974; Podgórski, 1985). Here, the angle ϕ has a simple physical interpretation of an internal friction angle.

2.5. Strength differential effect

Modelling of the material effort that takes into account the strength differential effect was a subject matter of many pioneers and veterans like Coulomb, Tresca, Lame, Clapeyron, de St. Venant, Navier. The first satisfying modification of the limit stresses hypotheses was proposed in 1856 year by R.W. Rankine as (Yu, 2004)

$$\sigma_1 \leqslant k_t \qquad \text{and} \qquad \sigma_3 \leqslant -k_c \tag{2.17}$$

where σ_1, σ_2, σ_3 are three principal values of the stress tensor – ordered from tension to compression. If $\varkappa = k_c/k_t = 1$ then this effort hypothesis reduces to the historical Galileo one. In the Heigh-Westergaard principal stress picture, the Rankine hypothesis is represented by six plains, double parallel, that build a cube with size $k_c + k_t$ located parallely to main axes – this model is pressure insensitive, of course.

Another hypothesis for the strength differential, but pressure insensitive material was thought as an extended Coulomb-Tresca by Otto Mohr in 1906 year (Mohr, 1906). He used his "geometrical approach" known as "curve envelope" (germ. Anstalung kurve) where the material in a

limit state undergoes slip due to maximal shear and, additionally, small cleavage due to presence of "longitudal stress" governed by $k_c - k_r$ difference. Analytically, the Mohr strength differential hypothesis reads

$$(\sigma_1 - \sigma_3)^2 + (k_s - k_t)(\sigma_1 + \sigma_3) = k_c k_t \tag{2.18}$$

In the Haigh-Westergaard principal stress picture this function forms a paraboloid moved to infinity, which possesses symmetry in the pressure axes. The mximum positive (tension) stresses in the Mohr paraboloid, located in the pressure axes, is obtained from the envelope with radius $a = 0.25(k_c - k_r)$ – in the case of $k_c = k_r$, the hypothesis turning into the classical Tresca-Guest one $\tau_{II} = 0.5(\sigma_1 - \sigma_3) = k_s$. Let us mention that the shear limit k_s is defined from the beginning as $k_s = 0.5k$ where $k = k_c = k_t$. The form of equation (2.18) implies that the limit on the pure shear is now defined to be

$$k_s = \frac{1}{2}\sqrt{k_c k_t} \tag{2.19}$$

It means that having measured three limits k_c, k_t, k_s and Mohr's relation (2.19), one can check correctives of hypothesis (2.18). Let us note, that Mohr's hypothesis (2.5) belonging to the so-called shearing-dominated mode of limit behaviour was also developed by Coulomb, Guest and others (Życzkowski, 1999; Hu *et al.*, 2017). This approach prefers octahedral state of stresses and τ_I, τ_{II}, τ_{III} shearing stresses that naturally leads to a non-smooth yield surface with edges. The edges are simple results of combination of non-continuous yield conditions, and the number of edges is a simple consequence of the number of field function. The fundamental are the six Tresca conditions

$$\tau_I \leqslant \mp k_s \qquad \tau_{II} \leqslant \mp k_s \qquad \tau_{III} \leqslant \mp k_s \tag{2.20}$$

written for no ordered main stresses σ_1, σ_2, σ_3. In this case, the number of edges (or pieces of limit surface) is six ($m = 6$). Similarly, for the Rankine hypothesis there is $m = 6$. But in the Galileo, Mariotte, and Ivliev hypothesis one has only three edges ($m = 3$), the Kolupaev hypothesis nine edges ($m = 9$), the Sokolovsky – eighteen edges ($m = 18$) (Kolupaev *et al.*, 2016).

2.6. Pressure sensitiveness within the maximum shear stresses treatment

A similar conclusion concerning the two-parameter shape function can be obtained by a direct analysis of the yield surfaces based examples on J_3 invariant, for instance, for pressure-insensitive $I_1^3/J_3 = k$ (Lade and Duncan, 1973); $J_2^3 + cJ_3^2 - k^6 = 0$ (Drucker, 1949); $J_2^{3/2} - cJ_3 - k^3 = 0$ (Cazacu and Barlat, 2004) as well as for pressure sensitive materials $a(bI_1^6 + 27J_2^3 + cJ_3^2)^{1/6} = 0$ (Gao *et al.*, 2011); a$[bI_1 + (J_2^{3/2} - cJ_3)^{1/3}] = 0$ (Yoon *et al.*, 2014) and $aI_1^2 + bJ_2 + cJ_3^{2/3} - 1 = 0$. Going to show this, let note that for isotropic, pressure-insensitive materials the above limit conditions can be written in terms of the second and third invariants as

$$f(J_2, J_3, k, c, n) = \sqrt{J_2^{3n}} - cJ_3^n - k^{3n} = 0 \tag{2.21}$$

There are only three parameter (k, c, n) yield surfaces – for $n = 1$, one obtain the Cazacu and Barlat (2004) condition, however, for $n = 2$ we have the condition of Drucker (Drucker,1973), yet a more simple example $J_2^3 - k^6 = 0$ was considered by Reuss (1928). Let us observe that for even exponents $n = 2, 4, 6, \ldots$ the yield condition predicts the same values of the yield stress in compression and tension $k_c = k_t$. Differently, for odd exponents $n = 1, 3, 5, \ldots$ the yield

condition provides different values of the yield stress, exhibiting the SD-effect. Next, reorganizing Eq. (2.21), one obtains

$$f(J_2, J_3, k, c, n) = \sqrt{J_2^{3n}}\Big[1 - c\Big(\frac{J_3}{\sqrt{J_2^3}}\Big)^n\Big] - k^{3n} = \big(\frac{1}{2}r\big)^{3n}\mathcal{P}'(\theta) - k^{3n} = 0 \tag{2.22}$$

where, now, the generalized shape function $\mathcal{P}'(\theta) = 1 - c_1\xi^{3n}$, $c_1 = c(2\sqrt{3}/9)^n$. It is easy to observe that condition, Eq. (2.22), has the same physical meaning as $f = r - k = r_0\mathcal{P}(\theta) - k = 0$ in the classical Coulomb-Mohr-like conditions, but the shape function $\mathcal{P}'(\theta)$ is only one-parametrical, whereas $\mathcal{P}(\theta)$ is two-parametrical.

There is in the literature unclearness concerning pressure sensitive materials. Namely, some authors, like Mirone and Corallo (2010), take into account the expression I_1J_2/J_3, which cannot lead to simple splitting of distortional and volumetric effects. But in the case when such splitting can be postulated, a yield function, written within the cylindrical (h, r, θ) coordinates has a form

$$f(I_1, J_2, J_3) = \Big(\frac{1}{2}r\Big)^{3n}\mathcal{P}'(\theta) - k_1h\mathcal{P}''(\eta) - k_2^{3n} = 0 \tag{2.23}$$

Here $\mathcal{P}'(\theta)$ is interpreted as the circumferential shape function and $\mathcal{P}''(\eta)$ is longitudinal along the hydrostatic axis $\sigma_1 = \sigma_2 = \sigma_2$ shape function, depending on the stress triaxiality $\eta = \sigma_m/\sigma_{HMH}$ that change with $h = I_1/\sqrt{3}$. In the simplest form this condition is a three-parametrical one. However, when the invariant J_2 is removed from considerations – like in $I_1^3/J_3 - c = 0$ – the above form should be modified to:

$$f(I_1, J_3) = h\mathcal{P}''(\eta)\mathcal{P}'''(\theta) - k = 0 \tag{2.24}$$

where both shape functions are multiplied. The similar form of the yield surfaces for a pressure sensitive material has been proposed by Barlat *et al.* (2003). In general, such a yield function depends only on I_1 and $J_2/\rho(\theta)$, where $\rho(\theta)\mathcal{P}^{-1}$ is the circumferential shape function – the deviatoric sections in planes with constant I_1 (or h) look similar but not necessarily coincide. The deviatoric sections are circular when the shape function does not depend on the Lode angle. Then $\rho(\theta) = 1$. Since θ varies between $-\pi/6$ and $\pi/6$ a shape function should contain periodic curves with the period $2\pi/3$. Some application of the $\rho(\theta)$ shape function has been applied to bone fracture mechanics by Pietruszczak *et al.* (1999) in the form

$$f = a_1\frac{\sqrt{3J_2}}{\rho(\theta)k_c} + a_2\frac{3J_2}{(\rho(\theta)k_c)^2} - \big(a_3 + \frac{I_1}{k_c}\big) = 0 \tag{2.25}$$

where a_1, a_2, a_3 are dimensionless constants and k_c is the yield strength in compression. This model can be prolonged onto anisotropic geomaterials (Piertuszczak and Mróz, 2001).

Now the examples of pressure sensitive materials like pressure modified Tresca

$$|\tau_n| = c - \frac{1}{3}\epsilon(\sigma_1 + \sigma_2 + \sigma_3) \tag{2.26}$$

or pressure modified Coulomb-Mohr

$$|\tau_n| + \sigma_n\tan\phi = c - \frac{1}{2}\epsilon(\sigma_1 + \sigma_2 + \sigma_3) \tag{2.27}$$

can be reformulated in terms of the shape function $\rho(\theta) \equiv \mathcal{P}^{-1}$. For instance, the pressure modified Coulomb-Mohr yield surface takes the following form (Barlat *et al.*, 1991)

$$\sqrt{J_2}(3\cos\theta - \sqrt{3}\sin\phi\sin\theta - \epsilon I_1\sin\phi - 3c\cos\phi = 0 \tag{2.28}$$

where ϵ is the pressure coefficient, ϕ is the friction angle and c is cohesion.

It has, generally, been assumed that the state of the maximum stress shearness does not need appearance of σ_2 in the mechanical or thermal effort modeling. Numerous continua, like dry sand, obeys the Tresca or Coulomb-Mohr conditions, in which failure does not need any previous plastic or inelastic deformation. When the stress shearness dominates over the stress intensity or stress triaxiality, the slip in the σ_1-σ_3 plane is independent of σ_2. Bur in many crystalline metals, when plastic flow occurs firstly, motion of dislocations on available slip planes within each individual crystal possesses at least five independent slip systems which are required in each grain if it is to undergo shape changes imposed by the crystals around it (Pęcherski, 1998). This means that even if the macroscopic plastic deformation is apparently confined in σ_1-σ_3 plane within a high proportion of individual grains, the plastic flow must be occurring on slip planes inclined to this, i.e. in slip systems driven by the intermediate main stress σ_2. It is therefore, physically reasonable that all three principal stresses should appear in form of a limit surface: proportional, yield, strength, fracture, and so on. But between the full appearance of σ_2 and the omitting of σ_2 – there is an open domain with a different level of the stress shearness which is represented via the Lode parameter. In another words, the passage from null-dependence to full-dependence of σ_2 within limit surface formulae can be realized continuously by using the Lode parameter μ (or the Lode angle θ, the normalized Lode angle $\overline{\theta}$, the normalized third invariant ξ).

2.7. An extension of the Mohr geometrical treatment of anisotropic materials

The Tresca and Coulomb-Mohr concept of maximum shear stresses can be applied also to anisotropic materials like wood, stones, clays and polycrystalline metals. To describe plastic anisotropy of rolled metal sheets, Hill (1948) developed an extension of the Mises-Hosford isotropic condition, Eq. (2.2), to the anisotropy case – this solution was based on 6 unknown coefficients. The rolling direction, the long transverse direction and short transverse direction are three main directions where the unknown parameter should be calibrated by the experimental data – generally one can obtain three-time more data than for isotropic bodies. Up to 1948 year, several methods had been proposed (Baltov and Sawczuk, 1965; Życzkowski, 1981; Mróz, 1967). A review of different solutions can be found in the paper by Oana Cazacu and Barlat (2004). They proposed, finally, to explore the technique of linear transformations $\sigma'_{ij} = L_{ijkl}\sigma_{kl}$ and $\sigma''_{ij} = C_{ijkl}\sigma_{kl}$ to define an anisotropic yield surface

$$|\sigma'_1 - \sigma''_2|^m + |\sigma'_2 - \sigma''_3|^m + |\sigma'_3 - \sigma''_1|^m = k_s^m \tag{2.29}$$

that needs fourteen parameters to be calibrated (six L_{ijkl}, six C_{ijkl} and m, k_s). For a pressure sensitive material the above criterion was extended by (Kuroda and Kuwabara, 2002) as

$$(1-c)\left(|\sigma'_1 - \sigma''_2|^m + |\sigma'_2 - \sigma''_3|^m + |\sigma'_3 - \sigma''_1|^m\right) + c\rho\left(|\sigma_1|^m + |\sigma_2|^m + |\sigma_3|^m\right) = k_s^m \tag{2.30}$$

where one additional constant c appears, and $\rho = 3^m(2^{m-1}+1)^{-1}$. The whole family of different dedicated yield surfaces is coming from the above condition, but calibration of such great numbers of parameters needs many experimental data connected with crystal plasticity calculations (Hu and Yoon, 2021).

From another point of view, the maximum shear approach cannot take into account the Lodge parameter or Lode angle directly. Thus, a way by using J_3 invariant is more promising. Cazacu and Barlat (2004) proposed a manner of introducing anisotropy into J_2 (6 parameters) and J_3 (11 parameters) invariants. A quite similar approach was developed by Yoon *et al.* (2014) where the concept of linear transformation was applied to all three invariants independently. It is the first invariant I_1*, second invariant J'_2 and third invariant J''_3, which are based on the transformed Cauchy stress tensor $\boldsymbol{\sigma}^* = \mathbf{H}\boldsymbol{\sigma} = h_1\sigma_1\mathbf{e}_1 \otimes \mathbf{e}_1 + h_2\sigma_2\mathbf{e}_2 \otimes \mathbf{e}_2 + h_3\sigma_3\mathbf{e}_3 \otimes \mathbf{e}_3$, $\boldsymbol{\sigma}' = \mathbf{L}'\boldsymbol{\sigma}$, $\boldsymbol{\sigma}'' = \mathbf{L}''\boldsymbol{\sigma}$ and its deviator. The Yoon model needs to calibrate $3+6+6=15$ unknown parameters of $\mathbf{H}$,

$\mathbf{L}'$, $\mathbf{L}''$. Among these anisotropic parameters, eight of them are related with in-plane properties while the other four parameters are used to describe the through-thickness behaviour of the metal. Finally, the yield function for pressure sensitive anisotropic metals reads in the form of (Yoon *et al.*, 2014)

$$I_1^* + \sqrt[3]{\sqrt{J_2'^3} - J_3''} = 1 \tag{2.31}$$

This yield function is able to describe: the stress triaxiality, the stress intensity and the stress shearness (Lode parameter) effects within the scope of different anisotropy of metals (Nixon *et al.*, 2010).

3. Energy-based limit approach

3.1. Pioneering steps – Maxwell, Beltrami, Huber, Schleicher, Burzyński

Recall, that the "father" of energy-based approach is James Clerk Maxwell who in 1856, in the letter to William Thomson, without introducing a concept of "distortional energy" as a part of elastic strain energy, wrote that it was the best candidate to be a measure of material effort (Maxwell, 1856; von Helmholtz, 1903; Pęcherski, 2008; Altenbach, 2010). Modern reconstruction of Maxwell's energy-based approach was worked out by Rychlewski (2011).

In the complete mathematical form, this energy-based approach was initiated by Beltrami (1885) in the form $W = \Phi \leqslant K$, where $\Phi = [(1-2\nu)/3E]I_1^2 + [(1+\nu)/3E]3J_2$ is some elastic deformation energy (volumetric density) describing a state of material effort, and K is critical value of this energy. Beltrami firstly found how critical energy K depended on the uniaxial yield k_t or torsion k_s (Becchi, 1994). This approach also opens a possibility of using many other experimental data like Vickers hardness, Charpy critical energy (fracture toughness), cohesiveness critical energy, and so on (Orłowski *et al.*, 2020, 2013).

Explicitly, the concept of "specific work of strain" Φ (internal energy, or specific work of stress) as a measure of material effort (germ. *Die Anstrengung*) was developed by Huber (1904), who was able to introduce a notion of "equivalent stress" (or reduced stress σ_{eq}). Also, he proposed the first picture of a limiting surface within the space of three principal stresses (Huber, 1904, his Fig. 2). Next, by comparison of the Beltrami and de Saint Venant criterions, Huber performed an exhausting discussion on the role of the Poisson ratio coefficient and its appearance in energy-based hypothesis. Next, he found that the Maxwell-Helmholtz decomposition of strain energy into purely volumetric and distortional parts $\Phi = \Phi_\nu + \Phi_f = [(1-2\nu)/3E]I_1^2 + [(1+\nu)/3E]3J_2$ could be useful in further research, and discussed a case when volumetric energy Φ_ν was negligible. Finally, Huber proposed a new combined criterion: if $p \geqslant 0$ then $\Phi \leqslant K$ and $p \leqslant 0$ then $\Phi_f \leqslant K_f$ – this condition leads to an elliptical-cylindrical yield surface and nowadays is called "the Beltrami-Huber combined condition".

3.2. Combined Huber's hypothesis

But presently and literally, when we go into the details of this reasoning, we have three hypotheses to offer: Beltrami's (ellipsoid), Huber's (part ellipsoid, part cylinder), and Mises's (all cylinder) (Table 1).

In the case of Beltrami (1885) measure, both components of energy are always used. In the case of the von Mises hypothesis, only the form energy is always used. In contrast, in the case of Huber's hypothesis, the energy form is used in the compressive stress region, and in the tensile region, the sum of volumetric and distortional internal energy $\Phi = \Phi_\nu + \Phi_f$ is used. In Fig. 7, the Beltrami, Huber, and Mises stress measures in the space of the principal axes of the stress

Table 1. Primary diagram of energy based hypothesis

Beltrami (1885)	$\Phi \leqslant K$ always	Fig. 7a
Huber (1904)	$\Phi_\nu + \Phi_f \leqslant K$ when $\sigma_1 + \sigma_1 + \sigma_3 > 0$ $\Phi_f \leqslant K$ when $\sigma_1 + \sigma_1 + \sigma_3 < 0$	Fig. 7b
Mises (1913)	$\Phi_f \leqslant K$ always	Figs. 7a,b

tensor, where the dominant is the hydrostatic axis on which the first invariant of the tensor, are depicted. For a constant critical energy K, the areas determined by the condition $f = W - K = 0$ can be obtained. In the case of the Beltrami measure, it is an ellipsoid hooked at the centre of the volume, in the case of von Mises, it is an infinite cylinder for extensions and compressions, and in the case of the Huber measure, it is the Misesian cylinder in the compression part and the Beltrami ellipsoid in the stretch part.

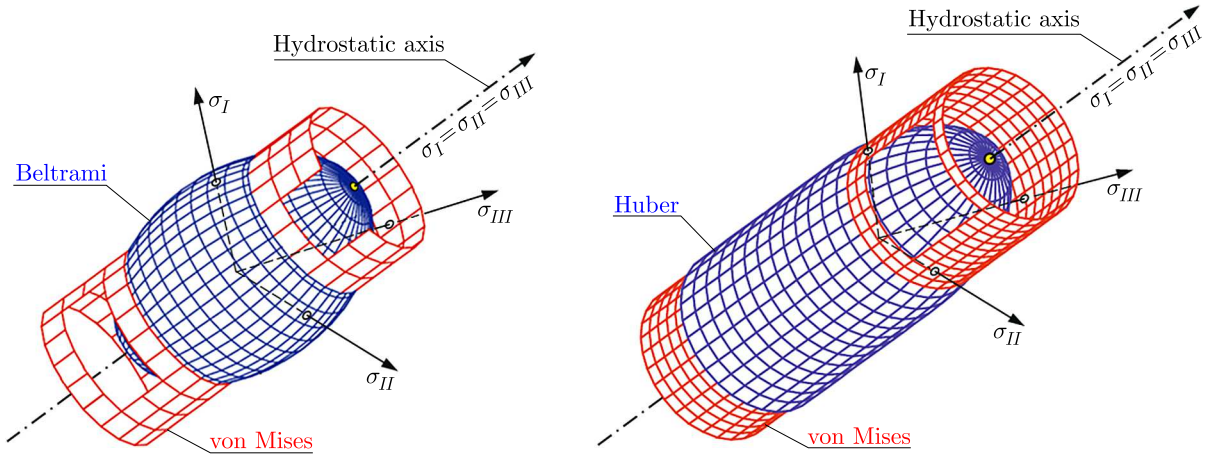


Fig. 7. Yield limits surfaces of (a) Beltrami (1884), (b) Huber (1904) and (a), (b) Mises (1913)

3.3. Revalorization of Huber's Combined Hypothesis

It was in (Schleicher, 1926) that the combined concept of Huber's effort measure was revalued for the first time after twenty-two years (Fig. 7). Schleicher, speaking about the combined Huber hypothesis, used two definitions of Huber's reduced stress, which is denoted by an additional H-index

$$\sigma_H = \begin{cases} \sqrt{2E\Phi} & \text{when} \quad I_1 < 0 \\ \sqrt{6G\Phi_f} & \text{when} \quad I_1 > 0 \end{cases} \tag{3.1}$$

Two years later, in 1928, it was criticised for the first time by Włodzimierz Burzyński (Burzyński, 1928). Huber's combined hypothesis has number C2 in his classification. The disadvantage of Huber's combined hypothesis is precisely its heterogeneity – because the states of pure stretching, two-way stretching and three-way stretching are described with full energy, and the states of single, double and three-way compression are described by form energy. Only the shear state satisfies both intervals. Burzyński, further wrote the boundary surface as shown in Fig. 8.

It shows that

$$k_s = \frac{k_t}{\sqrt{2(1+\nu)}} = \frac{k_c}{\sqrt{3}} \tag{3.2}$$

which means that the boundaries of k_t and k_c differ slightly. Using the Mohr envelope method, Burzyński showed a drawing of Huber's hypothesis (for $\nu = 1/4$) in the τ-σ plane – which

$$\left.\begin{array}{c} \dfrac{1}{2(1+\mu)}(\sigma_1^2+\sigma_2^2+\sigma_3^2)-\dfrac{\mu}{1+\mu}(\sigma_1\sigma_2+\sigma_2\sigma_3+\sigma_3\sigma_1)=k_s^2 \\ \text{dla: } \sigma_1+\sigma_2+\sigma_3\geq 0 \text{ i} \\ \dfrac{1}{3}(\sigma_1^2-\sigma_2^2+\sigma_3^2-\sigma_1\sigma_2-\sigma_2\sigma_3-\sigma_3\sigma_1)=k_s^2 \\ \text{przy: } \sigma_1+\sigma_2+\sigma_3\leq 0 \end{array}\right\} \text{(C2)}$$

Fig. 8. Combined von Huber hypothesis (Burzyński, 1928), (μ is $1/\nu$, where ν is the Poisson coefficient)

nowadays is called "the Burzyński plane". It is a drawing of a rotational ellipsoid (in the Beltrami part) and a rotating cylinder (in the von Mises part) – the intersection of the two solids takes place in the planes $\sigma_1+\sigma_2+\sigma_3=0$. For $\nu=0$, the Mises cylinder ends in a sphere, for $\nu\to 0.5$ the von Huber hypothesis becomes the von Mises hypothesis. Burzyński marks the latter as C3 and considers it separately.

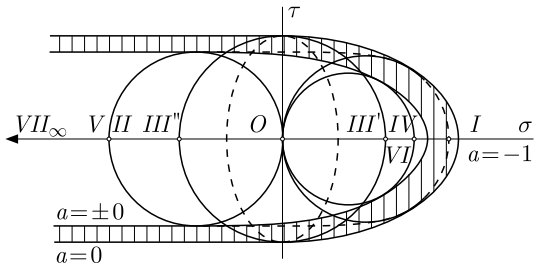


Fig. 9. Huber's cylinder-ellipsoid in the Burzyński plane (Burzyński, 1928). The Lode parameter is denoted here by letter a. By letters $I, II, III, \ldots, VII$ limit states $k_t, k_c, k_s, \ldots, k_{ccc}$ are denoted

Recently, researchers such as Altenbach, Bolchoun, Kolupaev (2016) have proposed to include Huber's hypothesis in the group of combined hypotheses. Holm Altenbach emphasizes additional advantages of Huber's hypothesis over Beltrami and Mises.

Reassuming, the most frequently used in the literature model of the J_2-plastic flow in fact possesses four different theoretical fundaments: these are (a) distortional energy source (Maxwell, 1856; Beltrami, 1885; Huber, 1904; Hencky, 1925; Schleicher, 1926; Burzyński, 1928, 1929a,b; Torre, 1947; Życzkowski, 1999), (b) von Mises norm of tensor – the second invariant of the stress deviator (Mises, 1913; Reuss, 1930; Kłębowski, 1958), (c) the stress intensity as a mean shear stress (Novozhilov, 1952), (d) the octahedral shear stress (Nadai, 1927; Zawadzki, 1954).

Only accidentally four different concepts have the same mathematical description. For instance, for the Nadai circular cone one finds

$$f=\tau_{oct}^{-2}-\frac{2}{9}(3C_0\sigma_{oct}-C_1)=0 \tag{3.3}$$

Coming from another line of reasoning (Nadai, 1927), where τ_{oct}, σ_{oct} are octahedral shear and normal stresses and C_0, C_1 are limiting constants, it describes the same cone as the Drucker-Prager $f=I_1-3a_0+9a_1(6J_2)^{-2}=0$ (Drucker and Prager, 1948).

It is worth noting that the Helmholtz decomposition $\Phi=\Phi_\nu+\Phi_f$ also leads to decomposition of critical energy into $K=K_\nu+K_f$. Beltrami (1885) would use the word "resilience" to denote the work necessary to be done on a body to overcome its elastic forces. The volumetric (cubical) resilience K_ν is a measure of the work necessary to be expended in compression in order to increase the density permanently. Distortional resilience K_f is the work required to be expended in pure distortion in order to produce a permanent change of form in the element – it is some limit that Φ_f can reach (Altenbach *et al.*, 2014).

3.4. Poisson coefficient dependent Schleicher's hypothesis

After many years, the concept of energy-based hypothesis of material effort was undertaken again by Schleicher (1926) who proposed to extend the Beltrami-Huber criterion $\Phi \leqslant K$ to pressure-sensitive materials: as a linear function of pressure $\sigma_{BH} = \sqrt{2E\Phi} = \sqrt{2EK}(1 - \epsilon p)$, or a parabolic function of pressure $\sigma_{BH} = \sqrt{2E\Phi} = \sqrt{2EK}(1 - \epsilon_1 p^2)$ – unfortunately, the Poisson coefficient was present in those solutions. Some response on Schleicher's paper was Burzyński's dissertation (see: Burzyński, 1928 – received to print December 13, 1927 in Lwów [now Lviv, Ukraine]), as well as in the papers (Burzyński, 1929a,b). Probably, under strong critics of Burzyński, F. Schleicher changed his measure of material effort and in the next paper, published on 13th of April, 1928 (Schleicher, 1928) he replaced σ_{BH} with the total elastic strain energy Φ by volumetric density of elastic distortion $\sigma_H = \sqrt{2E\Phi_f} = \sqrt{2EK}f(p)$. Working on brittle materials like marble, limestone, sandstone, Schleicher, finally fitted the experimental data to the cone in the $\sigma_H - p$ coordinates plane, what, in fact, was the first application of a model introduced by Drucker and Prager (1948). Note that the denotations: Beltrami-Huber σ_{BH} and Huber σ_H are originally coming from Schleicher (Dudda, 2021).

3.5. Burzyński pressure-size function hypothesis

Next, Burzyński proposed a modification for pressure-sensitive materials where the Poisson coefficient does not appear. Generally, he proposed a "size function" η_ν correcting the contribution of volumetric energy – contemporary, it is nothing else as introducing the stress triaxiality effect into the energy-based approach. It is one of the main achievements of Burzyński, since he solved a crucial question in such a way which does not disturb the scientific power of energy-based approach (Altenbach, 2010). The "size function" η_ν in the hypothesis making the volumetric energy "partially present", if $\eta_\nu = 0$ then the material is pressure-insensitive, if $\eta_\nu = 1$ then the material is fully pressure-sensitive.

Mathematically, the Burzyński hypothesis can be written as follows (Burzyński, 1928; Pęcherski, 2008; Pęcherski *et al.*, 2014)

$$\eta_\nu \Phi_\nu + \Phi_f = K \tag{3.4}$$

where a particular form of pressure dependency of the function η_ν was assumed as $\eta_\nu = \omega + \delta/p$ and ω, δ are unknown parameters. The core of Burzyński's idea is to express three unknown parameters ω, δ, K in terms of tripled of material limit constants k_c, k_t, k_s which are known from an experiment of uniaxial compression, tension and simple shear. The other forms of the size function could also be considered in order to find another state of the material like brittle and ductile failure or continuous damage. Then other experimental limits data can be used: bi-axial compression and tension k_{cc}, k_{tt}; tri-axial compression and tension k_{ccc}, k_{ttt}, and so on (Kordzikowski and Pęcherski, 2010).

3.6. The J_3 effect within the energy-based framework

Let us discuss shortly the possibility of appearance of J_3 invariant in the energy-based approach. It is a well known fact that, for common materials, the invariant J_3 (or the Lode parameter) does not appear in the elastic strain energy expression. The fundamental example is density of elastic strain energy for the so-called Hooke elastic material $\Phi = [(1-2\nu)/3E]I_1^2 + [(1+\nu)/3E]3J_2$. If such energy is a starting point to develop any independent approach, nowadays called an "energy-based" one, one can ask about a physical foundation of this. The rational arguments for defending the generality of energy-based formulations of yield conditions are the following.

Firstly, one must recall that from thermodynamical point of view (Casey and Sullivan, 1985; Enger *et al.*, 2018; Banaś and Badur, 2017) the internal energy of deformable continua depends on all intensive state parameters, not only on the elastic one. For instance, according to Taylor and Quinney (1935), the plastic deformation is also partially involved in energy Φ. According to the modern energy-based treatment, developed within the frame "strain energy density" (SED) (Łagoda and Ogonowski, 2005) the internal energy can be a function even of different internal fracture mechanisms and so on.

Secondly, papers concerning ductile failure and damage, e.g. (Bai and Wierzbicki, 2008) state that the stress triaxiality and stress shearness take part gradually, starting from initial yield surfaces completely based on classical elastic solutions. At a very low stress triaxiality state, the stress shearness (Lode parameter) has an important influence on failure behaviour. It means, that in the state of constant stress intensity, it is difficult to realize, and sophisticate damage and failure models should incorporate not-complete state of three-dimensionality of the stress state – this non-completeness, governed by the Lode parameter, is called here the stress shearness.

Speaking in terms of mathematics, the reason for adding the Lode parameter is that multiple stress state with different principal stress values can result in the same value of the stress triaxiality. Thus the stress intensity and stress triaxiality alone cannot completely describe the three-dimensional stress state and its effect on the fully developed state of damage and failure (Gao *et al.*, 2009, 2011). That means that the dimensionless invariants I_1, J_2, J_3 should be accounted in the models with plastic isotropic and kinematic hardening. On the other hand, Lou and Yoon (2017) showed that the stress shearness (the Lode parameter) had only a marginal effect on the macroscopic yield surface, whereas its influence on anisotropic damage is remarkable. Therefore, in a model that describes the starting moment of plasticity influencing the stress shearness can be neglected (Kowalczyk *et al.*, 2003).

From the above reasoning it follows that there is a physically acceptable manner of introducing J_3 into energy-based treatment. The examples can be found in Frąś *et al.* (2010, 2014), Frąś and Pęcherski (2010), Nalepka and Pęcherski (2002, 2003).

3.7. The Burzyński-Pęcherski hypothesis taking into account the third invariant

Thus, Pęcherski and his co-workers proposed to introduce the stress shearness effect into the energy-based approach and the Burzyński measure of material effort (Pęcherski *et al.*, 2011; Nowak *et al.*, 2011). Trying to find the influence of the Lode parameter, he proposed to introduce a some shape function η_f to made a partial (variable) contribution of the energy density of distortion. It means, that the extended material effort hypothesis in a case of variable energy, both with the stress triaxiality effect (volumetric energy) and the stress shearness effect (distortional energy) reads

$$\check{\eta}_\nu \Phi_\nu + \check{\eta}_f \Phi_f = K \tag{3.5}$$

where $\check{\eta}_\nu$, $\check{\eta}_f$ denote the size function and the shape function, respectively. Using the definition of Φ_ν, Φ_f in terms of invariants I_1, J_2, condition Eq. (3.5) can be expressed in the following way (Pęcherski *et al.*, 2011)

$$\eta_\nu(I_1)(3I_1)^2 + \eta_f(J_3)3J_2 = K^2 \tag{3.6}$$

The shape function (or the Lode influence function) can now be proposed in a mathematical form similar to $\mathcal{P}(\theta)$ in the maximum shearness approach. Taking, for instance, the two parameter Podgórski shape function (Podgórski, 1984)

$$\eta_f(J_3) = \frac{1}{\cos\left(\frac{\pi}{6} - \beta\right)} \cos\left[\frac{1}{3}\arccos(\alpha\cos(3\theta)) - \beta\right] \tag{3.7}$$

and two parametrical size function

$$\eta_\nu(I_1) = \omega + \frac{3\delta}{I_1} \tag{3.8}$$

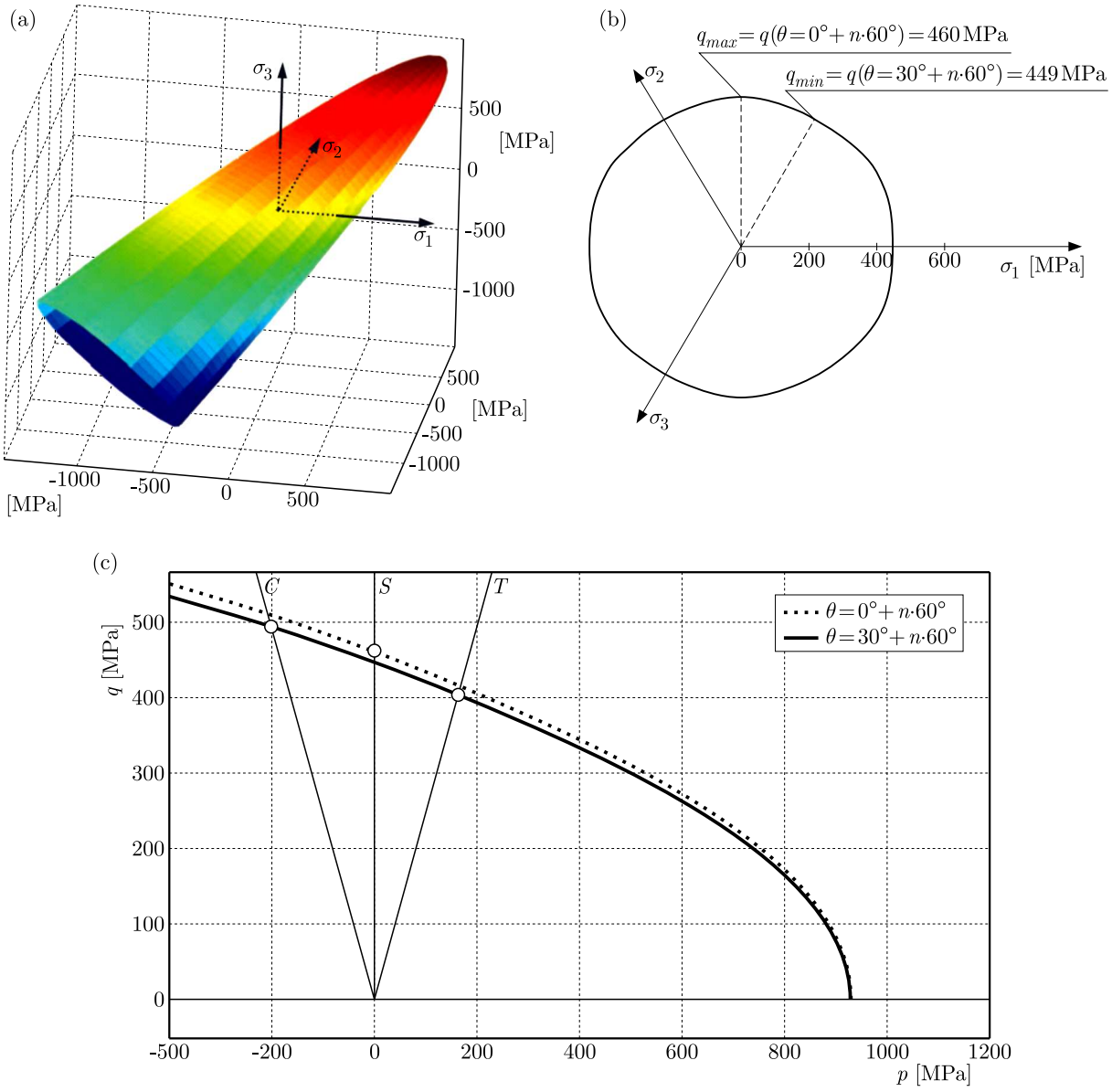


Fig. 10. The Burzyński-Pęcherski hypothesis (Pęcherski *et al.*, 2011): (a) a view to the Heigh-Westergaard cylindrical coordinates, (b) the perpendicular cross-section, (c) the Burzyński-Pęcherski hypothesis on the q-p plane

Pęcherski *et al.* (2011) obtained a five parametrical $(\alpha, \beta, \omega, \delta, K)$ yield surface, Eq. (3.5). The surface is paraboloidal and the cross-section parallel to the octahedral plane indicates that the Lode angle dependence has a hexagonal character. These five unknown parameters are to be expressed by five experimental data k_t, k_c, k_s, k_{cc}, k_{tt} – the procedure of fitting the paraboloidal yield surface, Eq. (3.6), has been made by using the Levenberg-Marquardt algorithm. Also Frąś *et al.* (2010) and Frąś and Pęcherski (2010) discussed several criteria developed within the energy-based approach in comparison with the Burzyński criterion.

The energy-like approach has some limitations (see Mucha *et al.*, 2018; Mróz and Seweryn, 1998). Let us discus a most important one coming from the fact that elastic storage energy Φ

not always is expressed in terms of J_2 invariant. For instance, the elastic energy of the gum metal is a function only of J_3 and σ_m invariants (Kowalczyk-Gajewska *et al.*, 2019).

3.8. The material anisotropy

Richard von Mises in June 1928 turned his point of view: from the Tresca-Mohr geometrical envelope approach to the energy-based approach. His aim was to construct a yield surface for generally anisotropic crystals (Mises, 1928). His mathematical concept lies on replacing $J_2 = s_{ij}s_{ji}/2$ invariant into a "weighted invariant" $J_2' = s_{ij}k_{ijkl}s_{kl}/2$ – then the anisotropic cylindrical surface is described by a dimensionless function $f = J_2' - 1 = 0$. By using many symmetry arguments, von Mises was able to reduce the number of independent coefficients of the compliance tensor k_{ijkl} to 15. It means that von Mises turn his $\tau_I^2 + \tau_{II}^2 + \tau_{III}^2 - k_s^2 = 0$ argumentation, called by him "die Guest-Mohrsche Bedingung", into energy-based argumentation.

In Section 3, von Mises, recalled the pioneering Beltrami, Huber, Schleicher, Burzyński papers, in the context of separation of the strain energy density into volumetric and distortional parts for anisotropic bodies. He proposed to start from $2\Phi = \boldsymbol{\sigma}\boldsymbol{\varepsilon} = (s\mathbf{I}+\mathbf{s})\cdot(e\mathbf{I}+\mathbf{e})$ decomposition of both tensors into: axiators and deviators (Mises, 1928, p. 170).

Next, Reuss (1930) improved the von Mises idea finding a more effective method of transforming the Cauchy stress tensor from Cartesian $\boldsymbol{\sigma} = \sigma_{ij}\mathbf{e}_i \otimes \mathbf{e}_j$, $i,j = x,y,z$ into the principal coordinates $\boldsymbol{\sigma} = \sigma_\chi \mathbf{e}_\chi \otimes \mathbf{e}_\chi$, $\chi = 1,2,3$, where he reduced the number of independent coefficients to be 3 plus 3 – Euler's angles that describe the rotation matrix between the principal axes of strain and stress. Few years later, Theodore Lehmann, proposed the yield surface of a generally anisotropic material in the form of a polynomial (Lehmann, 1964)

$$f = k + k_{ij}s_{ij} + k_{ijkl}s_{ij}s_{kl} + k_{ijklmn}s_{ij}s_{kls_mn} + \ldots \tag{3.9}$$

Here appears some analogy to the third anisotropic invariant J_3'. If it is possible to take into account the anisotropy induced by kinematic hardening of an initially isotropic material, the above polynomial can take a form (Baltov and Sawczuk, 1965)

$$f = \frac{1}{2}s_{ij}s_{ij} - s_{ij}\alpha_{ij} + A\alpha_{ij}\alpha_{kl}(s_{ij} - \alpha_{ij})(s_{kl} - \alpha_{kl}) + \ldots \tag{3.10}$$

where α_{ij} describes traceless kinematic hardening. Additionally, let us note that the energy-based approach can be also extended into initially anisotropic materials by the concept of spectral decomposition. It is based on the Rychlewski idea introducing anisotropy between three principal stresses and three principal strains by six single stress $\boldsymbol{\sigma}_\alpha$, $\alpha = 1,2,3,4,5,6$ (Rychlewski, 2011), determined by spectral decomposition of the compliance tensor. Ostrowska-Maciejewska *et al.* (2012, 2013) introduced a new proposition of the limit surface for anisotropic materials with asymmetry of limits, by introducing six influence functions $\eta_\alpha(\boldsymbol{\sigma}_\beta) = \eta_\alpha(I_{1(\beta)}, J_{2(\beta)}, J_{3(\beta)})$, $\alpha,\beta = 1,\ldots,6$, into the energy density of elastic strain

$$\eta_1\Phi_1(\boldsymbol{\sigma}_1) + \ldots + \eta_\chi\Phi_\chi(\boldsymbol{\sigma}_\chi) - 1 = 0 \qquad \chi \leqslant 6 \tag{3.11}$$

This concept was applied in Szeptyński's thesis (Szeptyński, 2017) who discussed misstatements appearing in the final form of the failure criterion formulation and by Ostrowska-Maciejewska in her articles (Ostrowska-Maciejewska *et al.*, 2012, 2013) and (Kowalczyk-Gajewska and Ostrowska-Maciejewska, 2005). Recently, the energy-based approach, developed within the Burzyński framework, was extended for symmetric anisotropic materials by Moayyedian and Kadkhodayan (2017, 2021) with different application to materials exhibiting bcc, fcc and hcp crystalline structures. It prolongs the line of reasoning by Hebda and Pęcherski (2005) and Vadilo *et al.* (2011).

3.9. The week material anisotropy – four parametrical hypothesis

Some special, one-parametrical anisotropy, called by Burzyński λ-anisotropy, was developed by Pęcherski *et al.* (2021). The original (Burzyński, 1928, p. 128) solution has the following form

$$\frac{1+\nu'}{3}{\sigma'_{HMH}}^2 + \frac{1}{3}(1-2\nu')(I'_1)^2 + (k_c - k_t)I'_1 - k_c k_t = 0 \tag{3.12}$$

where "prime", similar like von Mises, means week-anisotropy modification

$$\begin{aligned} 3J'_2 &= {\sigma'_{HMH}}^2 = 2(1-\lambda)(\sigma_2-\sigma_3)^2 + 2\lambda(\sigma_3-\sigma_1)^2 + 2(1-\lambda)(\sigma_1-\sigma_2)^2 \\ I'_1 &= 3p' = \frac{\lambda\sigma_1 + (1-\lambda)\sigma_2 + \lambda\sigma_3}{1+\lambda} \end{aligned} \tag{3.13}$$

Here appears the "week anisotropic plasticity coefficient" $\nu' = (1/\varphi^2)(k_c k_t)/(2k_s^2) - 1$, which governs the shape of limit surface (conical, cylindrical, paraboloidal, hyperbolical) also depends on λ as $\varphi^2 = (2/3)(1+\lambda)$. For identification of λ, Pęcherski *et al.* (2021) proposed to make a biaxial tension test.

3.10. The week material anisotropy – many-parameter hypothesis

Similarly, starting from the energy-based arguments, Cazacu and Barlat (2004), keeping the Mises-Reuss line of reasoning, proposed a simple paraboloidal yield surface

$$3J'_2 - \frac{m}{3}(I'_1)^2 - 1 = 0 \tag{3.14}$$

where invariants J'_2, I'_1 are typically calculated but for the transformed Cauchy stresses $\sigma'_{ij} = B_{ijkl}\sigma_{kl}$, where B_{ijkl} are anisotropy constants and m is a constant responsible for pressure sensitiveness. It is Burzyński's proposal, Eq. (3.10), which cannot be applied to the case with symmetry in compression and tension. For a material with transverse isotropy, like limestone, the B_{ijkl} reduces to only 5 independent components, for sn isotropic material Eq. $(3.13)_1$ reduces to the two-parameter yield surface $3J_2 + (k_c - k_t)I_1 - k_c k_t = 0$. This model needs six data to be calibrated: compressive and tensile strengths along the transverse direction $k_{\parallel t}, k_{\parallel c}, k_{\perp t}, k_{\perp c}$ and $k_{s\parallel}$, $k_{s\perp}$ – the shear strength in the symmetry plane and the shear strength in the normal transversal plane. In a general case, the Cazacu-Cristescu surface, Eq. $(3.13)_1$ needs a more sophisticated set of experimental data. But, in the paper by Kowalewski (1998), the practice manner to calibrate $12+6=18$ parameters of yield condition $(3.13)_1$ for 18G2A low-alloy steel was shown. This line of reasoning can be extended for cellular materials (Kordzikowski *et al.* 2005).

4. Further comments

Here we have presented the main substantive dispute that has been and still is going on between Mohr's geometrical approach and Huber's energy approach. In this dispute, the key is the Lode experiment, which showed that for most metals the Huber method is more justified and consistent with the experimental data.

Nevertheless, as the literature shows, the geometrical approach is still preferred by the majority of researchers. The achievements of Mohr's geometrical method are documented. This method can be implemented into numerical calculations with some difficulties. As a result, it is still valid and used.

It is also easy to see that the energetic approach was developed by Huber's disciples, laying the foundation for a scientific school called "the science of effort." That is why so many Polish

researchers can be found in the literature on the subject. In particular, we recall the achievements of Professor Ryszard Pęcherski (IPPT PAN, Warsaw).

But it is Holm Altenbach (Altenbach, 2010) who, preferring the energy-based approach, turns our attention on some, very emotional, scientific discussion in the triangle between William Thomson, W. Rankine, and young J. Clerk Maxwell (see: Maxwell, 1856). It was really Maxwell, who first underlined the role of distortional energy limit in fracture and plasticity. Therefore, taking Maxwell as the original author of the concept of limit distortional energy, Altenbach proposed to call it: "the Maxwell-Huber-Mises" hypothesis and to denote the equivalent stress by σ_{MHM}. In the present paper, according to the great academic tradition and with engineering experience we continue the use of "Huber-Mises-Hencky" hypothesis and σ_{HMH} denotation. This story was discussed in 1924 at the I-st International Congress of Applied Mechanics in Delft, and found a reflection in, perhaps the first, methodical elucidation of the mathematical theory of plasticity given in 1927 by H. Mierzejewski (1928). Yet other historical remarks are found in (Becchi, 1994; Altenbach, 2010; Rychlewski, 2011; Bruhns, 2014).

A historical review of the energy approach should begin with the work of Beltrami (1885). It initiated the energetic approach to the hypotheses of exertion – in particular, its development is marked by the following works:

- Beltrami (1885): $\Phi = \sigma_{ij}\varepsilon_{ji} = \Phi(\sigma_{ij}) = \Phi(\varepsilon_{ij}) \leqslant K$
- Huber (1904): $\Phi = \Phi_\nu + \Phi_f \leqslant K$ or $\Phi_f \leqslant K$
- Mises (1914), Hencky (1924): $\Phi_f \leqslant K$
- Schleicher (1926): $\Phi = \Phi_\nu(\nu) + \Phi_f(\nu) \leqslant Kf(p)$
- Burzyński (1928): $\Phi = \eta_\nu\Phi_\nu + \Phi_f \leqslant K$, $\eta_\nu = \omega + \delta/p$
- Zawadzki (1956): $\Phi = \Phi_\nu + \Phi_f + \Phi_{th} \leqslant K$
- Pęcherski (2011): $\Phi = \eta_\nu\Phi_\nu + \eta_f\Phi_f \leqslant K$, $\eta_f = 1 + \alpha[1 - e^{-\beta(1+\cos(3\theta)}]$

In the literature, the critical energy K is determined by the uniaxial tensile limit value $K = \sqrt{2Ek_t}$, $0.01 < K < 0.30\,\mathrm{MJ/m^3}$=MPa. For example, for $k_t = 700\,\mathrm{MPa}$, $E = 2.1\,\mathrm{GPa}$, we have $K = 0.23\,\mathrm{MJ/m^3}$. This line of thought, which return us to the original Huber concept expressed in Eq. (1.1), was recently re-evaluated by Nalepka and Pęcherski (2003). They proposed to calculate the critical energy K from the first principles, even those based on quantum thermodynamics. Nevertheless, in our opinion, the problem of experimental specification of the value of K is still open, and must be developed in details. Some methods developed by Ładoga and Ogonowski (2005) seem to be helpful.

Acknowledgements

We dedicate this work to the memory of the late Professor Ryszard Pęcherski (1949-2022). Professor Pęcherski devoted many years and a lot of creative effort to renewing and revitalizing the energetic approach to the science of the material effort. In particular, Ryszard Pęcherski spent lots of time on the proper and complete understanding of Tytus Huber's works. Thanks to his job, Huber's work has been made available to the international community. Next, prof. Pęcherski took up the development of the works of Huber's student – Włodzimierz Burzyński. For many years, Professor Pęcherski, as a member of the Presidium of the Polish Society of Theoretical and Applied Mechanics, promoted the achievements of Polish mechanics.

This work has been supported by the Ministry of Science of Poland as the statute subject. The authors thank Professors Artur Ganczarski, Jerzy Podgórski, Michel Feidt, Holm Altenbach for constructive discussion and comments.

References

1. Altenbach H., 2010, Strength hypotheses – a never ending story, *Czasopismo Techniczne*, Politechnika Krakowska, **107**, 20, 5-15

2. Altenbach H., Bolchoun A., Kolupaev V.A., 2014, Phenomenological yield and failure criteria, [In:] *Plasticity of Pressure-Sensitive Materials*, Holm Altenbach, Andreas Öchsner (Eds.), 49-152, Springer-Verlag Berlin Heidelberg
3. Bai Y., Wierzbicki T., 2008, A new model of metal plasticity and fracture with pressure and Lode dependence, *International Journal of Plasticity*, **24**, 1071-1096
4. Bai Y., Wierzbicki T., 2010, Application of extended Mohr-Coulomb criterion to ductile fracture, *International Journal of Fracture*, **161**, 1-20
5. Baltov, A., Sawczuk, A., 1965, A rule of anisotropic hardening, *Acta Mechanica*, **1**, 81-92
6. Banaś, K., Badur J., 2017, Influence of strength differential effect on material effort of a turbine guide vane based on thermo-elasto-plastic analysis, *Journal of Thermal Stresses*, **40**, 11, 1368-1385
7. Barlat F., Brem J.C., Yoon J.W., Chung K., Dick R.E., Lege D.J., Pourboghrat F., Choi S.H., Chu E., 2003, Plane stress yield function for aluminum alloy sheets – part 1: theory, *International Journal of Plasticity*, **19**, 1297-1319
8. Becchi A., 1994, I Criteri di Plasticita: Cento Anni di Dibattito (1864–1964), Doctor Thesis, Firenze
9. Beltrami E., 1885, Sulla conditioni di resistenza dei corpi elastici, Rend. Ist. Lomb., II, 18, 1885, [In]: Beltrami E. (1902-1920) Opere matematiche (4 vols), Hoepli, Milan, 180-189
10. Burzyński W., 1928, Studjum nad Hipotezami Wytężenia (Study on Material Effort Hypotheses), Nakładem Akademji Nauk Technicznych (issued by the Academy of Technical Sciences), Lwów, January 7, 1928, 1-192 (in Polish); reprinted in: Włodzimierz Burzyński *Dzieła Wybrane*, tom I, Polska Akademia Nauk, PWN Warszawa, 1982, 67-258 (in Polish)
11. Burzyński W., 1929a, Teoretyczne podstawy hipotez wytężenia, *Czasopismo Techniczne*, 1929, **47**, 1-41, Lwów (in Polish); reprinted in: Włodzimierz Burzyński *Dzieła Wybrane*, tom I, Polska Akademia Nauk, PWN Warszawa, 1982, 264-303 (in Polish), English translation: Theoretical Foundations of the Hypotheses of Material Effort, *Engineering Transactions*, **56**, 3, 269-305, 2008
12. Burzyński W., 1929b, Ueber die Anstrengungshypothesen, *Schweizerische Bauzeitung*, **94**, 21, 23, November 1929, 259-262; reprinted in Włodzimierz Burzyński *Dzieła Wybrane*, tom I, Polska Akademia Nauk, PWN Warszawa, 1982, 259-262
13. Bruhns O.T., 2014, Some remarks on the history of plasticity – Heinrich Hencky, a pioneer of the early years, [In:] *The History of Theoretical, Material and Computational Mechanics – Mathematics Meets Mechanics and Engineering*, Ed. Erwin Stein, Springer-Verlag Berlin Heidelberg, 133-152
14. Casey J., Sullivan T.D., 1985, Pressure dependency, strength-differential effects and plastic volume expansion in metals, *International Journal of Plasticity*, **1**, 1, 39-61
15. Cazacu O., Barlat F., 2004, A criterion for description of anisotropy and yield differential effects in pressure-insensitive metals, *International Journal of Plasticity*, **20**, 2027-2045
16. Drucker D.C., 1949, Relations of experiments to mathematical theories of plasticity, *Journal of Applied Mechanics*, **16**, 349-357
17. Drucker D.C., 1973, Plasticity theory, strength-differential (SD) phenomenon, and volume expansion in metals and plastics, *Metallurgical and Materials Transactions*, **4**, 667-673
18. Drucker D.C., Prager W., 1948, Soil mechanics and plastic analysis or limit design, *Quarterly of Applied Mathematics*, 157-165
19. Dubey P., Kopeć M., Łazińska M., Kowalewski Z.L., 2023, Yield surface identification of CP-Ti and its evolution reflecting pre-deformation under complex loading, *International Journal of Plasticity*, **167**, August, 103677
20. Dudda W., 2020, Mechanical characteristics of 26H2MF and St12T steels under compression and elevated temperatures, *Strength of Materials*, **52**, 2, 325-328
21. Dudda W., 2021, *Issues Concerning the Thermal Effort Concept in Heat Resistive Steels*, Press UWM, Olsztyn, 1-300
22. Egner W., Sulich P., Mroziński, S., Egner H., 2020, Modelling thermo-mechanical cyclic behavior of P91 steel, *International Journal of Plasticity*, **135**, 102820
23. Föppl, A., 1907, *Vorlesunfen über technische Mechanik*, V. Bd. Leipzig

24. Frąś T., Kowalewski Z., Pęcherski R.B., Rusinek A., 2010, Applications of Burzyński failure criteria – I. Isotropic materials with asymmetry of elastic range, *Engineering Transactions*, **58**, 1-10
25. Frąś T., Nishda M., Rusinek A., Pęcherski R.B., Fukuda N., 2014, Description of the yield state of bioplastics on examples of starch-based plastics and PLA/PBAT blends, *Engineering Transactions*, **62**, 329-354
26. Frąś T., Pęcherski R.B., 2010, Applications of the Burzyński hypothesis of material effort for isotropic solids, *Mechanics and Control*, **29**, 2, 45-50
27. Gao X. Zhang T., Hayden M., Roe C., 2009, Effects of the stress state on plasticity and ductile failure of an aluminum 5083 alloy, *International Journal of Plasticity*, **25**, 2366-2382
28. Gao X., Zhang T., Zhou J., Graham S., Hayden M., Roe C., 2011, On stress-state dependent plasticity modeling: Significance of the hydrostatic stress, the third invariant of stress deviator and the non-associated flow rule, *International Journal of Plasticity*, **27**, 217-231.
29. Geiringer H., Prager W., 1934, Mechanik isotroper Körper im plastischen Zustand, *Ergebnisse der exakten Naturwissenschaften*, **13**, 314-363
30. Gudehus, G., 1973, Elastoplastische Stoffgleichungen fur Trockenen Sand, *Ingenieur Archiv*, **42**, 3, 151-169
31. Hebda M., Pechęrski R.B., 2005, Energy-based criterion of elastic limit state in fibre-reinforced composites, *Archives of Metallurgy and Materials*, **50**, 1073-1088
32. Helmhotz H., 1902, *Dynamic continuerlich verbreiteten Massen*, Leipzig
33. Hencky H., 1924, Zur Theorie plastischer Deformationen und der hierdurch im Material hervorgerufenen Nachspannungen, *ZAMM*, **4**, 323-334
34. Hill R., 1948, A theory of the yielding and plastic flow of anisotropic metals, *Proceedings of the Royal Society, London*, **A193**, 281-297
35. Hosford W.F., Allen T.J., 1973, Twinning and directional slip as a cause for a strength differential effect, *Metallurgial Trensactions*, **4**, 1424-1425
36. Hu Q., Li X., Han X., Li H., Chen J., 2017, A normalized stress invariant-based yield criterion: modeling and validation, *International Journal of Plasticity*, **99**, 248-273
37. Hu Q., Yoon J.W., 2021, Analytical description of an asymmetric yield function (Yoon, 2014) by considering anisotropic hardening under non-associated flow rule, *International Journal of Plasticity*, **140**, 102978
38. Huber M.T., 1904, Właściwa praca odkształcenia jako miara wytężenia materjału, *Czasopismo Techniczne*, Lwów, (in Polish), English translation: Specific work of strain as a measure of material effort, *Archives of Mechanics*, **56**, 173-190, 2004
39. Huber M.T., 1948, *Kryteria wytrzymałościowe w stereomechanice technicznej*, Warszawa, PIW
40. Kłębowski Z., 1958, Przyrost właściwej energii swobodnej jako miara wytężenia, *Zeszyty Naukowe Politechniki Warszawskiej – Mechanika*, **37**, 79-85
41. Kolupaev V.A., Yu M.H., Altenbach H., 2016, Fitting of the strength hypotheses, *Acta Mechanica*, **227**, 1533-1556
42. Kordzikowski P., Janus-Michalska M., Pęcherski R.B., 2005, Specification of energy-based criterion of elastic limit states for cellular materials, *Archives of Metallurgy and Materials*, **50**, 621-634
43. Kordzikowski P., Pęcherski R.B., 2010, Assessment of the material strength of anisotropic materials with asymmetry of the elastic range, *Mechanics and Control*, **29**, 2, 57-62
44. Kowalczyk K., Ostrowska-Maciejewska J., Pęcherski R.B., 2003, An energy-based yield criterion for solids of cubic elasticity and orthotropic limit state, *Archives of Mechanics*, **55**, 5-6, 431-448
45. Kowalczyk-Gajewska K., Ostrowska-Maciejewska J., 2005, Energy-based limit criteria for anisotropic elastic materials with constraints, *Archives of Mechanics*, **57**, 133-155
46. Kowalczyk-Gajewska K., Pieczyska E.A., Golasiński K., Maj M., Kuramoto S., Furuta T., 2019, A finite strain elastic-viscoplastic model of Gum Metal, *International Journal of Plasticity*, **119**, 85-101

47. Kowalewski Z.L., 1998, Assessment of cyclic properties of 18G2A low-alloy steel at biaxial stress state, *Acta Mechanica*, **120**, 71-89
48. Krzyś W., Życzkowski M., 1962, *Sprężystości i Plastyczność. Wybór zadań i przykładów*, PWN, Warszawa
49. Kuroda M., Kuwabara T., 2002, Shear band development in polycrystalline metal with strength-differential effect and plastic volume expansion, *Proceedings of The Royal Society A, London*, **458**, 2243-2262
50. Lade P.V., Duncan J.M., 1973, Cubical triaxial tests on cohesionless soil, *Journal of the Soil Mechanics and Foundations Division, ASCE*, **99**, SM10, 793-812
51. Lehmann Th., 1964, Anisotrope plastische Formänderungen, *Rheologica Acta*, **3**, 281-285
52. Lode W., 1926, Versuche über den Einfluss der mittleren Hauptspannung auf das Fliessen der Metalle: Eisen, Kupfer und Nickel, *Zeitschrift für Physik*, **36**, 913-943
53. Lou Y.S., Yoon J.W., 2017, Anisotropic ductile fracture criterion based on linear transformation, *International Journal of Plasticity*, **93**, 3-25
54. Lou Y.S., Yoon J.W., Huh H., 2014, Modeling of shear ductile fracture considering a changeable cut-off value for the stress triaxiality, *International Journal of Plasticity*, **54**, 56-80
55. Lou Y., Zhang S., Yoon J.W., 2020, Strength modeling of sheet metals from shear to plane strain tension, *International Journal of Plasticity*, **134**, 102813
56. Łagoda T., Ogonowski P., 2005, Criteria of multiaxial random fatigue based on stress, strain and energy parameters of damage in the critical plane, *Materialwiss Werkstofftech*, **36**, 429-437
57. Maxwell J.C., 1856, A private letter to prospect Lord Kelvin, *Proceedings of the Cambridge Philosophical Society*, **32**, 1936 (cf. also: Origins of Clerk Maxwell's electric ideas as described in familiar letters to William Thompson, ed. by Sir J. Larmor, Cambridge at Univerdity Press, 1937)
58. Mierzejewski H., 1927, *Foundations of Mechanics of Plastic Solids* (in Polish), Warszawa
59. Mirone G., Corallo D., 2010, A local viewpoint for evaluating the influence of stress triaxiality and Lode angle on ductile damage and hardening, *International Journal of Plasticity*, **26**, 348-371
60. Mises R., 1913, Mechanik der festen Körper im plastisch-deformablen Zustand, Göttingen Nachrichten, *Mathematisch-Physikalische Klasse*, **4**, 1, 582-592
61. Mises R., 1928, Mechanik der plastischen Formänderung von Kristallen, *ZAMM*, **8**, 161-185
62. Moayyedian F., Kadkhodayan M., 2017, A modified Burzyński criterion for anisotropic pressure-dependent materials, *Sadhana – Academy Proceedings in Engineering Sciences*, **42**, 1, 95-109
63. Moayyedian F., Kadkhodayan M., 2021, Modified Burzyński criterion along with AFR and non-AFR for asymmetric anisotropic materials, *Archives of Civil and Mechanical Engineering*, **21**, 64, 1-19
64. Möhr O., 1882, *Über die Darstellung des Spannungszustandes und des Deformationsztanden eines Körperlementen*, Zivilingenieur (rep. O. Möhr, *Technische Mechanik*, Berlin, 1906)
65. Möhr O., 1900, Welche Umstände bedingen die Elastizitätsgrenze u. den Bruch eines Materials, *Zeitschrift des Vereines Deutscher Ingenieure*
66. Möhr O., 1906, *Abhandlungen aus dem Gebiete der Technischen Mechanik*, Verlag vön Wilhelm Ernst & Söhn, Berlin, 1-447
67. Mróz Z., 1967, On the description of anisotropic workhardening, *Journal of the Mechanics and Physics of Solids*, **15**, 163-175
68. Mróz Z., Seweryn A., 1998, Non-local failure and damage evolution rule: Application to a dilatant crack model, *Journal de Physique IV*, France, **8**, 257-268
69. Mucha M., Wcisło B., Pamin J., Kowalczyk-Gajewska K., 2018, Instabilities in membrane tension: Parametric study for large strain thermoplasticity, *Archives of Civil and Mechanical Engineering*, **18**, 1055-1067
70. Nádai A., 1927, *Der bildsame Zustand der Werkstoffe*, Berlin
71. Nalepka K., Pęcherski R.B., 2002, Fizyczne podstawy energetycznego kryterium wytężenia monokryształów, *XXX Szkoła Inżynierii Materiałowej*, red. J. Pacyno, Kraków-Ustroń-Jaszowiec, 1-4 X 2002, AGH, Kraków 2002, 311-316

72. Nalepka K., Pęcherski R.B., 2003, Energetyczne kryteria wytężenia. Propozycja obliczania granicznych energii z pierwszych zasad, *Rudy i Metale Nieżelazne*, **48**, 533-536
73. Nixon M.E., Cazacu O., Lebensohn R.A., 2010, Anisotropic response of high-purity-titanium: experimental characterization and constitutive modelling, *International Journal of Plasticity*, **26**, 516-532
74. Nowak M., Ostrowska-Maciejewska J, Pęcherski R.B., Szeptyński P., 2011, Yield criterion accounting for the third invariant of stress tensor deviator. Part I. Proposition of the yield criterion based on the concept of influence functions, *Engineering Transactions*, **59**, 4, 273-281
75. Nowak Z., Nowak M., Pęcherski R.B., 2014, A plane stress elastic-plastic analysis of sheet metal cup deep drawing processes, SSTA 2013, [In:] *10th Jubilee Conference on Shell Structures – Theory and Applications*, W. Pietraszkiewicz and J. Górski (Eds.), **3**, 129-132
76. Novozhilov V.V., 1952, On a physical meaning of the stress invariants, *Applied Mathematics and Mechanics (PMM)*, **5**, 16-30
77. Oliferuk W., Maj M., Raniecki B., 2004, Experimental analysis of energy storage rate components during tensile deformation of polycrystals, *Materials Science and Engineering: A*, **374**, 1-2, 77-81
78. Orłowski K.A., Ochrymiuk T., Atkins A., Chuchała D., 2013, Application of fracture mechanics for energetic effects predictions while wood sawing, *Wood Science and Technology*, **47**, 949-963
79. Orłowski K.A., Ochrymiuk T., Hlaskova L., Chuchała D., Kopecky Z., 2022, Revisiting the estimation of cutting power with different energetic methods while sawing soft and hard woods on the circular sawing machine: a Central Europe case, *Wood Science and Technology*, **54**, 457-477
80. Ostrowska-Maciejewska J., Pęcherski R.B., Szeptyński P., 2012, Limit condition for anisotropic materials with asymmetric elastic range, *Engineering Transactions*, **60**, 125-138
81. Ostrowska-Maciejewska J., Szeptyński P., Pęcherski R.B., 2013, Mathematical foundations of limit criterion for anisotropic materials, *Archives of Metallurgy and Materials*, **58**, 4, 1223-1235
82. Pełczyński T., 1959, O hipotezie wytężeniowej O. Mohra, *Przegląd Spawalnictwa*, **3**, 11, 74-78
83. Pęcherski R.B., 1998, Macroscopic effects of micro-shear banding in plasticity of metals, *Acta Mechanica*, **131**, 203-224
84. Pęcherski R.B., 2008, Burzyński yield condition vis-a-vis the related studies reported in the literature, *Engineering Transactions*, **56**, 4, 383-391
85. Pęcherski R.B., Nalepka K., Frąś T., Nowak M., 2014, Inelastic flow and failure of metallic solids. Material effort: study across scales, [In:] *Constitutive Relations under Impact Loadings*, T. Łodygowski, A. Rusinek (Eds.), CISM International Centre for Mechanical Sciences 552, Springer, Vienna
86. Pęcherski R.B., Rusinek A., Frąś T., Nowak M., Nowak Z., 2021, Energy-based yield condition for orthotropic materials exhibiting asymmetry of elastic range, *Archives of Metallurgy and Materials*, **65**, 2, 771-778
87. Pęcherski R.B., Szeptyński P., Nowak M., 2011, An extension of Burzyński hypothesis of material effort accounting for the third invariant of stress tensor, *Archives of Metallurgy and Materials*, **56**, 503-508
88. Pietruszczak S., Inglis D., Pande G.N., 1999, A fabric-dependent fracture criterion for bone, *Journal of Biomechanics*, **32**, 1071-1079
89. Pietruszczak S., Mróz Z., 2001, On failure criteria for anisotropic cohesive-frictional materials, *International Journal for Numerical and Analytical Methods in Geomechanics*, **25**, 509-524
90. Podgórski J., 1984, Limit state condition and the dissipation function for isotropic materials, *Archives of Mechanics*, **36**, 323-342
91. Podgórski J., 1985, General failure criterion for isotropic media, *Journal of Engineering Mechanics*, **III**, 2, 188-201
92. Prager W., Hodge P.G., 1954, *Theorie ideal plastischer Körper*, Wien, Springer-Verlag
93. Reuss A., 1929, Berechnung der Flissgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle, *ZAMM*, **9**, 49-58

94. Reuss A., 1930, Berücksichtigung der elastischen Formänderung in der Plastizitätstheorie, *ZAMM*, **10**, 266-274
95. Rychlewski J., 2011, Elastic energy decomposition and limit criteria, *Engineering Transactions*, **59**, 1, 31-63
96. Schleicher F., 1926, Der Spannungszustand an der Fließgrenze (Plastizitätsbedingung), *ZAMM*, **6**, 199-216
97. Schleicher F., 1928, Über die Sicherheit gegen Uberschreiten der Fliessgrenze bei statischer Beanspruchung, *Der Bauingenieur*, **9**, 15, 253-261
98. Skrzypek J., Ganczarski A.W., 2015, *Mechanics of Anisotropic Materials*, Springer International Publishing Switzerland, Cham
99. Skrzypek J., Ganczarski A., 2016, Constraints on the applicability range of pressure-sensitive yield/failure criteria: strong orthotropy or transverse isotropy, *Acta Mechanica*, **227**, 2275-2304
100. Szeptyński P., 2017, Energy-based yield criteria for orthotropic materials, exhibiting strength-differential effect. Specification for sheets under plane stress state, *Archives of Metallurgy and Materials*, **62**, 729-736
101. Timoshenko S., 1953, *History of Strength of Materials*, Mc Graw-Hill
102. Torre C., 1947, Ein flussdermittleren Hauptnormalspannung auf die Fließ- und Bruchgrenze, *Österreichisches Ingenieur-Archiv*, **I**, 4/5, 316-342
103. Vadillo G., Fernández-Sáez J., Pęcherski R.B., 2011, Some applications of Burzyński yield condition in metal plasticity, *Materials and Design*, **32**, 2, 628-635
104. Wierzbicki T., Bao Y., Lee Y.-W., Bai Y., 2005, Calibration and evaluation of seven fracture models, *International Journal of Mechanical Sciences*, **47**, 719-743
105. Willam K.J., Warnke E.P., 1974, Constitutive model for the triaxial behaviour of concrete, *Proceedings of the May 17-19, 1974, International Association of Bridge and Structural Engineers Seminar on Concrete Structures Subjected to Triaxial Stresses*, Bergamo, Italy
106. Yoon J.W., Lou Y.S., Yoon J.H., Glazoff M.V., 2014, Asymmetric yield function based on the stress invariants for pressure sensitive metals, *International Journal of Plasticity*, **56**, 184-202
107. Yu M.-H., 2004, *Unified Strength Theory and its Applications*, Springer-Verlag, Berlin, Heidelberg
108. Zawadzki J., 1954, Ciśnienie zredukowane jako jeden z parametrów wytężenia. Przyrost właściwej energii swobodnej jako miara wytężenia, *Rozprawy Inżynierskie*, **LXXIII**, 357-398
109. Życzkowski M., 1981, *Combined Loadings in the Theory of Plasticity*, PWN-Polish Scientific Publishers, Warszawa
110. Życzkowski M., 1999, Discontinuous bifurcations in the case of the Burzyński-Torre yield condition, *Acta Mechanica*, **132**, 19-35

Manuscript received November 23, 2023; accepted for publication October 8, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 827-839, Warsaw 2024
https://doi.org/10.15632/jtam-pl/193952

PREDICTION AND MITIGATION OF INSTABILITY IN ULTRA LONG DRILLING SHAFT LINING STRUCTURES USING THE CUSP CATASTROPHE MODEL

Jimin Liu, Ruixue Pan, Hua Cheng, Haixu Fan

School of Civil Engineering and Architecture, Anhui University of Science and Technology, Huainan, China; and State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mine, Huainan, China

corresponding author J.M. Liu, e-mail: jimliu@aust.edu.cn

This paper deals with the stability prediction of an ultralong drilling shaft lining structure and how to mitigate its structural instability. Based on catastrophic characteristics of the instability process, the catastrophe method and a cusp catastrophic model is applied in analyzing the instability optimization measure. The process and mechanism of catastrophe instability is analyzed, and its corresponding instability criterion is founded. A case study and numerical results show that this optimization measure can increase its critical depth by 45% and mitigate structural vertical instability, which provides a theoretical possibility for the stability control technology of ultra long shaft lining structures.

Keywords: cusp catastrophic model, critical depth, instability criterion, drilling shaft lining structure

1. Introduction

A shaft drilling method is a special construction method in coal and copper mine constructions all over the world, which is also used in subways and municipal engineering (Liu and Meng, 2015; Dorn and Kaledin, 2013; Hara *et al.*, 2019). In drilling shaft lining structures (DSLS) during its construction, buckling usually becomes a dominated failure pattern rather than damage due to material strength, which is often related to deformation of the whole DSLS experiencing a sudden and irreversible change when it reaches or exceeds the critical value. Any slight eccentricity or lateral force can cause shaft lining to suddenly tilt and slip, resulting in vertical instability of DSLS. Moreover, the longer the construction length of DSLS, the greater the possibility of vertical instability of the structure. Stability characteristics of DSLS have become a problem in shaft construction and seriously restricted its application in the deep coal mine shaft engineering. Therefore, it is necessary to take measures to predict and suppress the vertical instability of DSLS in ultra long coal mines and carry out corresponding theoretical research.

At present, many scholars have conducted numerous researches on stability characteristics of shaft lining structures. Most of them regarded DSLS as a slender compression rod hinged at both ends when it is sunk to the bottom of the well and before cementing and filling, and analyzed stability of the structure based on the energy method (Wegner and Kurpisz, 2017; Xing *et al.*, 2023). Hong (1980) initially mentioned that there exists a structural instability with DSLS and suggested adding counterweight water to control the instability. Niu *et al.* (2006) believed that the counterweight water should be regarded as lateral pressure and established a critical depth calculation method in constant cross-section DSLS. Cheng *et al.* (2008) used the same mechanical background and proposed the critical depth of variable cross-section DSLS. They all raised their instability criterion and critical depth based on the principle of minimum potential energy. Their research results have been well verified and applied in shaft engineering

at a depth of 600m. But the studies on predicting the vertical instability of ultra long DSLS with a construction depth exceeding 600m and the mitigating instability optimization measure is scare.

Furthermore, there are many studies concerning catastrophe theory and its application. Catastrophe theory is a fundamental method of scientific research, which uses mathematical models to explain various forms and structures of discontinuous catastrophe phenomena. It has many successful application examples in many disciplines such as stability of elastic structures, economy and natural science (Merli and Pavese, 2018; Karman and Pawlowski, 2019; Niu *et al.*, 2023). Especially in the field of civil engineering, catastrophe theory has helped to solve many problems such as the critical buckling load of pile foundation, soil slope and etc. (Lei *et al.*, 2022; Xu and Ni, 2019). Liu *et al.* (2020, 2022) established a catastrophe model for the instability of DSLS and analyzed its instability mechanics under THE traditional construction method. It can be found that the studies on catastrophe theory and its catastrophic models can help one to solve structural instability problems and have ideal application effects.

Motivated by this, this paper proposes a cusp catastrophic model based on catastrophe theory to reveal the characteristic of an ultra long shaft structure instability phenomenon when using instability optimization measures. Based on abrupt and irreversible characteristics during the instability process, the catastrophe method and a cusp catastrophic model was applied. Firstly, an instability optimization measure is proposed and a cusp catastrophe model is established. Based on the catastrophe method and catastrophic model, the instability characteristics of ultra long DSLS is analyzed. In the end, the critical depth affecting factors are discussed, and an actual engineering verifies the effectiveness of the instability optimization measure.

2. Theoretical model

2.1. Instability optimization measures in ultra long drilling shaft lining

In general, the traditional construction of the shaft drilling method concludes three steps: suspending and sinking the shaft, suspended and sunks to the bottom of the well but not yet filled and cemented, and filled cemented the shaft. Figure 1a shows a schematic representation and a scene diagram of the shaft construction during suspension and subsidence. When the last prefabricated shaft segments finish its connection, the whole DSLS lands on the bedrock directly. The moment when the shaft floats and sinks to the bottom and touches the bottom of the well on the bedrock but not yet filled and cemented, the horizontal and vertical displacement at the bottom of DSLS is limited and forms a bi-directional hinged support. At the same time the shaft is controlled by guided wood for lateral displacement at the wellhead position. The two ends of the entire DSLS are hinged and constrained. And geometric dimensions in the vertical direction are significantly higher than those in the horizontal direction, thus forming a slender rod with both ends hinged. At this moment, any slight eccentricity or lateral force during construction may cause structural sloping or sliding, resulting in overall instability of DSLS.

To improve the overall structural stability of DSLS in the ultra long coal mine shaft engineering, we propose the instability optimization measures of pre pouring cement slurry before the shaft touches the bottom. Before the shaft is suspended and sinks to the bottom of the well, cement mortar is injected into the well through a grouting pipe in advance. The speed of suspension and sinking of the shaft is controlled by controlling the injection amount and speed of counterweight water, so that the shaft is slowly embedded into the bottom of the well. 4-6 grouting pipes are evenly arranged around the well, with diameters ranging from 50 mm to 250 mm and a length equal to depth of the shaft. The grouting pipe gradually increases as the cement mortar level rises and is removed from the top. Figure 1b shows a schematic representation and a scene diagram of the instability optimization measure. Through applying this optimization

measure, the bottom of the shaft is fixed. The constraint manner at the bottom is intervened and optimized from hinged to fixed.

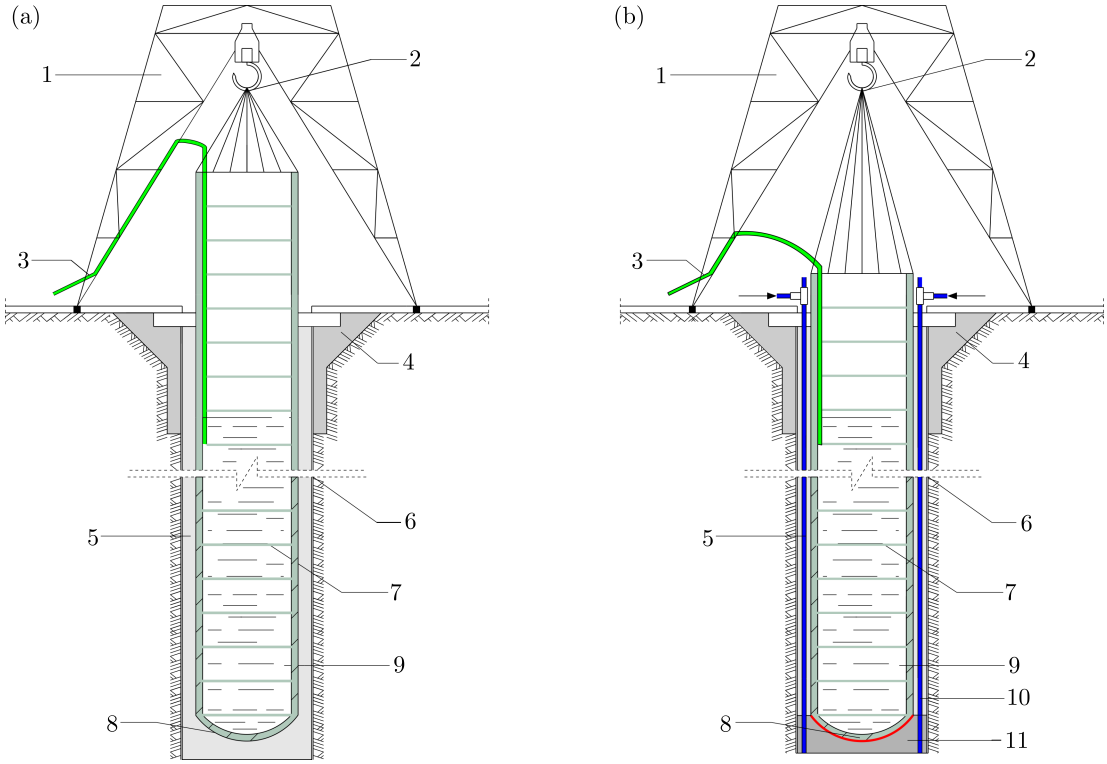


Fig. 1. Schematic representation and the scene diagram of the shaft drilling method: (a) traditional construction measure, (b) instability optimization measure; 1 – gantry crane, 2 – drilling platform, 3 – main water pipe, 4 – guided wood at wellhead, 5 – mud, 6 – well holes, 7 – shaft segment, 8 – shaft lining bottom, 9 – counterweight water, 10 – pumping pipe of cement slurry, 11 – cement slurry

2.2. Basic assumption and mechanical background

For convenience of the discussion, we assume that the inner diameter of the shaft remains unchanged and the shaft material is described by a single specification. The counterweight water level has not reached the wellhead position. Furthermore, we make the following basic assumptions:

Assumption 1. The shaft material is linearly elastic and follows Hooke's law.

Assumption 2. DSLS can be regarded as a slender rod with the bottom fixed and top hinged, and its deflection curve equation satisfies $y = \delta[\cos(3\pi x/2H) - \cos(\pi x/2H)]$, where δ is the maximum displacement of deflection curve, H is height of the shaft structure (Yang, 2019).

Assumption 3. DSLS obeys the theory of small deflection of structural stability (Ozbasaran, 2018). The rod has no initial defects or stresses. During buckling, the rod only undergoes planar bending deformation, and the bending deformation is small.

The force acting on DSLS includes the self-weight of shaft P_c, the lateral pressure of mud on the outer surface of shaft P_m, the lateral pressure of counterweight water on the inner surface of shaft P_w, the reverse force on the top of shaft R_B, and the reverse force on the bottom of the shaft R_A, which is shown in Fig. 2a.

The lateral pressure caused by the mud unit length on the outer surface of shaft p_m can be expressed as $p_m = (1/4)\pi D^2 \gamma_m \sin\alpha$. D is outer diameter of the shaft, γ_m is weight of mud, α is inclination of the deflection curve. The force direction is along the normal direction of the deflection curve. The lateral pressure caused by unit length counterweight water on the inner surface of shaft p_w can be expressed as $p_w = (\pi/4)d^2\gamma_w \sin\alpha$. d is inner diameter of the shaft, γ_w is weight of water. Its force direction is also along the normal direction of the deflection curve. The self-weight of the reinforced concrete shaft per unit length is $p_c = (\pi/4)(D^2 - d^2)[\gamma_c(1-\rho)+\gamma_s\rho]$. γ_c and γ_s are weights of concrete and steel, ρ is the reinforcement ratio of the reinforced concrete shaft. The direction is vertical downwards. Due to the absence of displacement at supports A and B, the external force potential energy generated by R_A and R_B is 0. We will not discuss the reverse force on the bottom of the shaft R_A and the reverse force on the top of the shaft R_B. Based on Assumption 2 and the mechanics of background, the deformation deflection curve and stability analysis calculation diagram of DSLS after using the instability optimization measure can be obtained, as shown in Fig. 2b.

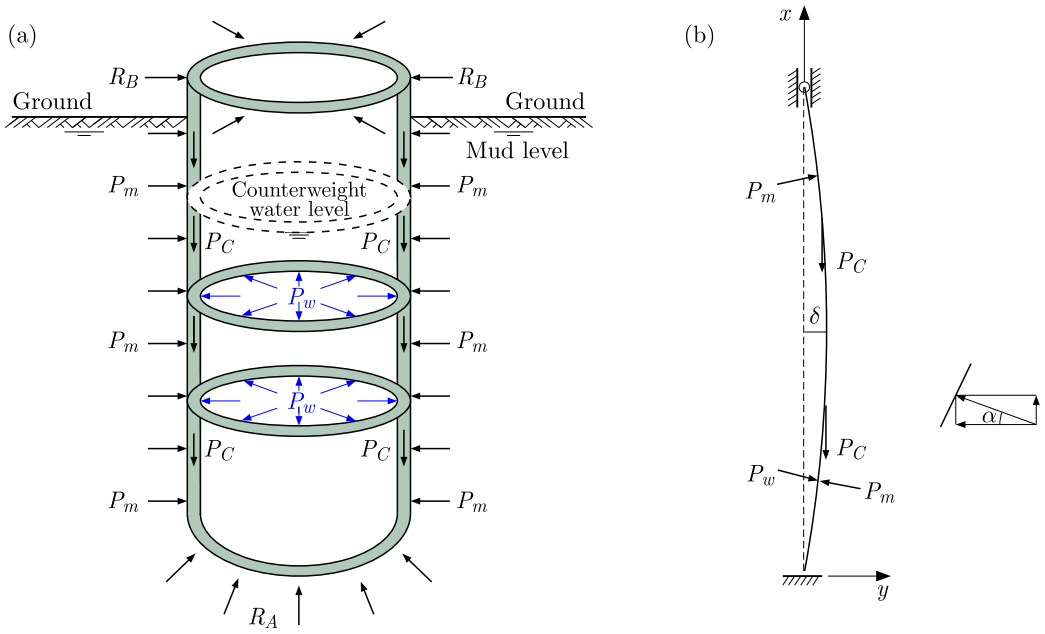


Fig. 2. Force and stability analysis of the shaft by using the instability optimization measure: (a) force analysis of shaft, (b) schematic diagram of stability analysis

2.3. Total potential energy equation

The total potential energy Π of DSLS is the sum of strain energy U released by the vertical bending deformation of the structure and external force potential energy V, expressed as

$$\Pi = U + V \tag{2.1}$$

According to Assumption 3, DSLS obeys the theory of small deflection of structural stability. In curvature equation $1/\rho = y''/\sqrt{(1+y'^2)^3}$, compared to 1, y'^2 can be ignored, the curvature can be approximated by y''. The in-plane bending moment can expressed as $M = 1/\rho = y''$. Based on the deflection curve equation in Assumption 2, the total strain energy of DSLS is

$$U = \int_0^H \frac{M^2}{2EI}\,dx = \frac{1}{2}\int_0^H EIy''^2\,dx = \frac{41EI\pi^4}{32H^3}\delta^2 \tag{2.2}$$

The external force potential energy V acting on the system is the total of external force potential energy V_y caused by the vertical force P and external force potential energy V_x caused by the horizontal force. V_y can be determined by $V_y = -p\lambda$, where λ is small deformation generated by the system under vertical force p. According to Assumption 3, the small deformation satisfies $\lambda = d_s - d_x = y'^2 d_x/4$. On the other hand, the external force potential energy V_x caused by the vertical force includes external force potential energy caused by horizontal component of mud lateral pressure $(V_m)_x$ and external force potential energy caused by the horizontal component of counterweight water pressure $(V_w)_x$. So the external force potential energy can be expressed as

$$\begin{aligned} V = V_y + V_x &= -P_c\lambda_c - (P_m)_y\lambda_m(P_w)_y\lambda_w + (V_w)_x + (V_m)_x = \Big(\frac{1-5}{16}\pi^2\Big)p_c\delta^2 \\ &+ \frac{\pi^3 D^2\gamma_m}{4}\Big(\frac{183\pi^2}{256H^2} + \frac{308}{75}\Big)\delta^4 - \frac{\pi d^2\gamma_w}{2}\sin^2\frac{\pi H_w}{2H}\sin^2\frac{\pi H_w}{H}\delta^2 + Q_m\delta \\ &- \frac{\pi^5 d^2\gamma_w\delta^4}{128H^2}\Bigg\{ - \frac{1}{\pi^2}\Big[\frac{17657}{1200}\cos\frac{\pi H_w}{H} - \frac{9913}{600}\cos\frac{2\pi H_w}{H} \\ &+ 9\Big(-\frac{1267}{120} + \frac{549}{1200}\cos\frac{3\pi H_w}{H} + \frac{1}{100}\cos\frac{4\pi H_w}{H} + \frac{1}{8}\cos\frac{5\pi H_w}{H}\Big)\Big]\sin^2\frac{\pi H_w}{2H} \\ &+ \frac{15}{\pi}\Big(- \frac{59}{15}\sin\frac{\pi H_w}{H} + \frac{83}{48}\sin\frac{2\pi H_w}{H} - \frac{37}{30}\sin\frac{3\pi H_w}{H} + \frac{27}{80}\sin\frac{4\pi H_w}{H} \\ &- \frac{9}{50}\sin\frac{5\pi H_w}{H} + \frac{9}{80}\sin\frac{6\pi H_w}{H}\Big)\frac{H_w}{H} + \frac{183}{8}\Big(\frac{H_w}{H}\Big)^2\Bigg\} \end{aligned} \tag{2.3}$$

Take

$$\begin{aligned} A_1 = -\frac{1}{\pi^2}\Big[&\frac{17657}{1200}\cos\frac{\pi H_w}{H} - \frac{9913}{600}\cos\frac{2\pi H_w}{H} + 9\Big(-\frac{1267}{120} + \frac{549}{1200}\cos\frac{3\pi H_w}{H} \\ &+ \frac{1}{100}\cos\frac{4\pi H_w}{H} + \frac{1}{8}\cos\frac{5\pi H_w}{H}\Big)\Big]\sin^2\frac{\pi H_w}{2H} + \frac{15}{\pi}\Big(-\frac{59}{15}\sin\frac{\pi H_w}{H} + \frac{83}{48}\sin\frac{2\pi H_w}{H} \\ &- \frac{37}{30}\sin\frac{3\pi H_w}{H} + \frac{27}{80}\sin\frac{4\pi H_w}{H} - \frac{9}{50}\sin\frac{5\pi H_w}{H} + \frac{9}{80}\sin\frac{6\pi H_w}{H}\Big)\frac{H_w}{H} + \frac{183}{8}\Big(\frac{H_w}{H}\Big)^2 \end{aligned} \tag{2.4}$$

By substituting formulas (2.2)-(2.4) into (2.1), we obtain the total potential energy function

$$\begin{aligned} \Pi = &\Big[-\frac{A_1\pi^5 d^2\gamma_w}{128H^2} + \frac{\pi^3 D^2\gamma_m}{4}\Big(\frac{183\pi^2}{256H^2} + \frac{308}{75}\Big)\Big]\delta^4 \\ &+ \Big(\frac{41EI\pi^4}{32H^3} + \frac{1-5}{16}\pi^2\Big)p_c - \frac{\pi d^2\gamma_w}{2}\sin^2\frac{\pi H_w}{2H}\sin^2\frac{\pi H_w}{H}\Big)\delta^2 + Q_m\delta \end{aligned} \tag{2.5}$$

2.4. A cusp catastrophic model for DSLS

Do equivalent transformation on (2.5), if

$$\begin{aligned} A_2 &= -\frac{A_1\pi^5 d^2\gamma_w}{128H^2} + \frac{\pi^3 D^2\gamma_m}{4}\Big(\frac{183\pi^2}{256H^2} + \frac{08}{75}\Big) \\ A_3 &= \frac{41EI\pi^4}{32H^3} + \Big(\frac{1-5}{16\pi^2}\Big)p_c - \frac{\pi d^2\gamma_w}{2}\sin^2\frac{\pi H_w}{2H}\sin^2\frac{\pi H_w}{H} \\ A_4 &= Q_m = \int_0^H p_m\cos\alpha\, d_x = 0 \end{aligned} \tag{2.6}$$

Meanwhile, the dimensionless parameters x, m and n are introduced as $x = 4\sqrt{4A_2}\delta$, $m = A_3/\sqrt{A_2}$, $n = A_4/\sqrt[4]{4A_2}$. The total potential energy function of DSLS can be expressed as

$$\Pi(x) = \frac{1}{4}x^4 + \frac{1}{2}mx^2 + nx \tag{2.7}$$

This equation has two control variables m and n, and one state variable x. According to Thom's classification theorems (Pacoste, 1993), DSLS conforms to the cusp catastrophe mathematical model (Li *et al.*, 2016).

Let the first derivative of the total potential function be zero

$$\Pi'(x) = x^3 + mx + n = 0 \tag{2.8}$$

The pints that satisfy equation (2.8) are critical points. But a critical point may not necessarily be a stable point, it may make the system stable or unstable. Only the point where the potential function takes a unique extremum is a stable point. So in order to make the structure stable, it is necessary to simultaneously satisfy the equation

$$\Pi''(x) = 3x^2 + m = 0 \tag{2.9}$$

Combine equations (2.8) and (2.9) to obtain the bifurcation set equation

$$\Delta = 4m^3 + 27n^2 = 0 \tag{2.10}$$

The points that satisfy equation (2.10) form the singularity set S. The projection of singularity set S in control variable plane C is called the bifurcation points set B. Figure 3 shows the singularity set S, bifurcation point set B and equilibrium surface M of the cusp catastrophe model (Saunders, 1980). The equilibrium surface M is made of all critical points that fulfill equation (2.8). M is a well-known surface made up of the control variables m and n and the state variables x.

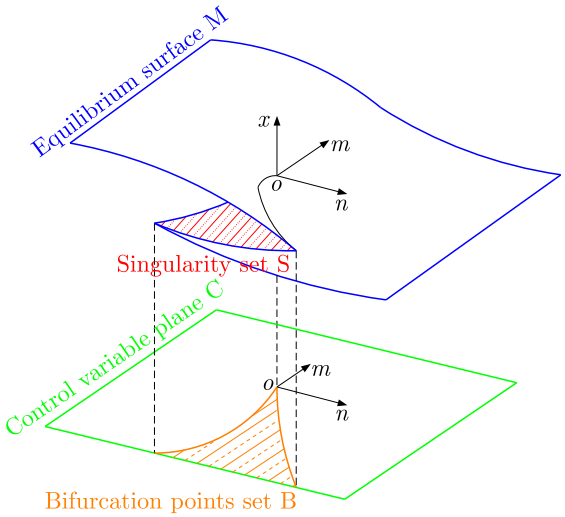


Fig. 3. Diagram of the equilibrium surface and the bifurcation set of cusp catastrophe model

3. Analysis and discussion

3.1. Catastrophic instability analysis

As the total potential energy function of DSLS conforms to the cusp catastrophe model, it will be analyzed by the cusp catastrophe model. In the cusp catastrophic model, the values of control variables m and n directly determine the stable state of the shaft. It is only when the equilibrium point crosses the bifurcation points set B when there is a possibility of vertical instability. The bifurcation set equation composed of control variables is the key to determining system instability (Saunders, 1980; Zhao *et al.*, 2023). When $\Delta > 0$, the control variables (m, n)

fall outside the bifurcation points set B, the system is located on the upper and lower leaves of the equilibrium surface M, the system (here refers to ultralong DSLS) is in a stable equilibrium state. When $\Delta < 0$, the control variables (m, n) fall within the bifurcation points set B, the system is located in the middle of the equilibrium surface M, and ultra long DSLS is in an unstable equilibrium state. When $\Delta = 0$, the control variables (m, n) fall on the boundary of the bifurcation points set B, the ultralong DSLS is in a critical stable equilibrium state.

The vertical instability and failure of ultralong DSLS is an evolutionary process from gradual to sudden changes. With a continuous connection of the shaft segment and the increasing injection of counterweight water into the shaft, the shaft structure begins to accumulate elastic potential energy. The total potential energy accumulated by the shaft structure maintains a stable equilibrium with dynamics and uncertainty. When the elastic potential energy of the shaft accumulates to a certain extent, the structure in a dynamic equilibrium state will induce structural instability by certain conditions.

3.2. Design of critical depth for instability and instability criterion of ultra long DSLS

Observe equations (2.6), as $A_4 = 0$, so $n = 0$ and

$$\Delta = 4m^3 = 4\Big(\frac{A_3}{\sqrt{A_2}}\Big)^3 = 4\left(\frac{\frac{41EI\pi^4}{32H^3} + \Big(1 - \frac{5}{16}\pi^2\Big)p_c - \frac{\pi d^2\gamma_w}{2}\sin^2\frac{\pi H_w}{2H}\sin^2\frac{\pi H_w}{H}}{\sqrt{-\frac{A_1\pi^5 d^2\gamma_w}{128H^2} + \frac{\pi^3 D^2\gamma_m}{4}\Big(\frac{183\pi^2}{256H^2} + \frac{308}{75}\Big)}}\right)^3 \tag{3.1}$$

When ultra long DSLS is in a critical stable equilibrium state, then the $\Delta = 0$. From Eq. (3.1) one can obtain the critical depth H_{cr} for instability of ultra long DSLS

$$H_{cr} = \sqrt[3]{\frac{41EI\pi^4}{(10\pi^2 - 32)p_c + 16\pi d^2\gamma_w \sin^2\frac{\pi H_w}{2H}\sin^2\frac{\pi H_w}{H}}} \tag{3.2}$$

When ultra long DSLS is in a stable equilibrium state, the $\Delta > 0$. From Eq. (3.1) one can find that the construction height of ultra long DSLS should satisfy

$$H < \sqrt[3]{\frac{41EI\pi^4}{(10\pi^2 - 32)p_c + 16\pi d^2\gamma_w \sin^2\frac{\pi H_w}{2H}\sin^2\frac{\pi H_w}{H}}} \tag{3.3}$$

On the right-hand side of inequality (3.3) there is just the critical depth H_{cr}, so inequality (3.3) can also be expressed as the relationship between construction height and critical depth $H < H_{cr}$.

When ultra long DSLS is in an unstable equilibrium state, the $\Delta < 0$. From Eq. (3.1) one can find that the construction height of ultra long DSLS should satisfy

$$H > \sqrt[3]{\frac{41EI\pi^4}{(10\pi^2 - 32)p_c + 16\pi d^2\gamma_w \sin^2\frac{\pi H_w}{2H}\sin^2\frac{\pi H_w}{H}}} \tag{3.4}$$

Simultaneously with Eq. (3.2), the instability conditions of ultra long DSLS can also be expressed as the relationship between construction height and critical depth $H > H_{cr}$.

Thus, the following can be used to convey ultra long DSLS instability criterion when applying the instability optimization measure

$$\begin{aligned} &H < H_{cr} \quad \text{ultra long DSLS is stable} \\ &H = H_{cr} \quad \text{ultra long DSLS is in critical instability} \\ &H > H_{cr} \quad \text{ultra long DSLS is unstable} \end{aligned} \tag{3.5}$$

3.3. Discussion of critical depth

In this part, a numerical study is conducted to investigate the influence effect of different factors on the critical depth. It can be seen from Eq. (3.2) that the factors affecting the critical depth include the shaft elastic modulus E, inertia moment I, inner diameter d, self-weight per unit length p_c and height of counterweight water H_w. But the shaft inertia moment is determined by the inner and outer diameters, while the outer diameter is generally a fixed value of 8 m or more, so the moment of inertia is determined by the inner diameter. The actual influence factors only contain E, d, p_c and H_w. Among them, E, d, p_c are the internal characteristics. H_w is the external factor. The process leading to the instability of ultra long DSLS when applying the instability optimization measure is a result of the combined action of its internal characteristics and external factors.

In ultra long shaft engineering, these four influence factors have a certain range. As the shaft segment is generally made of C35C̃80 concrete (Fang *et al.*, 2023), E ranges from 31.5 GPa to 38 GPa. The outside diameter of the shaft is often greater than 8 m. The thickness ranges from 500 mm to 850 mm according to the shaft strength requirement. So d ranges from 6.3 m to 7 m. p_c is related to the making method of the shaft segment. Generally, it is no less than $3 \cdot 10^5$ N/m. H_w is required to meet the shaft suspension and anti-overturning standards. The minimum and maximum heights of H_w is restricted. Usually, H_w will be greater than half of the shaft construction height. Here we choose three shaft construction heights, 650 m, 700 m and 750 m, and calculate the critical depth for fifteen independent variables within their reasonable range. The basic parameters are set to be $D = 8$ m, $d = 6.3$ m, $p_c = 3 \cdot 10^5$ N/m, $H_w = 380$ m. The critical depth with different elastic modulus is shown in Fig. 4a. The calculation results under three construction heights all show that the larger the elastic modulus, the greater the critical depth. And the critical depths are all greater than the corresponding construction height, the shaft is vertically stable. It illustrates that the use of high-strength materials is beneficial for improving the vertical stability of ultra long DSLS.

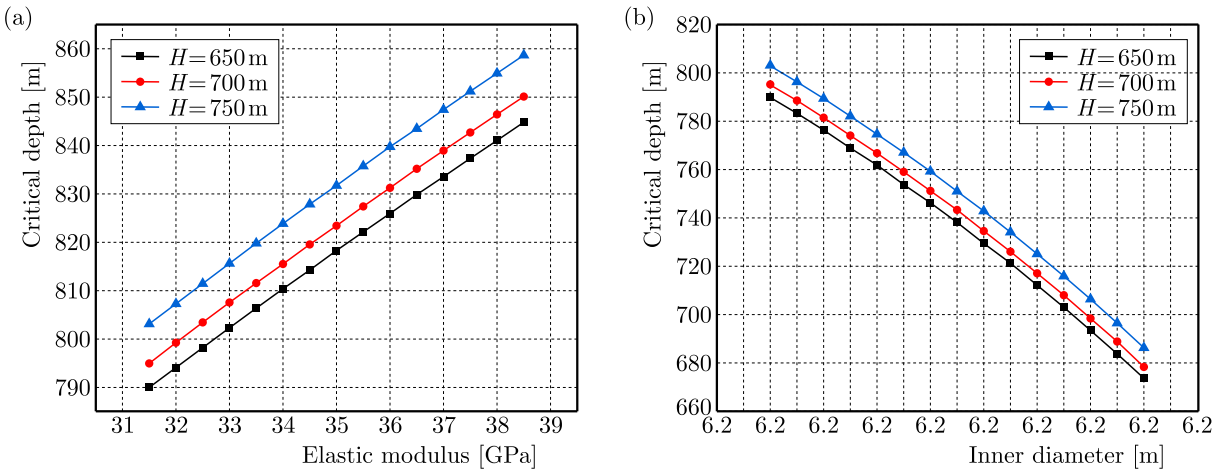


Fig. 4. Critical depth with different elastic moduli (a) and inner diameter (b) for three construction heights

Another investigation was devised to investigate the relationship between critical depth and inner diameter. When taking $E = 31.5$ GPa, other parameters are also set to be $D = 8$ m, $p_c = 3 \cdot 10^5$ N/m, $H_w = 380$ m. The critical depth with different inner diameters is shown in Fig. 4b. The calculation results under three construction heights all show that the smaller the inner diameter, the greater the critical depth. For the construction height of 650 m, the critical depth is greater than construction height and the shaft is vertically stable. But for the construction height of 700 m and 750 m, the critical depth is not always greater than construction

height and the structure may not necessarily be stable. Due to relatively fixed outer diameter in ultralong DSLS, the size of inner diameter is closely related to thickness of the shaft. It illustrates that using a smaller inner diameter or thicker shaft is more conducive to structural stability. When construction height exceeds 700 meters, it is necessary to control its inner diameter or thickness.

Another investigation was developed to explore the relationship between critical depth and unit length self-weight. When taking $D = 8\,\text{m}$, $d = 6.3\,\text{m}$, $E = 31.5\,\text{GPa}$, $H_w = 380\,\text{m}$, the critical depth with different unit length self-weight is shown in Fig. 5a. The calculation results under three construction heights all show that the larger the inner diameter, the smaller the critical depth. For the construction height of 650 m, the critical depth is greater than construction height and the structure is vertically stable. But for the construction height of 700 m and 750 m, the critical depth is not always greater than construction height and the structure may not necessarily be stable. It suggests that using a lightweight shaft material is more conducive to structural stability. When construction height exceeds 700 meters, controlling its unit length self-weight is critical.

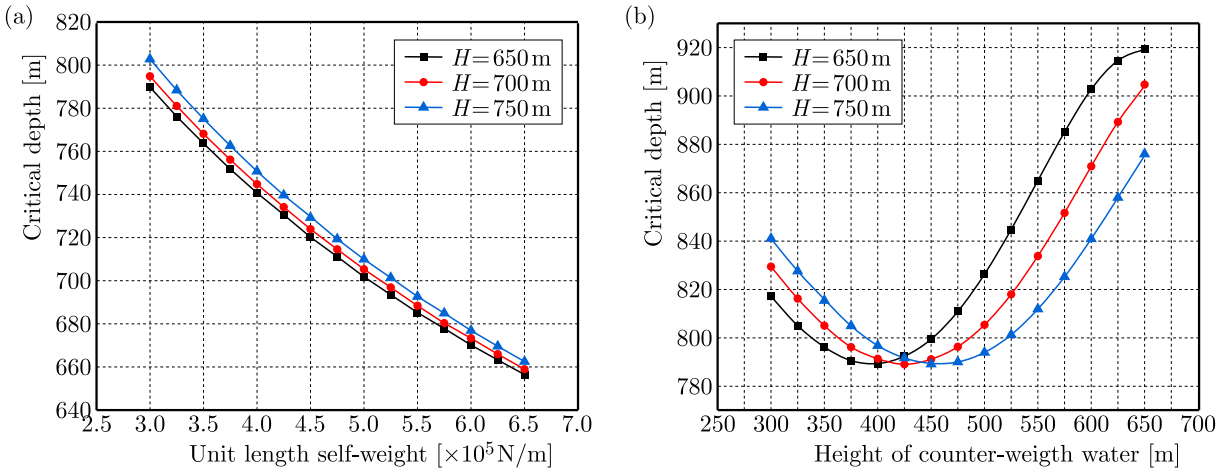


Fig. 5. Critical depth with different self-weights per unit length (a) height of counterweight water (b) at three construction heights

An additional study was carried out the to explore the relationship between critical depth and counterweight water height. When taking $D = 8\,\text{m}$, $d = 6.3\,\text{m}$, $E = 31.5\,\text{GPa}$, $p+c = 3\cdot10^5\,\text{N/m}$, the critical depth with different counterweight water heights is shown in Fig. 5b. The calculation results under three construction heights all show that the critical depth basically increases with the increase of counterweight water height, but there is a sudden change in the critical depth when counterweight water height reaches about 425 m. Critical depths for three counterweight water heights are all greater than construction height and the structure is vertically stable. It suggests that controlling counterweight water height within a reasonable range is favorable for ultra long DSLS.

Multifactor sensitivity analysis (Wang *et al.*, 2023) is a technique for assessing the effect of several uncertain elements changing simultaneously in a process. It can help one to analyze the effect of each factor on analytical indicators at various amplitudes, Figure 6 gives the calculation results of sensitivity coefficients for four influencing factors on the critical depth. The results reveal that factors described in order of the degree of influence are the inner diameter, elastic modulus, unit length self-weight, and counterweight water height sequentially. And the order of influencing factors forr three construction heights is exactly the same.

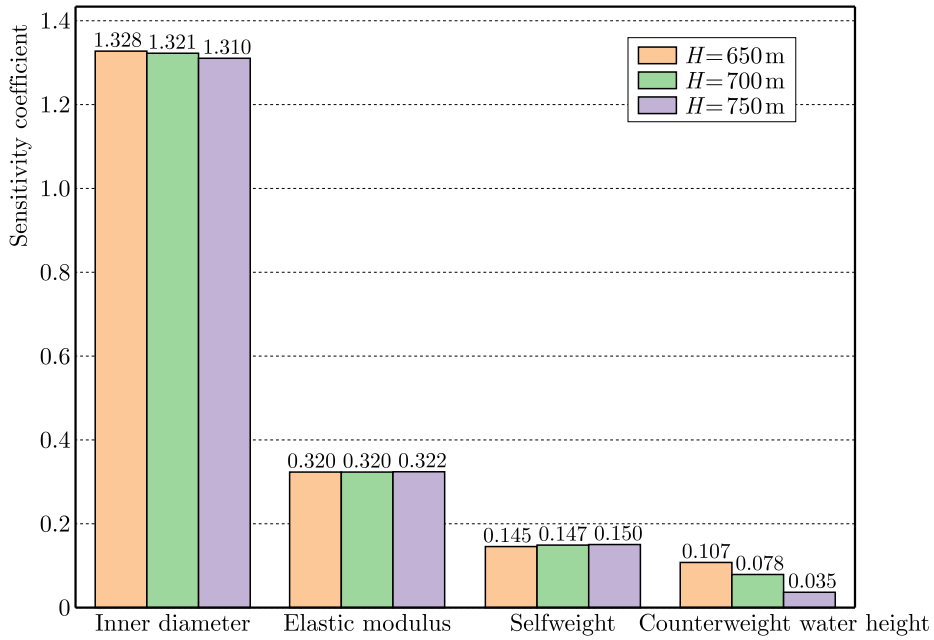


Fig. 6. Multi factor sensitivity analysis

4. Validation

4.1. Engineering background

North air shaft of Taohutu coal mine is located in the central eastern part of the Maowusu Desert in northeast China. The shaft is constructed by the drilling method, with a borehole diameter of 9.4 m, outer diameter of 8.1 m, minimum inner diameter of 6.5 m, total construction height of 751 m. Based on the engineering geological and hydrogeological conditions of the shaft passing through the strata, as well as the situation where the shaft will bear loads, it is determined that the DSLS adopts a combination of reinforced concrete shaft structure, double-layer inner and outer steel plates-reinforced concrete composite shaft structure. In the range of elevation $\pm 0 \sim -396$ and $-747 \sim -751$, a reinforced concrete shaft structure is used. Between $-396 \sim -747$, a double-layer steel plate reinforced concrete composite shaft structure is used. Concrete grade has a minimum strength of C40 and a maximum strength of C75.

4.2. Numerical validation

The parameters of shaft engineering are determined by calculation and normalization. The calculated parameters for the shaft elastic modulus, inner diameter and unit length self-weight are obtained by the ratio of cumulative sum of the product of each shaft section height and elastic modulus, inner diameter, or unit length self-weight to the total length of the shaft. So $E = 39.48$ GPa, $d = 6.8$ m, $p_c = 4.45 \cdot 10^5$ KN/m, $H_w = 423.8$ m. Following substituting into formula (3.2), the critical depth is 758 m when using the instability optimization measure. Due to the critical depth being greater than the total construction height, the shaft structure is stable. Theoretically, the shaft will not experience vertical instability after implementing the instability optimization measure. To further verify the effectiveness of instability optimization measure comparatively, we calculate the critical depth under traditional construction techniques according to literature in which the critical depth is also determined by cusp catastrophic model. It is found that the critical depth is 522.7 m, which is significantly smaller than the total construction height. So if traditional construction techniques are used, the shaft will experience vertical instability. Table 1 lists the comparison calculation results between the instability optimization measure and traditional techniques for Taohutu North air shaft.

As shown in Table 1, the critical depth found by using instability optimization measure is 1.45 times greater than that of traditional construction techniques. The critical depth of instability optimization measure increases by 45% compared to traditional construction techniques. It validates that the instability optimization measure effectively mitigates the instability of DSLS in Taohutu North air shaft.

Table 1. Comparison between instability optimization measure and traditional techniques for Taohutu North air shaft

	Instability optimization measure	Traditional construction techniques
Basis for critical depth calculation	Equation (3.2)	Liu *et al.* (2.20)
Critical depth calculation results	758 m	522.7 m
Instability criterion	$H < H_{cr}$	$H > H_{cr}$
Theoretical analysis conclusion	stable	unstable

5. Conclusion

By considering the characteristics of obvious suddenness and irreversible damage, the catastrophe method and cusp catastrophe model is employed to describe the instability process and mechanism of ultralong DSLS when applying the instability optimization measures. The following conclusions can be drawn from the numerical analysis:

- The cusp catastrophe model proposed in this paper can accurately describe the structural buckling behavior of ultralong DSLS under the instability optimization measure. The bifurcation set equation composed of control variables helped one to analyze the stability characteristics.
- The ultra long DSLS vertical instability and failure are the result of a gradual trend toward abrupt changes. Quantitative calculation of the critical depth helps one to evaluate stability of the shaft. Only when the critical depth reaches construction height, the shaft is in a critical stable condition.
- The critical depth of this instability optimization measure is dependent on the material properties and external factors. It decreases with increasing inner diameter and self-weight per unit length, and rises with increasing elastic model and counterweight water height. Lightweight and high-strength materials, as well as thick shaft structures are more popular in utilizing the instability optimization measure.
- Adopting instability optimization measure can increase the critical depth by approximately 45% compared to the traditional construction method. This instability optimization measure can theoretically mitigate the structural instability of DSLS, and its feasibility would be approved by more and more engineering.

Acknowledgement

This research was supported by Scientific Research Foundation for High-level Talents of Anhui University of Science and Technology (2024yjrc04).

References

1. Cheng H., Liu J.-M., Rong C.-X., Yao Z.-S., 2008, Variable cross section shaft drilling lining's vertical stability in thick alluvium, *Journal of China Coal Society*, **33**, 12, 1351-1357

2. Dorn E., Kaledin O., 2013, Modern and innovative shaft sinking and construction technology with examples from current projects/Moderne und innovative Schachtbautechnik am Beispiel aktueller Abteufprojekte, *Geomechanics and Tunnelling*, **6**, 5, 574-581

3. Fang Y., Yao Z., Xu Y., Li X., Liu X., Huang X., 2023, Permeability evolution and pore characteristics of reactive powder concrete of drilling shaft with initial salt erosion damage, *Construction and Building Materials*, **403**, 133141

4. Hara T., Yashima A., Sawada K., Kariya K., Tsuji H., Soga N., 2019, Small-diameter vertical shafts constructed in the shallow space of steep mountainous areas, *Underground Space*, **4**, 3, 235-250

5. Hong B.-Q., 1980, Axial stability of deep shaft wall in mud by drilling method, *Coal Science and Technology*, **9**, 22-25

6. Karman A., Pawlowski M., 2022, Circular economy competitiveness evaluation model based on the catastrophe progression method, *Journal of Environmental Management*, **303**, 114223

7. Lei Y., Li P.-J., Liu Z.-Y., Li J.-Z., Hu W., 2022, Method for calculation of buckling critical load of pile foundation crossing karst cave in karst area, *Rock and Soil Mechanics*, **43**, 3347-3356

8. Li Y., Tang N., Jiang X., 2016, Bayesian approaches for analyzing earthquake catastrophic risk, *Insurance: Mathematics and Economics*, **68**, 110-119

9. Liu J., Cheng H., Rong C., Wang C., 2020, Analysis of cusp catastrophic model for vertical stability of drilling shaft lining, *Advances in Civil Engineering*, **2020**, 1, 8891751

10. Liu J., Liu Y., Fu X., 2022, Mechanism analysis of catastrophic instability of drilling shaft lining based on Python language, *Coal Science and Technology*, **50**, 9, 75-81

11. Liu Z., Meng Y., 2015, Key technologies of drilling process with raise boring method, *Journal of Rock Mechanics and Geotechnical Engineering*, **7**, 4, 385-394

12. Merli M., Pavese A., 2018, Electron-density critical points analysis and catastrophe theory to forecast structure instability in periodic solids, *Acta Crystallographica. Section A: Foundation and Advances*, **74**, 102-111

13. Niu J.M., Wu J.H., Liu X.L., Yuan M.Q., Wang L.B., 2023, Quantitative analysis of acoustic black hole property by the catastrophe theory, *International Journal of Mechanical Sciences*, **259**, 108621

14. Niu X.-C., Hong B.-Q., Yang R.-S., 2006, Study on vertical structural stability of bored shafts filled part of water, *Rock and Soil Mechanics*, **27**, 1897-1901

15. Ozbasaran H., 2018, Optimal design of I-section beam-columns with stress, non-linear deflection and stability constraints, *Engineering Structure*, **171**, 385-394

16. Pacoste C., 1993, *On the Application of Catastrophe Theory to Stability Analyses of Elastic Structures*, Royal Institute of Technology, Stockholm

17. Saunders P.T., 1980, *An Introduction to Catastrophe Theory*, Cambridge University Press, Cambridge

18. Wang E., Zhang L., Chen H., Zhang X., 2023, Collinearity-oriented sensitivity analysis for patterning energy factor significance in buildings, *Journal of Building Engineering*, **73**, 106685

19. Wegner T., Kurpisz D., 2017, An energy-based method in phenomenological description of mechanical properties of nonlinear materials under plane stress, *Journal of Theoretical and Applied Mechanics*, **55**, 1, 129-139

20. Xing X., Ma B., Du J., Liu Y., Liu Z., Han X., 2023, An energy method for predicting and suppressing the instability of a three-dimensional thermoacoustic coupling system with a microperforated plate, *Journal of Theoretical and Applied Mechanics*, **61**, 4, 755-768

21. Xu J., Ni Y., 2019, Prediction of grey-catastrophe destabilization time of a granite residual soil slope under rainfall, *Bulletin of Engineering Geology and the Environment*, **78**, 5687-5693

22. Yang S., 2019, Influence analysis of buckling stability of super-long piles under different working conditions based on energy method, Master's Thesis, China University of Geosciences, Wuhan, China

23. Zhao K., Wu J., Yan Y., Yang J., Tian X., *et al.*, 2023, The cusp catastrophe warning model for easily weathered granite considering energy accumulation and dissipation, *Environmental Science and Pollution Research*, **30**, 45888-45898

Manuscript received February 3, 2024; accepted for publication September 30, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 841-854, Warsaw 2024
https://doi.org/10.15632/jtam-pl/194502

EVALUATION OF TRIBOLOGICAL BEHAVIOR OF TRIBOPAIRS IN WATER HYDRAULIC RADIAL PISTON PUMPS WITH MULTIPLE COATINGS AND LUBRICATION

Ran Li

Graduate School, China Coal Research Institute, Beijing, China, and

Beijing Tianma Intelligent Control Technology Co., Ltd., Beijing, China, and

College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao, China

Wenshu Wei, Hao Liu, Xiangyu Geng, Jianfeng Li, Yuehua Lai, Rongming Chen, Dongze Li, Tianliang Yu

Beijing Tianma Intelligent Control Technology Co., Ltd., Beijing, China

corresponding author Yuehua Lai, e-mail: laiyh@mktm.com.cn

Dalong Wang

School of Mechanical Electronic and Information Engineering, China University of Mining and Technology-Beijing, Beijing, China

Jian Ye, Shoubin Li, Wei Wang, Huigang Wu

Beijing Tianma Intelligent Control Technology Co., Ltd., Beijing, China

In order to explore the tribological behavior of slipper/eccentric cam tribopairs under different lubrication conditions at power ends of water hydraulic radial piston pumps, friction-wear tests were carried out. It is found that tribopairs under water-based lubrication exhibited better friction-reducing performances than those under water lubrication alone. The average steady-state friction coefficient under water-based lubrication was between 0.17 and 0.23. HMn62-3-3-0.7/17-4PH tribopair using water-based lubrication exhibited the lowest wear rate of $1.1 \cdot 10^{-12}\,\mathrm{mm}^3/\mathrm{N{\cdot}m}$. The findings of this study offer a reference for the identification of potential material pairs for slipper/eccentric cam tribopairs.

Keywords: water hydraulic radial piston pump, tribopair, friction-wear, HMn62-3-3-0.7, water-based lubrication

1. Introduction

Water hydraulic radial piston pumps (WHRPPs) are widely used in various industrial applications, such as seawater desalination, descaling, and coal mining high hydraulic power (Cotter, 2000; Li *et al.*, 2023, 2024; Zhao *et al.*, 2017). WHRPPs use water or seawater both as a working medium at the liquid ends and as a lubrication medium at the power ends for low-speed and low-pressure applications (Zhang *et al.*, 2017). For high-speed and high-pressure applications, mineral oils are generally used as lubrication media at the power ends of WHRPPs separated by dynamic seals. Damage or failure of dynamic seals results in emulsification of the lubricating oil, which may lead to possible failure of the components, such as slipper/eccentric cam tribopairs at the power ends of the WHRPPs. Significant advancements have been achieved in the design and application of WHRPPs in recent years owing to creation of novel lubricating media that combine exceptional anti-corrosion, anti-wear, and environmentally friendly properties. Therefore, it is important to understand and predict tribological behavior of slipper/eccentric cam tribopairs under different lubrication conditions at the power ends of WHRPPs.

Many studies have been conducted to investigate tribological or lubricating behavior of slipper/swashplate tribopairs in piston pumps under water or seawater lubrication conditions (Nie *et al.*, 2006; Yinshui *et al.*, 2009; Ma *et al.*, 2015; Liang *et al.*, 2023; Nie *et al.*, 2019; Yin *et al.*, 2023; Kou *et al.*, 2022; Wu *et al.*, 2018). Recently, Wei *et al.* (2019) conducted frictional corrosion tests on PEEK/AISI 630 tribopairs on representative slipper/swashplate tribopairs in water hydraulic axial piston pumps under dry, pure water, and seawater lubrication conditions. Wu *et al.* (2020) conducted comparative studies on wear tests for three types of slipper materials, 1Cr18Ni9Ti, ZQSn10-1, and ZQAl10-4-4, and two types of swashplate materials, 38CrMoAl and 45 steel, under dry sliding conditions. Despite the inherent safety and environmental protection of water and seawater lubrication, their applications for long-term and high-efficiency operations are still limited owing to insufficient lubrication performance and low viscosity. Therefore, the addition of functional additives to water-based lubricants, such as water-glycol hydraulic fluids, to improve their tribological and anticorrosive behavior is one of the focal points of research. Zhou *et al.* (2010) studied friction and wear behavior of a-CNx/SiC tribopairs in an ethylene glycol (EG) aqueous solution and found that the concentration of 10 vol.% EG gave the lowest friction coefficient of 0.019. Espinosa *et al.* (2014) showed that an ultra-low minimum friction coefficient of 0.0001 could be obtained for a sapphire/AISI 316 L contact lubricated by a water +1 wt% solution of bis(2-hydroxyethylammonium) succinate.

Several studies have been conducted to develop a novel WHRPP in the field of underground coal mining using environmentally friendly and anticorrosion lubrication (Li *et al.*, 2023, 2024). In this study, friction and wear tests were conducted using pin-on-disk specimens to represent slipper/eccentric cam tribopairs in a WHRPP under various lubrication conditions. Three Cu-based slipper materials and four types of eccentric cam materials/coatings were used. The material and lubrication dependencies were determined and analyzed based on the friction coefficients and wear rates for different pin-on-disk tribopairs. The wear mechanisms of the three Cu-based pin specimens sliding against the 42CrMo4 lower specimens with quench polish quench (QPQ) coatings, which are specialized types of nitrocarburizing case hardening that increases corrosion resistance, were characterized using scanning electron microscopy (SEM) and energy-dispersive X-ray spectroscopy (EDS) analyses. The novel contribution of this study is the identification of potential material pairs for application as slipper/eccentric cam tribopairs in water hydraulic piston pumps under water-based lubrication.

2. Experiments

The effects of lubricants on the tribological behavior of slipper/eccentric cam tribopairs in the WHRPPs were investigated by pin-on-disk wear tests using SRV IV friction tester shown in Fig. 1 (frequency = 50 Hz, stroke = 1 mm, test load = 100 N, test duration = 1 h, and temperature = 50°C). The pin-on-disc contact surfaces were wetted with 0.5 mL lubricant before each test.

As shown in Fig. 2, the upper specimens were designed with a diameter of 9.3 mm which tapered to a diameter of 2 mm at the contact surface, and a total height of 10 mm. The diameter and height of the lower specimens were 24.4 mm and 7.9 mm, respectively.

Three Cu-based materials were selected that were potentially suitable for slipper/eccentric cam tribopairs, i.e., HMn62-3-3-0.7, QAL10-5-5 and ZCuPb15Sn8.

The lower specimens consisted of different steel: martensitic precipitation-hardening stainless steel, 17-4PH, alloy steel, 42CrMo4; and nitriding steel 38CrMoAl. A QPQ coating was used to increase the hardness and corrosion resistance of 42CrMo4 and 38CrMoAl lower specimens at a low cost and with low part distortion. The 42CrMo4 lower specimen was also surface-treated with 15 wt.% tungsten carbide (WC)-reinforced Ni-based coating to improve its wear and corrosion resistance when subjected to water-based lubrication (Liu *et al.*, 2001).

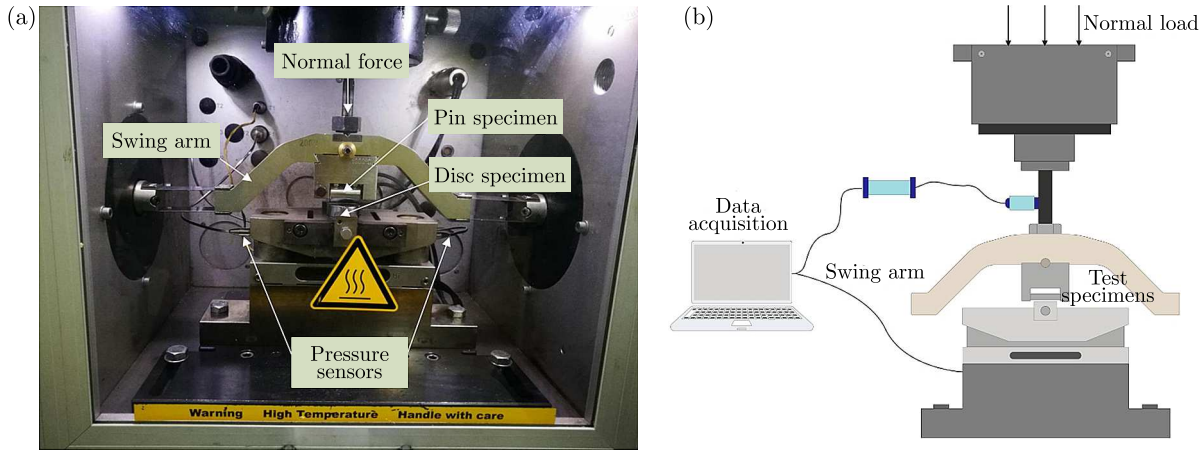


Fig. 1. (a) Test chamber of SRV IV test apparatus; (b) schematic representation of SRV IV test apparatus

Fig. 2. Representation of slipper/eccentric cam pin-on-disc tribopairs

3. Results and analysis

3.1. Coefficient of friction

The relationship between the friction coefficients and sliding time of HMn62-3-3-0.7 upper specimens sliding against four types of lower specimens under different lubrication conditions: distilled water, GL-5 85 W-90 oil, and 8 vol.% FF330 water-based lubrication, were determined experimentally, as shown in Fig. 3. The friction coefficients for HMn62-3-3-0.7 sliding against the four types of lower specimens under oil-lubricated conditions reached an approximately steady state, with values in the range of 0.16-0.20. In the water-lubricated condition, the friction coefficients for HMn62-3-3-0.7 sliding against the lower specimens 17-4PH, 42CrMo4-QPQ, and 38CrMoAl-QPQ were in the range of 0.3-0.32, a factor of 1.6 times of those for the oil-lubricated condition. A significant fluctuation in the friction coefficient for the HMn62-3-3-0.7/38CrMoAl-QPQ tribopair, with a peak value of approximately 0.80, indicated that severe wear may occur for each fluctuation. For a 8 vol.% FF330 water-based lubrication, the steady-state friction coefficients for the HMn62-3-3-0.7 sliding against the four types of lower specimens were nearly identical to those under the oil lubricating condition, except for a small amount of fluctuations in the steady states.

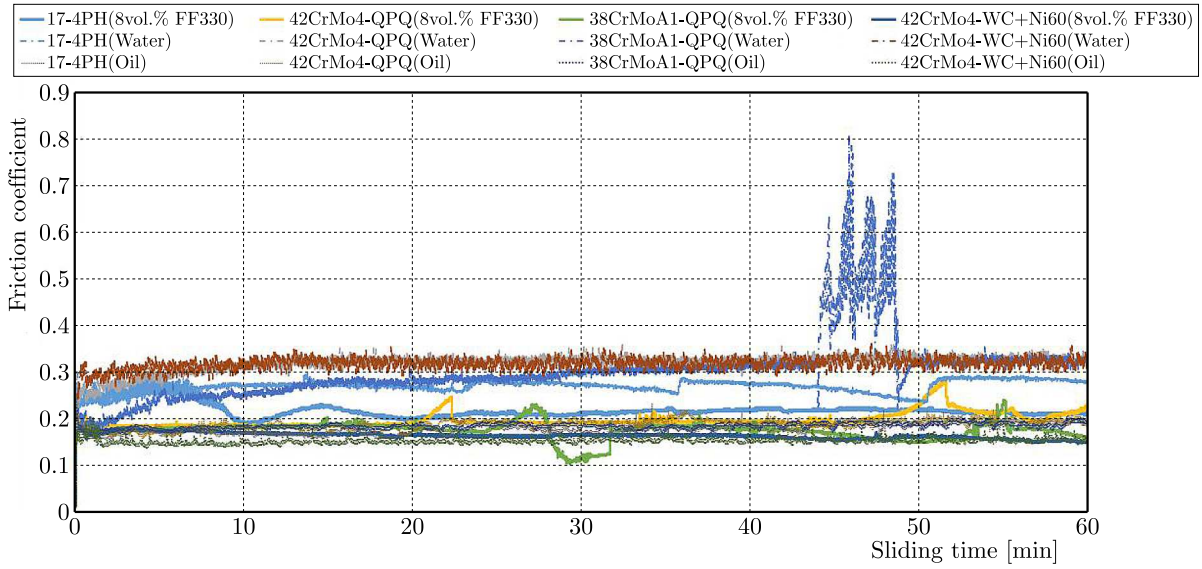


Fig. 3. Variation in friction coefficients of HMn62-3-3-0.7 upper specimens sliding against four types of lower specimens under different lubricating conditions

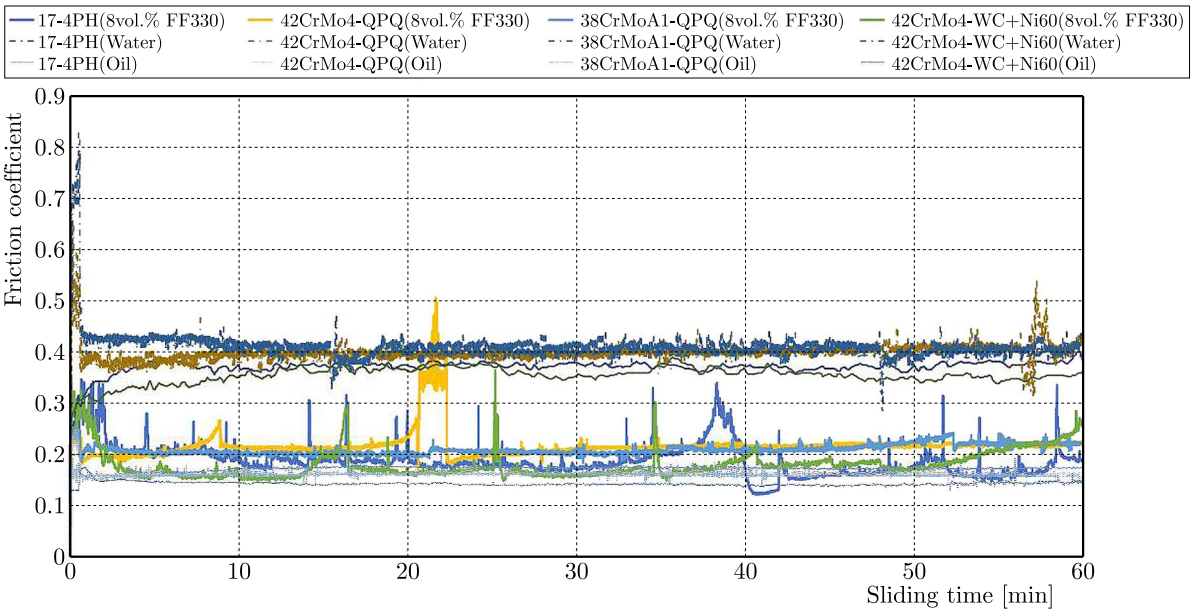


Fig. 4. Variation in friction coefficients of QAL10-5-5 upper specimens sliding against four types of lower specimens under different lubricating conditions

Figure 4 shows the relationship between friction coefficients and sliding time of QAL10-5-5 upper specimens sliding against the four types of lower specimens under three different lubrication conditions. The trend of variation in friction coefficients for QAL10-5-5 upper specimens under oil-lubricated conditions was comparable to that for HMn62-3-3-0.7 upper specimens, which fluctuated greatly in the initial stage, then gradually increased, and finally reached an approximate steady state. The friction coefficients for the oil-lubricated condition are in the range of 0.14-0.17. For HMn62-3-3-0.7 tribopairs in the water-lubricated condition, the steady-state friction coefficients are in the range of 0.35-0.41 for QAL10-5-5 upper specimens against the four types of lower specimens, except for the occurrence of a peak fluctuation of a value of 0.537 for the QAL10-5-5/42CrMo4-QPQ. Numerous fluctuations in the steady-state friction coefficients were observed for the QAL10-5-5/17-4PH, QAL10-5-5/42CrMo4-QPQ, and QAL10-5-

-5/42CrMo4-WC+Ni tribopairs under 8 vol.% FF330 lubrication. The 8 vol.% FF330 lubrication provides good anti-wear behavior for the QAL10-5-5/38CrMoAl-QPQ tribopair, which gives a steady state friction coefficient value of 0.22.

The ZCuPb15Sn8 upper specimens sliding against the four types of lower specimens under water-lubricated conditions showed poor anti-wear behavior, and strong fluctuations occurred during the test, as shown in Fig. 5. Fluctuations in the friction coefficient also occurred during the tests of the 8 vol.% FF330 lubricated tribopairs, especially for the ZCuPb15Sn8/42CrMo4-QPQ tribopair which showed strong fluctuations and a peak friction coefficient of 0.47. Oil lubrication is the only method suitable for preventing the ZCuPb15Sn8 from wear for four types of eccentric cam materials. The steady-state friction coefficients for the ZCuPb15Sn8 upper specimens under oil-lubricated conditions are in the range of 0.16-0.21.

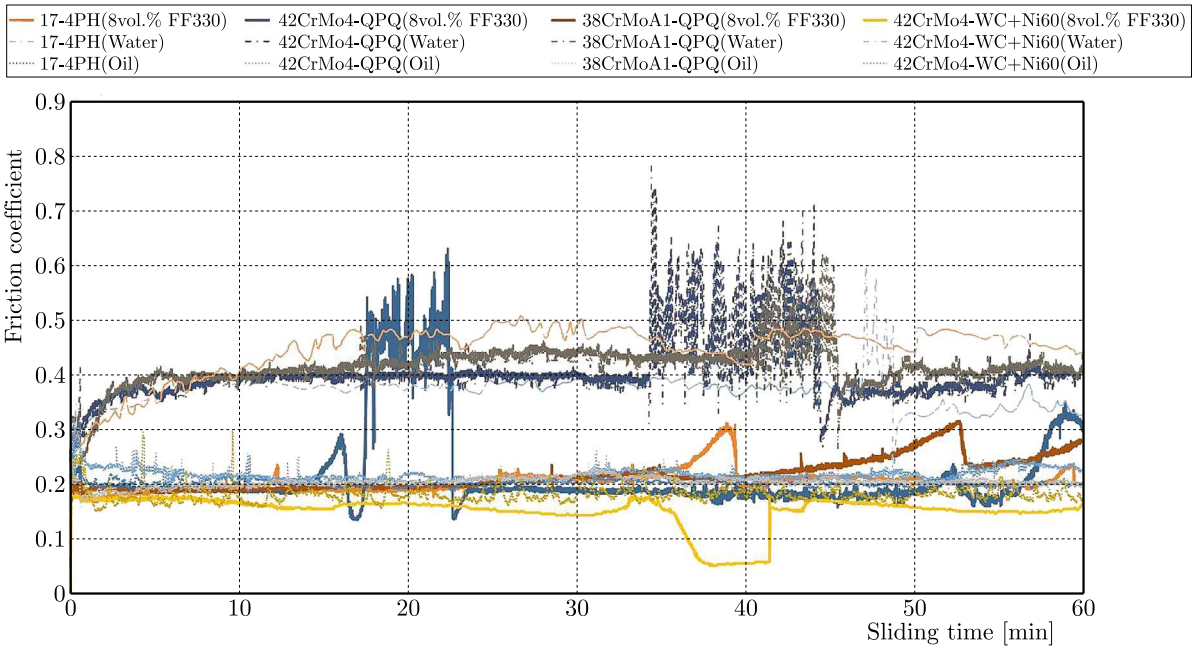


Fig. 5. Variation in friction coefficients of the ZCuPb15Sn8 upper specimens sliding against four types of lower specimens under different lubricating conditions

3.2. Wear rate

The wear rate W_s [mm^3/Nm] of the materials is given by Eq. (3.1), as proposed in previous studies (Zhang *et al.*, 2015; Yin *et al.*, 2021)

$$W_s = \frac{\Delta m}{\rho N L} \tag{3.1}$$

where ρ denotes density of the worn material, N the normal load, L the total sliding distance, and Δm the wear weight. The weight analysis method based on ISO-14242/2 (2016) was used to measure the wear weights of the upper and lower specimens.

The wear rate comparisons among different tribopairs are shown in Fig. 6. Figures 6a-6c show the wear rates of the HMn62-3-3-0.7, QAL10-5-5, and ZCuPb15Sn8 upper specimens against the four types of lower specimens under oil, water, and 8 vol.% FF380 lubrication, respectively. For the HMn62-3-3-0.7 material, it can be confirmed that lubrication with 8 vol.% FF330 results in lower wear rates for all four types of lower specimens, especially for 38CrMoAl-QPQ which has a much smaller wear rate compared to oil and water lubrication. The QAL10-5-5/17-4PH tribopair showed good anti-wear behavior with all three lubricants, as shown in Fig. 6b. Water

lubrication led to substantially higher wear rates for the alloy steels with different coatings. The 8 vol.% FF330 lubrication for the QAL10-5-5 material produced comparable wear rates compared to the values for oil lubrication. Figure 6c shows that the ZCuPb15Sn8/42CrMo4--WC+Ni tribopair has the lowest wear rates for the three lubricants compared to the wear rates for the ZCuPb15Sn8 upper specimens and the other three steel lower specimens. The ZCuPb15Sn8/42CrMo4-QPQ tribopair exhibited inadequate resistance to wear when lubricated with 8 vol.% FF330. This finding aligns with the friction-coefficient results shown in Fig. 5. With the exception of the ZCuPb15Sn8/42CrMo4-QPQ tribopair, the 8 vol.% FF330 lubrication results in lower wear rates for the other tribopairs compared to the values for oil and water lubrication.

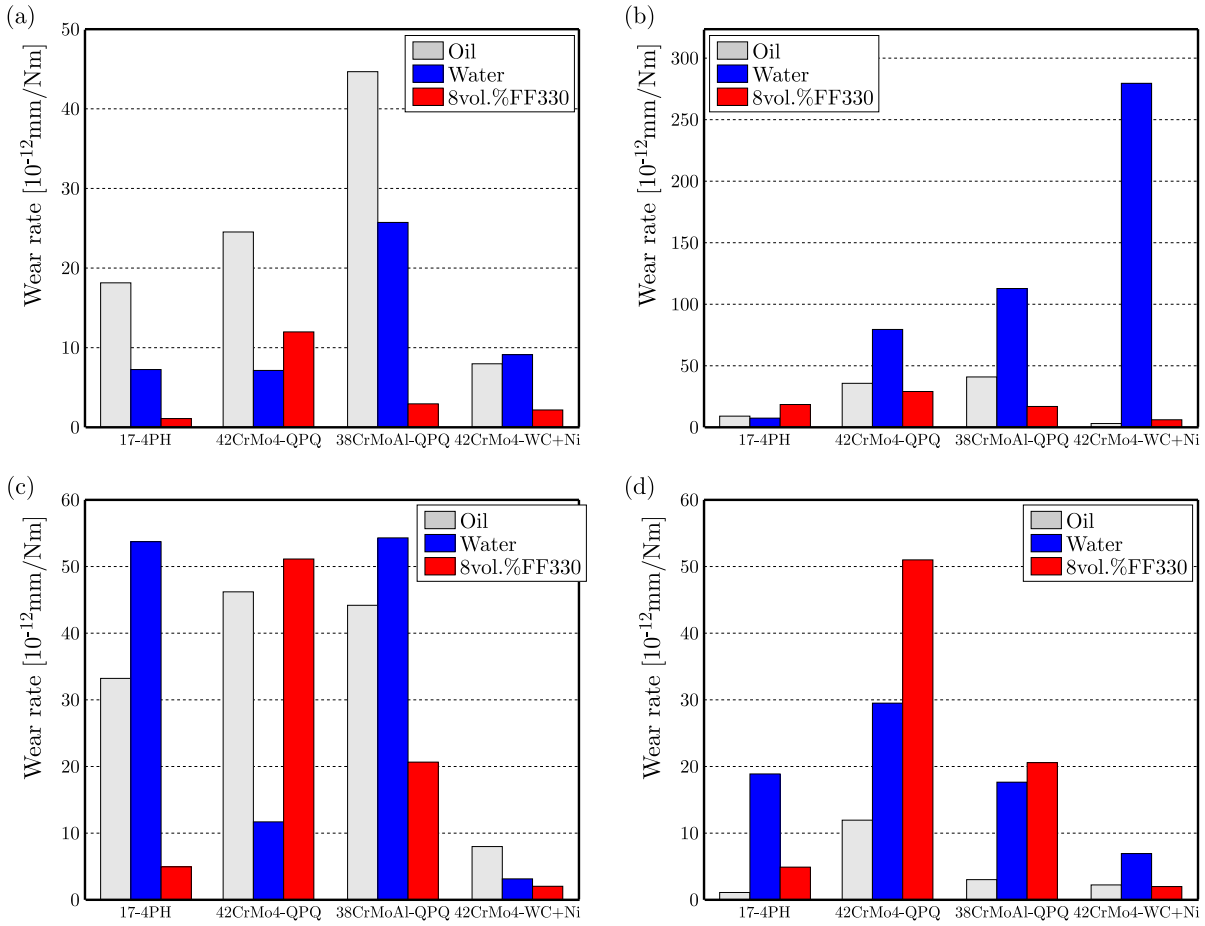


Fig. 6. Wear rate comparisons among different tribopairs: (a) HMn62-3-3-0.7, (b) QAL10-5-5, (c) ZCuPb15Sn8, (d) three Cu-based materials

Figure 6d shows the wear rates of the three Cu-based materials sliding against the four types of eccentric cam materials under 8 vol.% FF330 lubrication. The HMn62-3-3-0.7 material seems to be a good choice for the fabrication of slippers, providing good anti-wear behavior for all four types of lower specimens. Among the four types of lower specimens, the 42CrMo4 alloy steel treated with the spray-welded 15 wt% WC-reinforced Ni-based coating appeared to be the most suitable for making an eccentric cam, which had the lowest wear rates for all three Cu-based upper specimens. The 42CrMo4 alloy steel treated with the QPQ coating should not be used to make an eccentric cam, as it had the highest wear rates for all three Cu-based upper specimens.

3.3. SEM morphology analyses

SEM analyses were performed on the worn surfaces of the three Cu-based upper specimens sliding against the 42CrMo4-QPQ lower specimens under oil, water and 8 vol.% FF380 lubrication to characterize the effect of 8 vol.% FF380 lubrication on the wear mechanism of the tribopairs. Figures 7a to 7f show the worn surfaces of the HMn62-3-3-0.7/42CrMo4-QPQ tribopairs under different lubrication conditions.

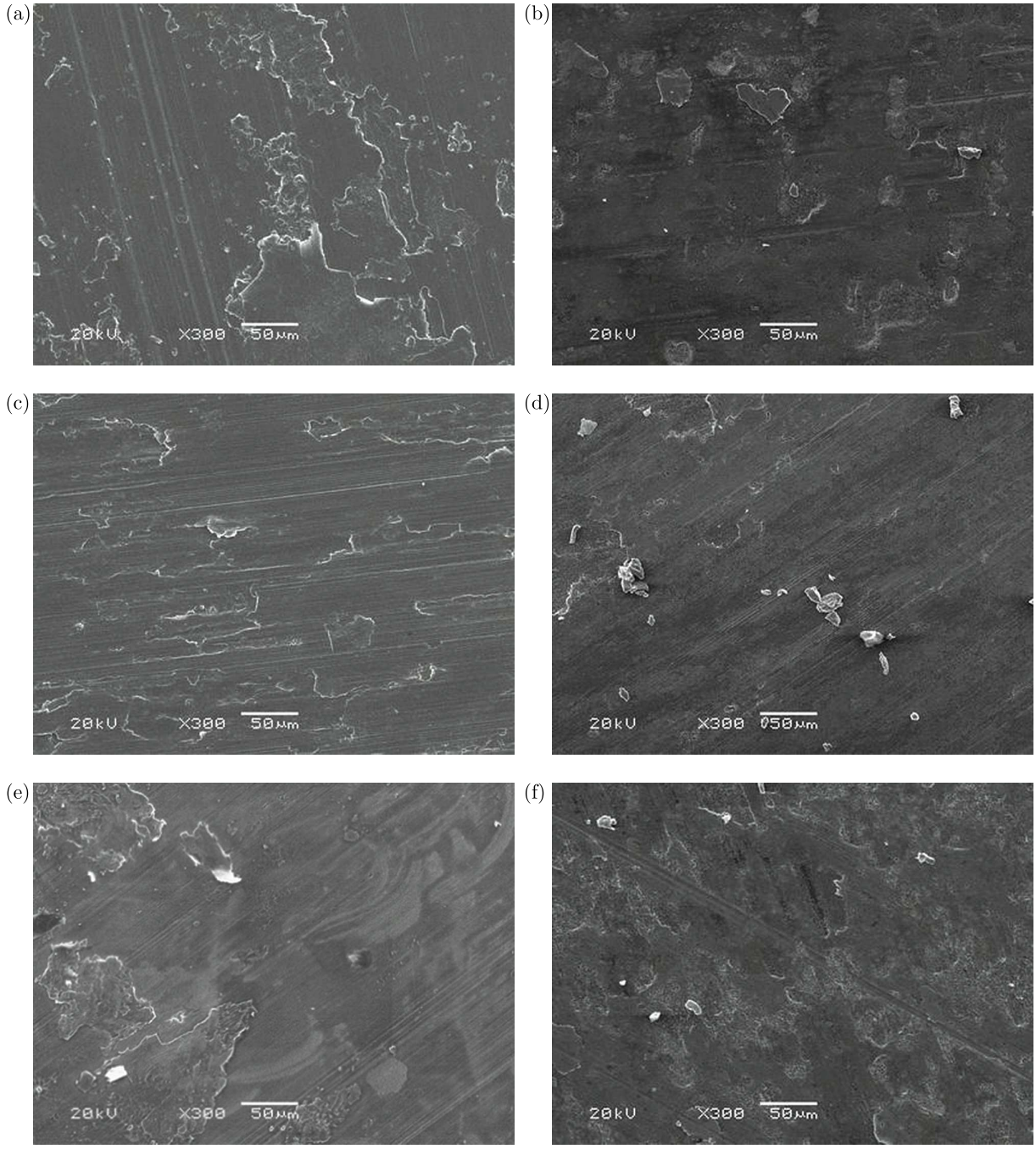


Fig. 7. Worn surfaces of HMn62-3-3-0.7 upper specimens sliding against 42CrMo4-QPQ lower specimens under different lubricating conditions: (a) HMn62-3-3-0.7(oil), (b) 42CrMo4-QPQ(oil), (c) HMn62-3-3-0.7(water), (d) 42CrMo4-QPQ(water), (e) HMn62-3-3-0.7(8vol.% FF330), (f) 42CrMo4-QPQ(8vol.% FF330)

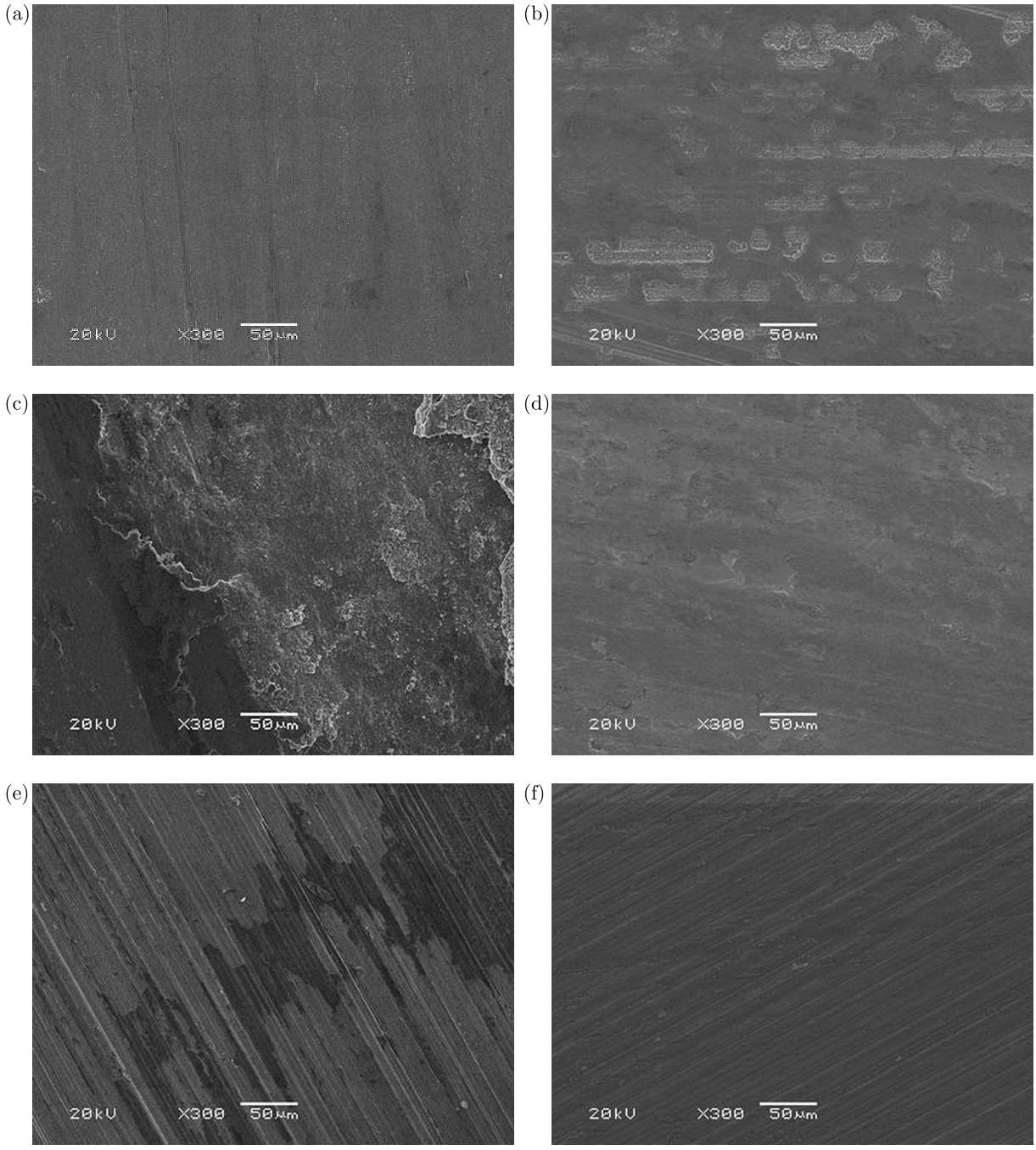


Fig. 8. Worn surfaces of QAL10-5-5 upper specimens sliding against 42CrMo4-QPQ lower specimens under different lubricating conditions: (a) QAL10-5-5(oil), (b) 42CrMo4-QPQ(oil), (c) QAL10-5-5(water), (d) 42CrMo4-QPQ(water), (e) QAL10-5-5(8vol.% FF380), (f) 42CrMo4-QPQ(8vol.% FF380)

As shown in Fig. 7a, wear pits and slight furrows appeared on the worn surface of the HMn62-3-3-0.7 upper specimen when subjected to oil lubrication, which was caused by abrasive wear on the counter surfaces between the HMn62-3-3-0.7 and 42CrMo4-QPQ specimens. There were many grooves on the worn surface of the HMn62-3-3-0.7 upper specimen, indicating that two-body abrasive wear occurred owing to the hard particles embedded in the counter surface of the 42CrMo4-QPQ lower specimen, as shown in Fig. 7c. Plastic deformation occurred around the grooves and oxidation was apparent in these grooves. A large amount of wear debris accumulated on the worn surface of the 42CrMo4-QPQ lower specimen, as shown in Fig. 7d. Adhesive

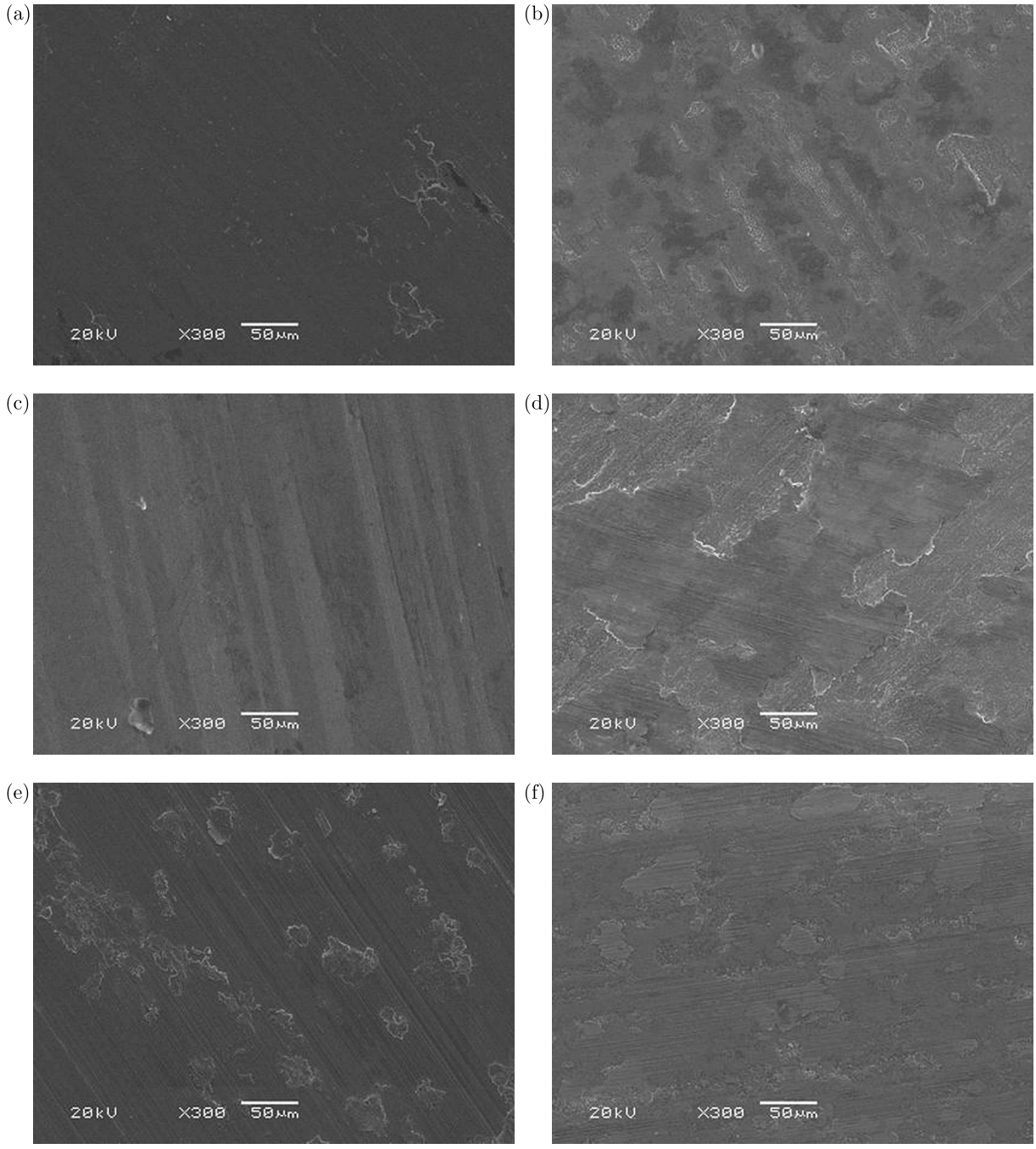


Fig. 9. Worn surfaces of ZCuPb15Sn8 upper specimens sliding against 42CrMo4-QPQ lower specimens under different lubricating conditions: (a) ZCuPb15Sn8(oil), (b) 42CrMo4-QPQ(oil), (c) ZCuPb15Sn8(water), (d) 42CrMo4-QPQ(water), (e) ZCuPb15Sn8(8vol.% FF380), (f) 42CrMo4-QPQ(8vol.% FF380)

wear occurred on the worn surfaces of the HMn62-3-3-0.7 upper specimen and the 42CrMo4--QPQ lower specimen with 8 vol.% FF380 lubrication, as shown in Fig. 7e, which may result in fluctuations in the friction coefficients of the HMn62-3-3-0.7/42CrMo4-QPQ tribopair.

Figures 8a to 8f show the worn surfaces of the QAL10-5-5 upper specimens sliding against the 42CrMo4-QPQ lower specimens under oil, water, and 8 vol.% FF330 lubricating conditions. Figure 8a shows that less wear occurred on the test surface of the QAL10-5-5 upper specimen under oil lubrication. Figure 8c shows that severe adhesive wear occurred on the test surface of the QAL10-5-5 upper specimen under water lubrication. As shown in Fig. 8e, scratches and

oxidation occurred on the worn surfaces of the QAL10-5-5 upper specimen under 8 vol.% FF330 lubrication. A comparison of Fig. 8b, 8d and 8f shows that almost no adherent QAL10-5-5 materials were transferred to the counter surface of the 42CrMo4-QPQ lower specimen under 8 vol.% FF330 lubrication, and the largest amount of QAL10-5-5 debris accumulated on the counter surface of 42CrMo4-QPQ lower specimen when lubricated with water.

The worn surfaces of the ZCuPb15Sn8/42CrMo4-QPQ tribopairs under oil, water and 8 vol.% FF330 lubrication are given in Figs. 9a to 9f. A comparison of Figs. 9a, 9c and 9e shows that water lubrication leads to the best anti-wear behavior, and the 8 vol.% FF330 lubrication provides the worst anti-wear behavior for the ZCuPb15Sn8/42CrMo4-QPQ tribopairs, which shows a good agreement with the wear-rate results for the ZCuPb15Sn8/42CrMo4-QPQ tribopair. Scratching and adhesion occurred on the tested surface of the ZCuPb15Sn8 upper specimen, as shown in Fig. 9e. Adherent ZCuPb15Sn8 materials were found on the counter surfaces of the lower specimens under oil, water and 8 vol.% FF330 lubrication, as shown in Figs. 9b, 9d and 9f.

3.4. EDS analyses

EDS analyses, as shown in Figure 10, were performed on the worn areas of the 42CrMo4-QPQ lower specimens sliding against HMn62-3-3-0.7, QAL10-5-5, and ZCuPb15Sn8 lower specimens under 8 vol.% FF330 lubrication. The HMn62-3-3-0.7 debris transferred to the 42CrMo4-QPQ lower specimen can be detected in Fig. 10a. The X-ray intensities of Cu and Zn for FF330 lubrication were much lower than those for water lubrication. In addition, Pb particles were detected in the HMn62-3-3-0.7/42CrMo4-QPQ tribopair under water-lubricated conditions. As shown in Fig. 10b, the QAL10-5-5 debris was transferred to the worn areas of the 42CrMo4--QPQ lower specimen, which contained Cu and Al particles. The X-ray intensities of Cu for the water-lubricated condition are the largest, which is consistent with the wear rate results shown in Fig. 6c. As shown in Fig. 10c, the debris in the worn areas contained Cu and Zn particles transferred from the ZCuPb15Sn8 upper specimens for all three lubrications, with Pb particles detected only for the water-lubricated condition. The X-ray intensities of Cu is highest for the 8 vol.% FF330 lubrication, which shows a good agreement with the wear-rate results given in Fig. 6d.

4. Discussion

4.1. Material and surface treatment dependence of tribological behavior of tribopairs

Upon examination of the friction coefficient and wear rate data presented in Section 3, it is evident that the upper specimens of HMn62-3-3-0.7 exhibit consistent steady-state friction coefficients and comparatively modest wear rates across nearly all four lower specimen types tested under the three lubrication conditions. Except for the QAL10-5-5/42CrMo4-WC+Ni tribopair under water lubrication, the spray-welded 15 wt% tungsten carbide (WC)-reinforced Ni-based coatings show the lowest wear rates compared with the other tribopairs with different materials, surface treatments and lubrication. The anti-wear behavior of the 38CrMoAl and 42CrMo4 lower specimens with QPQ coatings were worse than that of the 17-4PH and 42CrMo4-WC+Ni lower specimens under the three lubrication conditions.

4.2. Effect of lubrication on the tribological behavior of tribopairs

The kinematic viscosity of the 8 vol.% FF330 water-based lubrication at the test temperature, i.e. 50°C, is significantly lower than that of the GL-5 85W-90 oil. The steady-state friction coefficients for the three Cu-based upper specimens sliding against the four types of lower specimens under 8 vol.% FF330 lubrication are close to the tribopairs under GL-5 85W-90 oil-lubricated conditions, except for the presence of some fluctuations. Figure 10 shows that the wear rates

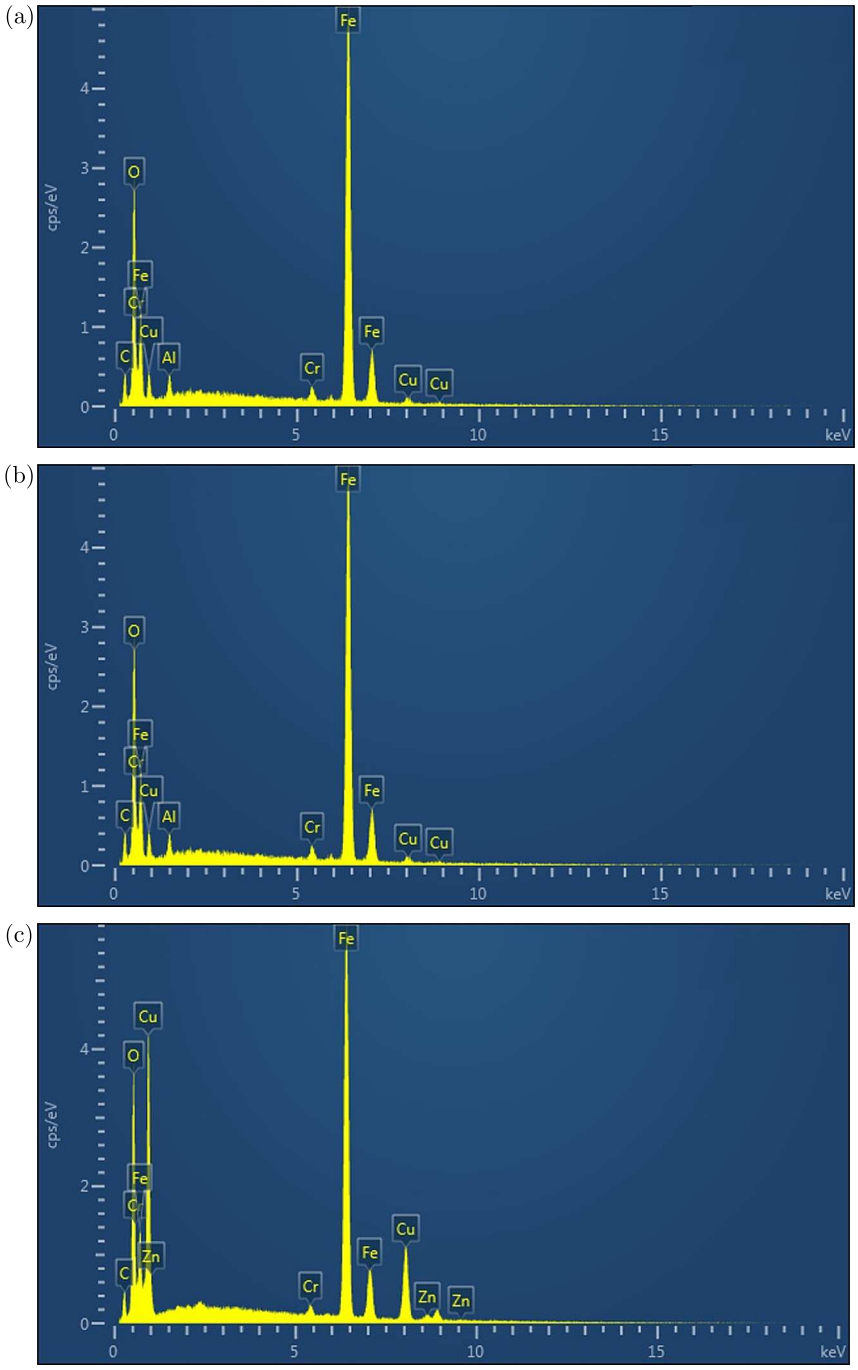


Fig. 10. EDS analyses on worn areas among different tribopairs: (a) 42CrMo4-QPQ against HMn62-3-3-0.7, (b) 42CrMo4-QPQ against QAL10-5-5, (c) 42CrMo4-QPQ against ZCuPb15Sn8

for all tribopairs under 8 vol.% FF330 lubrication are less than 55 mm^3/Nm, much lower than those for the tribopairs under water lubrication. The result of this investigation can be explained by previous studies (Matta *et al.*, 2008; Zhang *et al.*, 2020), which reported that the excellent tribological performance was attributed to the tribochemical transformation of alcohols.

4.3. Effect of initial surface roughness and hardness on tribological behavior of tribopairs

Two surface treatment methods, i.e., the QPQ coating and spray-welded 15 wt% tungsten carbide (WC)-reinforced Ni-based coatings were used to improve the anti-corrosion behavior of representative eccentric cam materials, i.e., 38CrMoAl and 42CrMo4. However, the initial

surface roughness values of the 38CrMoAl and 42CrMo4 lower specimens coated using the quench polishing method were much higher than those of the 42CrMo4 lower specimens coated using the spray welding method. Previous studies have proved that the initial surface roughness values have a significant effect on the "running-in" friction behavior of tribopairs in dry conditions. However, the dominating effect of the initial surface roughness on the friction and wear behavior has not been confirmed under lubricated condition. Dib *et al.* (2020) reported that no large difference in friction coefficients and wear losses could be found for the AISI 304L stainless steel/AISI 52100 alloy steel tribopairs at Ra = 0.034 μm and Ra = 0.72 μm when subjected to oil lubrication. Meanwhile, the hardness values for the QPQ coated lower specimens, in the range of 600 HV to 650 HV, are much lower than those for the 15 wt%WC + Ni coated lower specimens having about 950 HV.Therefore, it is likely that the hardness may have an effect on the tribological behavior of the coated specimens.

5. Conclusions

In this study, wear tests were performed using pin-on-disk specimens with different coatings and lubrication conditions to evaluate the feasibility of applying water-based lubrication for slipper/eccentric cam tribopairs in WHRPPs. The conclusions are as follows:

- The friction-reducing performance of all tribopairs is found to be superior when lubricated with 8 vol.% FF330 water, which is closely consistent with the test results for friction coefficients under oil lubrication. The steady-state friction coefficients, when lubricated with 8 vol% FF330 water, exhibit mean values ranging from 0.17 to 0.23.
- In general, tribopairs under 8 vol.% FF330 water-based lubrication showed lower wear rates than those under oil and water lubrication. The lowest wear rate was obtained, i.e., $1.1 \cdot 10^{-12}$ mm^3/N·m, for the HMn62-3-3-0.7 /17-4PH tribopair under 8 vol.% FF330 water-based lubrication.
- The most severe wear was observed on the worn surface of the QAL9-4-4 specimen under water lubrication. Adhesive wear dominated the wear mechanism of QAL9-4-4 sliding against the 42CrMo4-QPQ disc specimens under water lubrication.

This paper provides a concise overview of the impact of material, surface treatment method, lubrication, initial surface roughness, and hardness on the tribological behavior of tribopairs in WHRPPs. We suggest that the tribochemical mechanism of HMn62-3-3-0.7/WC-reinforced Ni-based coated surfaces lubricated with water-ethylene glycol be accounted for in future studies.

Acknowledgment

This research was supported by CCTEG Project (2023-TD-MS015, 2023-TD-QN004) and TMIC Project (2022TM-167M) funding, China.

References

1. Cotter M., 2000, Rethinking the piston pump – a physicist's convictions, *World Pumps*, **2000**, 401, 32-33
2. Dib J., Hereñu S., Alí D., Pellegri N., 2020, Influence of load, surface finish and lubrication on friction coefficient of AISI 304 stainless steel, *Journal of Materials Engineering and Performance*, **29**, 5, 2739-2747
3. Espinosa T., Jiménez, M., Sanes J., Jiménez A.E., Iglesias M., Bermúdez M.D., 2014, Ultra-low friction with a protic ionic liquid boundary film at the water-lubricated sapphire-stainless steel interface, *Tribology Letters*, **53**, 1-9

4. ISO-14242-2, 2016, Implants for surgery-wear of total hip-joint prostheses – Part 2: Methods of measurement
5. Kou B., Li R., Li Z., Hao R., 2022, Experimental study on the tribological properties of high water-based piston friction pairs with different ceramic materials, *Advances in Mechanical Engineering*, **14**, 8
6. Li R., Wei W., Liu H., Geng X., Cui J., *et al.*, 2024, Design and experimental validation of a valve-distributed water radial piston pump, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, **238**, 12, 5488-5497
7. Li R., Wei W., Liu H., Geng X., Wang D., *et al.*, 2023, Tribological behavior of tribo-pairs in water hydraulic radial piston pumps, *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, **237**, 7, 1603-1623
8. Liang Y., Wang W., Zhang Z., Xing H., Wang C., *et al.*, 2023, Effect of material selection and surface texture on tribological properties of key friction pairs in water hydraulic axial piston pumps: A review, *Lubricants*, **11**, 8, 324
9. Liu S., Yu Z., Yi D., Li Y., 2001, Chemical pretreatments of the surface of WC-15wt%Co with diamond coatings, *Acta Metallurgica Sinica (English Letters)*, **14**, 2, 143-147
10. Ma J., Chen J., Li J., Li Q., Ren C., 2015, Wear analysis of swash plate/slipper pair of axis piston hydraulic pump, *Tribology International*, **90**, 467-472
11. Matta C., Joly-Pottuz L., De Barros Bouchet M.I., Martin J.M., Kano M., *et al.*, 2008, Superlubricity and tribochemistry of polyhydric alcohols, *Physical Review B*, **78**, 8, 085436
12. Nie S., Lou F., Ji H., Yin F., 2019, Tribological performance of CF-PEEK sliding against 17-4PH stainless steel with various cermet coatings for water hydraulic piston pump application, *Coatings*, **9**, 7, 436
13. Nie S.L., Huang G.H., Li Y.P., 2006, Tribological study on hydrostatic slipper bearing with annular orifice damper for water hydraulic axial piston motor, *Tribology International*, **39**, 11, 1342-1354
14. Wei L., Zhang Z., Nie S., Wu X., 2019, Direct and indirect corrosion wear performance of AISI 630 steel for the slipper/swashplate pair in a water hydraulic pump, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, **233**, 10, 1605-1615
15. Wu H., Zhao L., Ni S., He Y., 2020, Study on friction performance and mechanism of slipper pair under different paired materials in high-pressure axial piston pump, *Friction*, **8**, 5, 957-969
16. Wu Z., Zhou F., Wang Q., 2018, Friction and wear properties of CrSiCN/SiC tribopairs in water lubrication, *Journal of Materials Engineering and Performance*, **27**, 6, 2885-2898
17. Yin F., He Z., Nie S., Ji H., Ma Z., 2023, Tribological properties and wear prediction of various ceramic friction pairs under seawater lubrication condition of different medium characteristics using CNN-LSTM method, *Tribology International*, **189**, 108935
18. Yin F., Wang Y., Ji H., Ma Z., Nie S., 2021, Impact of sliding speed on the tribological behaviors of cermet and steel balls sliding against SiC lubricated with seawater, *Tribology Letters*, **69**, 39
19. Yinshui L., Wu D., He X., Zhuangyun L., 2009, Materials screening of matching pairs in a water hydraulic piston pump, *Industrial Lubrication and Tribology*, **61**, 3, 173-178
20. Zhang R., Chen Q., He Z., Xiong L., 2020, In situ friction-induced amorphous carbon or graphene at sliding interfaces: effect of loads, *Applied Surface Science*, **534**, 146990
21. Zhang T., Zhang X., Chen X., Jia Y., 2017, Numerical investigation on the effect of labyrinth seal configuration on leakage, *MATEC Web of Conferences*, EDP Sciences, **128**, 02013
22. Zhang Z., Nie S., Yuan S., Liao W., 2015, Comparative evaluation of tribological characteristics of CF/PEEK and CF/PTFE/graphite filled PEEK sliding against AISI630 steel for seawater hydraulic piston pumps/motors, *Tribology Transactions*, **58**, 6, 1096-1104

23. Zhao S., Guo T., Yu Y., Dong P., Liu C., Chen W., 2017, Design and experimental studies of a novel double-row radial piston pump, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, **231**, 10, 1884-1896

24. Zhou F., Yue B., Wang Q., Wang X., Koshi A., Kato K., 2010, Tribological properties of a-CNx coatings sliding against SiC balls in ethylene glycol aqueous solution, *Lubrication Science*, **22**, 6-7, 225-236

Manuscript received January 11, 2024; accepted for publication October 11, 2024

JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 4, pp. 855-873, Warsaw 2024
https://doi.org/10.15632/jtam-pl/194235

DESIGN METHOD OF DIAMOND BUCKLING RESTRAINED BRACED FRAME STRUCTURE BASED ON NORMAL FORCE BALANCE OF COLUMN JOINTS

Qing Zhang, Ji Yao

Kunming University of Science and Technology, Faculty of Civil Engineering and Mechanics, Kunming, China
e-mail: evanzhangqing@126.com

The diamond buckling restrained braced frame (DBRBF) structure is a novel form of an anti-lateral force support system. A design method for DBRBF structures suitable for various structural types is proposed based on the equilibrium of normal forces in column joints. Taking into account the impact of linear stiffness ratio between upper and lower buckling restrained braced (BRB) elements on the unbalanced forces in column joints of DBRBF sructures, a design method is presented to ensure the normal force balance before and after BRB yield. Building upon this, the calculation method for determining the optimal linear stiffness ratio of two types of BRB elements and the corresponding design method are provided along with the deduced matching conditions for the two types of BRB designs. Furthermore, the relationship between the energy dissipation capacity of BRB elements and deformation of the primary structure under different seismic levels is examined, establishing the condition under which BRB elements serve as the first line of defense prior to yielding of the primary structure. Additionally, the conditions for energy dissipation and ductility guarantee of BRB elements under rare earthquakes are specified. An equation is derived for the length of the working section of BRB elements in DBRBF structures. The results demonstrate that the DBRBF structure can effectively mitigate the adverse effects of BRB elements on the columns through a well-designed approach, enabling realization of matched product designs, and providing a theoretical reference for similar engineering design projects.

Keywords: diamond buckling restrained braced structure, linear stiffness ratio, energy dissipation, design method

1. Introduction

BRB is a type of dissipative element that prevents buckling of the compression steel brace by means of an external restraining mechanism (Yoshino and Karino, 1971; Wakabayashi *et al.*, 1973). It can effectively dissipate energy under both tension and compression. The research has demonstrated that BRB energy dissipation can significantly reduce seismic damage to the primary structure (Di Sarno and Manfredi, 2010). In the structural seismic design, BRBs can enhance the lateral stiffness of the structure, similar to conventional supports, prior to yielding. After yielding, they can undergo plastic deformation to dissipate energy and increase structural damping. Due to its dual functionality of the "bearing" and "energy dissipation", BRB has gained widespread recognition and application once it came out (e.g., Black *et al.*, 2002; Khampanit *et al.*, 2014; among others).

According to the Chinese Seismic Design Code (GB50011-2010, 2010), the primary arrangements for BRB supports include $\wedge$-shaped, $\vee$-shaped, and paired single diagonal bar supports. When the BRB is integrated with the frame structure, the support generates internal force that is transferred through the gusset plate to the beam column joints, potentially causing plastic

hinges. This is particularly evident when the axial force of the support is transferred to concrete members with weak tension, resulting in a more intricate force distribution in the beam column joint area. Altering the connection form or force transfer mode between the gusset plate and the frame is currently a common solution (Benavent-Climent *et al.*, 2015). Furthermore, lateral deformation of the structure leads to compression or tension on the gusset plate due to deformation of the beam and column. This causes stress concentration on the gusset plate, which may lead to premature failure and the subsequent withdrawal of the BRB from operation (Maheri and Ghaffarzadeh, 2008). Simultaneously, the reaction of the gusset plate significantly impacts the beam column joints in an adverse manner. Therefore, several innovative design methods have been proposed to mitigate the negative effects of gusset plates on frame members (Berman and Bruneau, 2009; Lin *et al.*, 2014).

To address the issue of adverse interaction between the brace and structural members, Qu *et al.* (2013, 2015) proposed a zigzag arrangement of BRBs in a reinforced concrete frame, with the removal of reinforced concrete beams in the supporting span. Under lateral deformation, the horizontal force components of adjacent BRBs axial forces counteract each other, effectively mitigating or even eliminating the horizontal force components at the nodes. Additionally, Qu *et al.* (2017) and Xie (2016) introduced a double K-shaped support layout (Fig. 1) as an alternative solution. In this configuration, any two adjacent BRBs always exert opposing forces. At each node, one BRB experiences tension while the other undergoes compression. The resulting normal forces at the node offset each other, thereby reducing the normal unbalanced force and preventing stress concentration in the node plate. Moreover, the BRB node is positioned away from potential "plastic hinges" at the ends of reinforced concrete beams and columns. This ensures that the plastic behavior of the main structural members does not adversely interact with the BRB joints, thus avoiding any detrimental impact on the gusset plates.

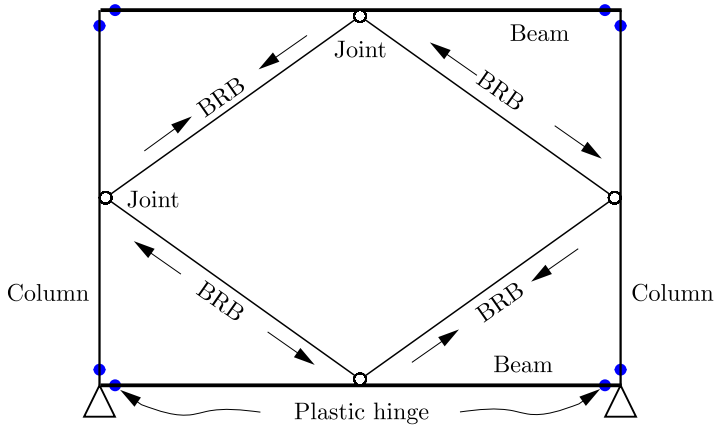


Fig. 1. Reinforced concrete frames braced by BRBs in double-K configuration

In ANSI/AISC 341-02 (2002), a K-braced frame is defined as an OCBF (ordinary concentrically braced frame) with a pair of diagonal braces connected to a single point within the clear height of the column on one side. The definition in ANSI/AISC 341-10 (2010) evolves to describe a braced frame arrangement where the brace is connected to a column with no out of plane brace. Based on these definitions, it is evident that the support structure depicted in Fig. 1 falls under the category of K-bracing. However, the Chinese Seismic Design Code (GB50011-2010, 2010) and ANSI/AISC341-10 (2010) do not recommend the use of K-bracing arrangements for supports. This recommendation also applies to the layout of BRBs due to the adverse effect of the normal unbalanced force on the column caused by the support at the node, which accelerates column failure (Sabelli *et al.*, 2013). In the elastic stage of ordinary steel supports, the normal stress at the beam and column nodes can be offset if the tension and compression members

produce the same strain and stress. In this case, the support only generates an additional axial force on the beam and column. However, as the transverse load and horizontal displacement increase, the compression member will buckle and cease to function. At this point, the brace will induce additional axial force and bending moment on the beam and column, resulting in detrimental effects on the seismic performance of the main structural members.

However, in Japan, there is always the possibility of connecting braces within the uninterrupted span of the column. For instance, Japan proposed the employment of a minimal-intrusion arm damper, which establishes a connection between the midspan of the beam and the upper section of the column. When the damper is activated, it generates a normal shear force on the column, and the magnitude of the detrimental shear force on the column relies on the normal component force generated at the node during energy dissipation of the damper (Kurata *et al.*, 2016). Numerous scholars have conducted valuable investigations on similar matters as well. For instance, Fan *et al.* (2021) examined the rhombic configuration of conventional steel braced steel frame structures. In comparison to the configuration of $\wedge$-shaped and singly inclined bars, they discovered that the rhombic arrangement enhanced redundancy and ductility of the structure. Taking into consideration the adverse effects of unbalanced support forces on the frame beams and columns, they proposed a performance-based seismic design approach for rhombic grid braced frame structures. Qu *et al.* (2017) and (Xie, 2016) conducted an experimental study on the damage mechanism, seismic performance, and mechanical performance of joints of RC frames with double K shaped BRBs. This arrangement effectively enhances the lateral stiffness of the structure and mitigates the shear force on the connection interface of concrete members caused by the braces.

The DBRBF structural system is a novel type of lateral force-resistant BRB supporting structure. It employs a configuration where four BRBs are arranged in a single-story and single-span layout, with their axes combined to form a diamond shape. This arrangement effectively keeps the BRB nodes away from regions with high bending moments, thereby reducing the adverse effects on structural members. Additionally, it enhances the redundancy and ductility of the structure (Fig. 1). Geometrically, each pair of BRBs and columns forms a K-shaped node. However, unlike the K-shaped layout, the diamond-shaped layout has a fixed node position at the midpoint of the column. Consequently, the axial force distribution of the pull rod and the compression rod cannot be adjusted by changing the node position. Nevertheless, the diamond-shaped layout facilitates the achievement of reasonable included angles between the BRBs, beams, and columns, thus providing convenience during construction.

On the other hand, the performance of BRB support differs from that of conventional steel support. BRB has the ability to bear yield due to its ability to continue functioning even after yielding, regardless of tension and compression. As the structure can be seen as a vertical cantilever beam with shear and bending deformation, the structural displacement is a combination of shear and bending deformation. The upper part of the structure is deformed as Δ_1, while the lower part is deformed as Δ_2. Whether the BRB layout is double K or diamond, Δ_1 is not equal to Δ_2 (Fig. 2). Therefore, based on the deformation coordination, the tensile and compressive deformations of the two upper $\wedge$-shaped BRBs are the same, denoted as δ_1, and the two lower $\vee$-shaped BRBs have the same deformation, denoted as δ_2, but δ_1 is not equal to δ_2. In other words, the deformation of BRB is symmetrical from the left to right and from the top to bottom, resulting in a complex stress state of the column node and an unbalanced force of the column node. Thus, according to the Chinese Seismic Design Code (GB50011-2010, 2010) for steel structures, the normal unbalanced force caused by brace buckling should not be considered for frame beams connected with $\wedge$-shaped and $\vee$-shaped BRB braces, and frame columns should not have K-shaped joints. Therefore, in order to eliminate the unbalanced force of column joints caused by bracing in the DBRBF structural system and ensure safety of

columns, solving the problem of column safety is the first and key issue in such a structural design.

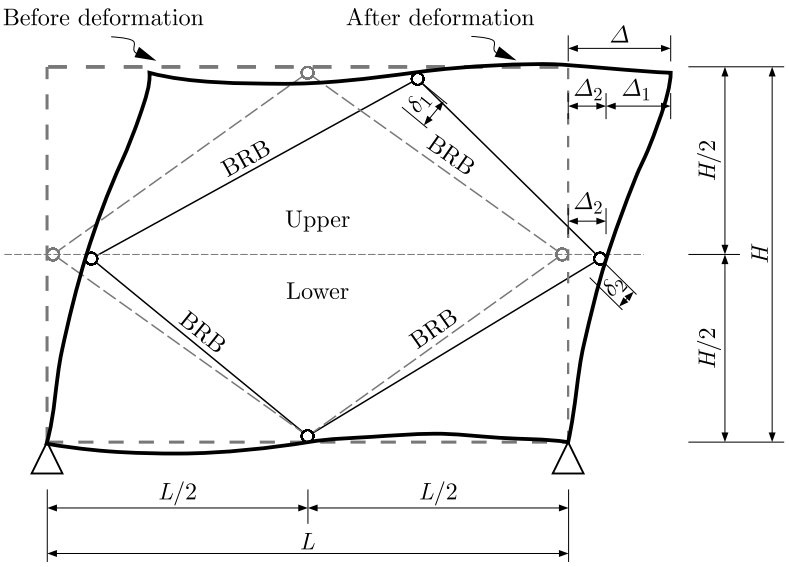


Fig. 2. Deformation of DBRBF structure depicted in a schematic manner

In this paper, the theoretical establishment of the design method for DBRBF is presented, focusing on the geometric configuration characteristics. The adverse effects of the normal unbalanced force of the node on the column are addressed, and the design method for matching BRB products with the structural performance target needs is discussed. The constraints on a BRB product design under different levels of a seismic action are explored, leading to valuable conclusions that facilitate the design of DBRBF. These findings hold significant implications for similar engineering applications.

2. Effect of the BRB linear stiffness ratio on the unbalanced force of DBRBF column joints

Based on the inherent static characteristics of the DBRBF structural system, it exhibits both symmetry and antisymmetry (Fig. 3). From the aforementioned analysis, it can be observed that if the linear stiffness of the upper two BRB_1 braces is equal, and the linear stiffness of the lower two BRB_2 braces is equal, the normal force at the beam node caused by the BRB under horizontal seismic force will naturally equilibrate, effectively nullifying the normal forces at the beam node and exerting no contribution to the column axial force. However, this is not the case for the column joints. If the linear stiffness of BRB_1 and BRB_2 is equal, the axial forces of BRB_1 and BRB_2 will differ in the elastic stage due to the uneven deformation between the upper and lower portions of the structure, resulting in an imbalance of the normal forces at the column joints. It has been proven that conventional design methods, which assume equal BRB linear stiffness in the same layer and span, lead to the same axial internal force values F_1 for the upper two BRB_1 braces and the same axial internal force values F_2 for the lower two BRB_2 braces, but $F_1 \neq F_2$, indicating that the normal force caused by the BRB at the column node cannot be balanced. Consequently, the conventional design method fails to eliminate the adverse effects of unbalanced forces caused by the BRB at the node on the column. Therefore, finding a rational design method that ensures column safety is crucial in the design of the DBRBF structural system.

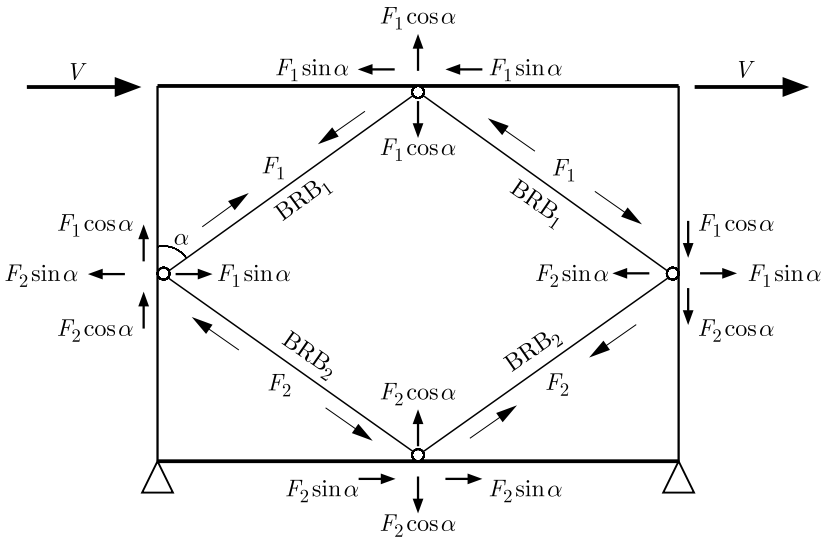


Fig. 3. Equilibrium of internal forces structural model of DBRBF

This paper aims to utilize the approach of adjusting the linear stiffness ratio of BRB_1 and BRB_2 in order to achieve equal axial internal forces among the four BRBs in the elastic stage. The objective is to eliminate the unbalanced force at the column joint prior to BRB yielding. Let K1 represent the linear stiffness of BRB_1 and K2 represent the linear stiffness of BRB_2 and define the linear stiffness ratio K_i as K_1/K_2. To analyze the impact of changes in the linear stiffness ratio on the unbalanced force at column joints caused by BRB internal forces, a frame structure example is examined.

The structure is a single span reinforced concrete frame (DBRBF) shown in Fig. 2. The height of each floor is H=3600mm, with column sections measuring 400 mm×400 mm. The beam span is $L = 4200$ mm, with a section of 200mm×400 mm and C30 concrete. The linear stiffness of BRB_1, denoted as K_1, varies according to different stiffness ratios. On the other hand, the linear stiffness of BRB_2, denoted as K_2, is fixed. The section of BRB_2 is fixed at 30 mm×30 mm, with a vertical force of $V = 200$ kN. By conducting a static analysis of the structure, the variations of BRB_1 and BRB_2 axial internal forces with respect to the linear stiffness ratio K_i are obtained and shown in Fig. 4.

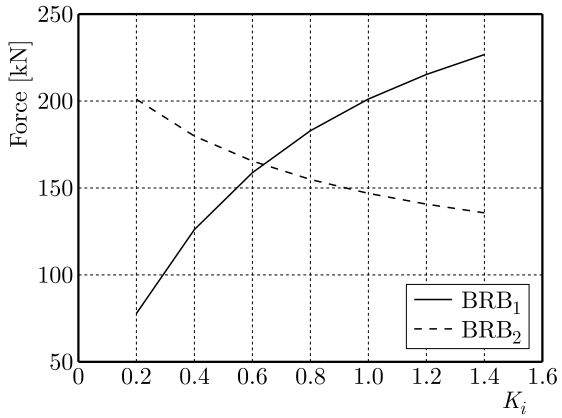


Fig. 4. A diagram of the relationship between the value of the BRB axial force and the linear stiffness ratio K_i

The analysis findings indicate that in the case of a DBRBF structural system, the selection of structural type design (such as storey height, span, section of structural members, material

properties, etc.) greatly influences the internal forces of BRB_1 and BRB_2 when the linear stiffness ratio K_i is changed. There exists a specific linear stiffness ratio (referred to as the optimal linear stiffness ratio K_r) which ensures that the axial forces of BRB_1 and BRB_2 are equal. It has been demonstrated that when $K_i = K_r$, the normal forces at the column nodes cancel each other out, achieving a balance. The larger the deviation between the actual linear stiffness ratio K_i of BRB_1 and BRB_2 and the optimal linear stiffness ratio K_r, the greater the difference in the axial internal forces of BRB_1 and BRB_2, and the more significant the unbalanced force at the column joints. If BRB_1 and BRB_2 have the same linear stiffness ($K_i = 1$), there may be a considerable difference in the axial internal forces between them, and the unbalanced force caused by BRB_1 and BRB_2 at the column joints cannot be neglected.

3. Design method for the normal force balance of column joints

The working performance of BRB can be divided into two stages: the elastic stage and the yield stage. In the elastic stage, when the linear stiffness ratio is K_r, four BRBs exhibit the same bearing capacity F ($F = F_1 = F_2$). The normal forces at the beam and column joints caused by BRB are in equilibrium. Determining the linear stiffness ratio K_r at this stage is a challenging task for engineers. As the lateral deformation of the structure increases, BRBs transition into the yield stage, this is the second stage. Since the linear stiffness of BRB_1 and BRB_2 may differ in the first stage, ensuring that all four BRBs have the same yield bearing capacity and yield simultaneously is also a problem faced by engineers. The next step involves analyzing the two-stage design method of the DBRBF structure separately.

3.1. Normal force balance condition of the column node before BRB yielding

The preceding analysis demonstrates that the DBRBF structural system, with its distinctive geometric characteristics of BRB arrangement and structural deformation properties, can effectively mitigate the unbalanced forces at column joints by adjusting the stiffness ratio of BRB_1 and BRB_2. Consequently, the equilibrium condition of normal forces at the column node during the initial stage can be expressed as

$$K_1 = K_r K_2 \tag{3.1}$$

At its current stage, the BRB is in the elastic phase and exhibits similar working performance to regular supports. The single degree of freedom system of the DBRBF is analyzed using the force method for static analysis. Due to symmetry and antisymmetry of the structure, it is sufficient to analyze only half of the structure, as depicted in Fig. 5. The BRB axes have length of LB, and BRB_1 and BRB_2 have axial tension and compression stiffness values of EA_1 and EA_2, respectively. The linear stiffness values are $K_1 = EA_1/L_B$ and $K_2 = EA_2/L_B$ for BRB_1 and BRB_2, respectively. The bending stiffness of the beam and column are E_1I_1 and E_2I_2, respectively. The beam has a span of L, the structural storey height is H, and α represents the included angle between the BRB and the column. Under the horizontal force V, let X_1 and X_2 be the axial forces generated by BRB_1 and BRB_2, respectively, and X_3 be the reaction force at the beam node. The mechanical equation can then be expressed as

$$\begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} \Delta_{1p} \\ \Delta_{2p} \\ \Delta_{3p} \end{bmatrix} = \mathbf{0} \tag{3.2}$$

One can get Eqs. (3.3) and (3.4) from Fig. 5, according to the Mohr integration method. Due to the relatively small influence of the axial force on deformation of components, in order to

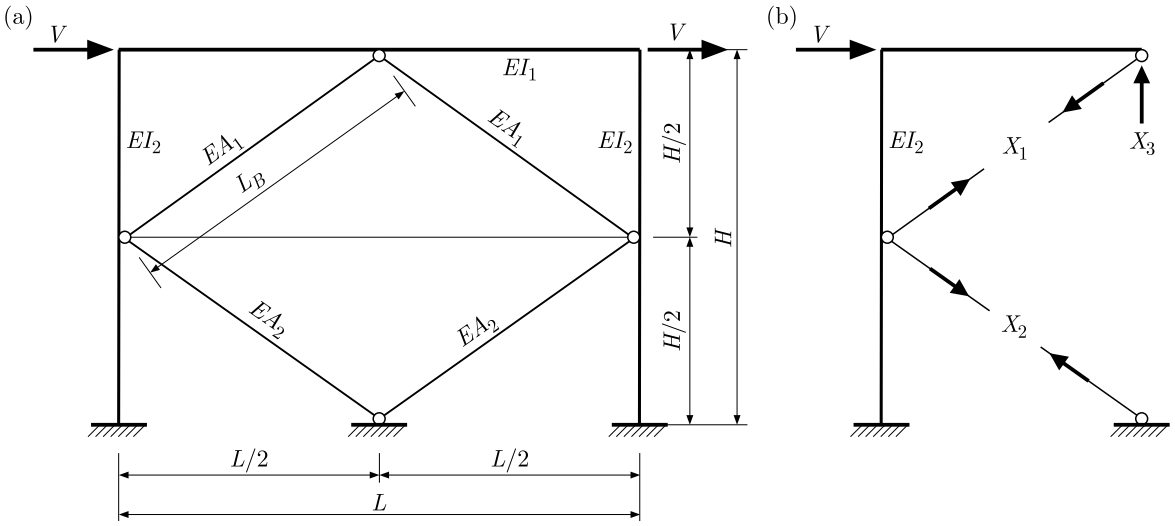


Fig. 5. Structural calculation parameter model of DBRBF: (a) parameter model of the structure, (b) semi structural mechanical model

facilitate calculation, the influence of the axial force on beam column deformation was ignored when deriving the formula

$$\begin{aligned}
&\delta_{11} = \frac{H^3 \sin^2\alpha}{24E_2I_2} + \frac{LH^2\sin^2\alpha}{24E_1I_2} + \frac{\sqrt{L^2+H^2}}{2EA_1} \qquad \delta_{22} = \frac{H^3\sin^2\alpha}{24E_2I_2} + \frac{\sqrt{L^2+H^2}}{2EA_2} \\
&\delta_{33} = \frac{L^3}{24E_1I_1} + \frac{2L^2H}{24E_2I_2} \qquad \delta_{12} = \delta_{21} = 0 \\
&\delta_{13} = \delta_{31} = -\frac{L^2H\sin\alpha}{24E_1I_1} - \frac{LH^2\sin\alpha}{16E_2I_2} \qquad \delta_{23} = \delta_{32} = -\frac{LH^2\sin\alpha}{16E_2I_2}
\end{aligned} \tag{3.3}$$

and

$$\Delta_{1p} = \frac{VH^3\sin\alpha}{48E_2I_2} \qquad \Delta_{2p} = \frac{5VH^3\sin\alpha}{48E_2I_2} \qquad \Delta_{3p} = -\frac{VLH^2}{4E_2I_2} \tag{3.4}$$

Solve Eq. (3.2) to get

$$\begin{aligned}
X_1 &= -\frac{\Delta_{1p}\delta_{23}^2 - \Delta_{2p}\delta_{13}\delta_{23} + \Delta_{3p}\delta_{13}\delta_{22} - \Delta_{1p}\delta_{22}\delta_{33}}{\delta_{22}\delta_{13}^2 + \delta_{11}\delta_{23}^2 - \delta_{11}\delta_{22}\delta_{33}} \\
X_2 &= -\frac{\Delta_{2p}\delta_{13}^2 - \Delta_{1p}\delta_{13}\delta_{23} + \Delta_{3p}\delta_{13}\delta_{11} - \Delta_{2p}\delta_{11}\delta_{33}}{\delta_{22}\delta_{13}^2 + \delta_{11}\delta_{23}^2 - \delta_{11}\delta_{22}\delta_{33}} \\
X_3 &= -\frac{\Delta_{1p}\delta_{13}\delta_{22} + \Delta_{2p}\delta_{11}\delta_{23} - \Delta_{3p}\delta_{11}\delta_{22}}{\delta_{22}\delta_{13}^2 + \delta_{11}\delta_{23}^2 - \delta_{11}\delta_{22}\delta_{33}}
\end{aligned} \tag{3.5}$$

To counteract the imbalanced forces at the joints, it is necessary to determine the internal forces $X_1 = X_2$ for BRB_1 and BRB_2, which can be obtained from Eqs. (3.5)

$$\begin{aligned}
&\Delta_{1p}\delta_{23}^2 - \Delta_{2p}\delta_{13}\delta_{23} + \Delta_{3p}\delta_{13}\delta_{22} - \Delta_{1p}\delta_{22}\delta_{33} \\
&\quad = -(\Delta_{2p}\delta_{13}^2 - \Delta_{1p}\delta_{13}\delta_{23} + \Delta_{3p}\delta_{13}\delta_{11} - \Delta_{2p}\delta_{11}\delta_{33})
\end{aligned} \tag{3.6}$$

By substituting Eqs. (3.2) and (3.4) into Eq. (3.6), we can derive

$$\frac{1}{EA_1}\Big(\frac{5\tan\alpha}{3E_1I_1} + \frac{4}{E_2I_2}\Big) - \frac{1}{EA_2}\Big(\frac{11\tan\alpha}{3E_1I_1} + \frac{4}{E_2I_2}\Big) = 0 \tag{3.7}$$

Hence, the optimal linear stiffness ratio for the single degree of freedom systems BRB_1 and BRB_2 can be expressed as

$$K_r = \frac{EA_1/L_B}{EA_2/L_B} = \frac{5E_2I_2\tan\alpha + 12E_1I_1}{11E_2I_2\tan\alpha + 12E_1I_1} \tag{3.8}$$

If the DBRBF structural system employs a steel or reinforced concrete beam and a column with the same grade of concrete, the simplified equation can be denoted as

$$K_r = \frac{K_1}{K_2} = \frac{5I_2\tan\alpha + 12I_1}{11I_2\tan\alpha + 12I_1} \tag{3.9}$$

In Eq. (3.9), $\tan\alpha$ represents the ratio of the structural span to the storey height, denoted as L/H. Equation (3.8) demonstrates that the optimal linear stiffness ratio of the DBRBF structural system BRB_1 and BRB_2 is determined by the inherent properties of the structure, relying on its unique geometric characteristics and the bending stiffness of its structural members. This ratio is unaffected by the horizontal force and is solely influenced by the three parameters: E_1I_1, E_2I_2, and α. This characteristic facilitates the simplified calculation of the DBRBF structural system analysis. Once the structural type selection is determined for a project, the optimal stiffness ratio of BRB_1 and BRB_2 in DBRBF structures remains a fixed value. Under predefined structural layout conditions, the bending stiffness of the beams and columns becomes the decisive factor for modifying the optimal linear stiffness ratio K_r. This feature can be utilized to develop programs that calculate the optimal linear stiffness ratio for multi-degree of freedom systems and continuous beams.

The recommended range of the included angle α between the BRBs and the column, as suggested by The Code for Seismic Design of Buildings (GB50011-2010, 2010), is 35°-55°. However, in this study, the included angle α is appropriately increased to 22.5°-67.5° ($\pi/8$ $3\pi/8$). From Fig. 6, it can be observed that for the DBRBF structure, under the same bending stiffness ratio of the beam and column, the optimal linear stiffness ratio K_r decreases as the included angle α increases. Conversely, when the geometric characteristics of the structural layout are

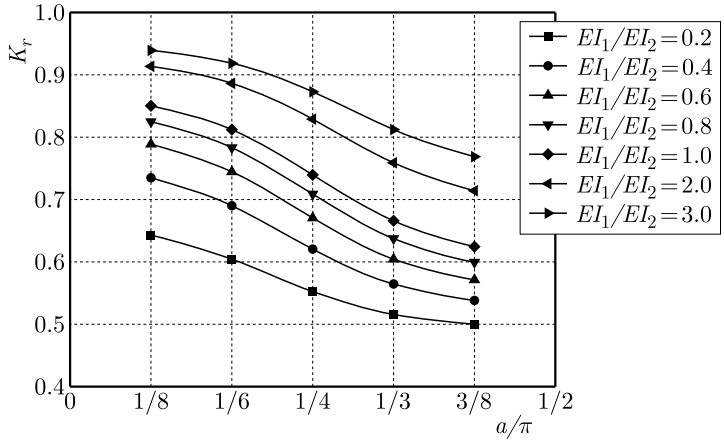


Fig. 6. Optimal linear stiffness ratio curve for different angles α

determined, with the same included angle α, the optimal linear stiffness ratio K_r increases with an increase in the bending stiffness ratio of the beam to the column. In the engineering design process, the structural system can be adjusted comprehensively based on the aforementioned change rules. Considering the recommended range of the included angle α in the specification and the requirement for strong columns and weak beams, the common range of the optimal linear stiffness ratio K_r in actual projects is 0.55-0.75. It has been demonstrated that although the axial force values of the four BRBs can be equalized by adjusting the stiffness ratio, the stiffness

of BRB_1 and BRB_2 are not equal due to the inherent properties of the DBRBF structure. Hence, in addition to the usual equivalent design based on calculation parameters, the design of BRB products also needs to consider matching the performance of BRB_1 and BRB_2 products. This ensures that BRB_1 and BRB_2 can have the same yield bearing capacity and meet structural deformation characteristics under different stiffness conditions.

3.2. Normal force balance conditions of column joints after BRB yielding

Due to inadequate capability of mainstream design software in China (e.g. YJK, PKPM) to accurately simulate the mechanical behavior of BRBs and node domains, the design software typically substitutes their mechanical behavior with two-force bars and achieves product design through parameter equivalence. The key parameters of the product mainly encompass axial stiffness, yield bearing capacity, yield displacement, and post-yield stiffness. Based on the previous analysis, it is observed that, under the optimal linear stiffness ratio K_r, BRB_1 and BRB_2 exhibit the same bearing capacity but different linear stiffness in the first stage of the DBRBF structure. Consequently, the separate and arbitrary equivalence design of BRB_1 and BRB_2 products is not viable. This approach may lead to mismatched mechanical behavior, such as yield bearing capacity and yield displacement between BRB_1 and BRB_2 in the second stage, resulting in unbalanced forces at the column joint. Therefore, in order to ensure that BRB_1 and BRB_2 possess equivalent bearing capacity after yielding and yield simultaneously, it is imperative to conduct performance matching design for major equivalent parameters, including the yield force and yield displacement of BRB_1 and BRB_2 products.

The primary focus of this paper is to examine the design of performance matching for BRB_1 and BRB_2 products. To simplify the analysis, the mechanical behavior of the BRB is divided into three parts: the energy dissipation section and two connection sections. The energy dissipation section has a length of L_e and a section area of A_e. The connection section has a length of L_j and an equivalent section area of A_j, as illustrated in Fig. 7. The connection section comprises various elements such as concrete components, gusset plates, core material elastic sections, and transition sections. It has been demonstrated that the actual length, stiffness, and cross-sectional area of BRB products differ from those of the two-force bar, but they exhibit equivalent performance during operation.

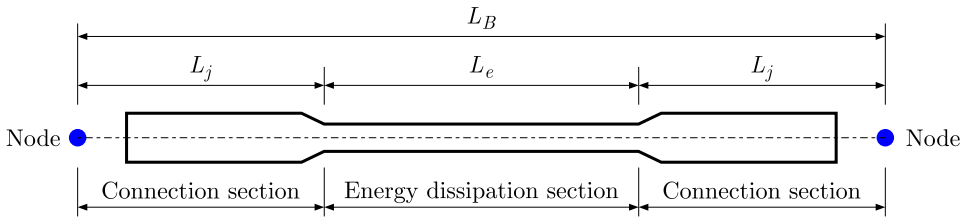


Fig. 7. BRB axial series composition

LY100, LY160, LY195, LY225, Q235, Q345, Q390, etc. are commonly utilized materials in the energy dissipation section of BRB (Buckling Restrained Brace) in Chinese projects. When designing BRB products, designers have some flexibility in selecting different steel yield strengths to determine the appropriate section area for the energy dissipation section and the individual areas of each part in the connecting section. Subsequently, they can adjust the lengths of the energy dissipation section and each part of the connecting section to achieve equivalence with the design parameters of a two-bar system. This ensures the mechanical behavior equivalence. Therefore, the design of a single BRB product results in various BRB products with different core materials that are equivalent to the same two-bar parameters (Wu *et al.*, 2021).

The results depicted in Fig. 7 exhibit that the linear stiffness of BRB products can be represented as

$$\frac{1}{K} = \frac{1}{K_e} + \frac{2}{K_j} \tag{3.10}$$

In Eq. (3.10), K represents the equivalent linear stiffness of the axis as calculated by BRB. K_e and K_j, on the other hand, refer to the linear stiffness of the energy dissipation section and the equivalent stiffness of the connection section, respectively

$$K_{ie} = \frac{EA_{ie}}{L_{ie}} \qquad K_{ij} = \frac{EA_{ij}}{L_{ij}} \qquad i = 1, 2 \tag{3.11}$$

By substituting Eq. (3.11) and $K = EA/L_B$ into Eq. (3.10), we can derive the ratio of the core section area to the core section length for both BRB_1 and BRB_2

$$\frac{A_{1e}}{L_{1e}} = \frac{1}{L_B - 2L_{1j}\frac{A_1}{A_{1j}}} A_1 \qquad \frac{A_{2e}}{L_{2e}} = \frac{1}{(L_B - 2L_{2j}\frac{A_2}{A_{2j}}} A_2 \tag{3.12}$$

The BRB yield bearing capacity F_y is

$$F_y = \sigma_y A_e = \eta f_y A_e \tag{3.13}$$

The parameter η represents the coefficient of super strength in the core energy dissipation section of steel, which is incorporated to account for the Bauschinger effect (as shown in Table 1).

Table 1. The Bauschinger coefficient

Steel	LY100	LY160	LY195	LY225	Q235	Q345	Q390
η	1.1	1.1	1.15	1.1	1.25	1.1	1.05

When both BRB_1 and BRB_2 exhibit an equal yield strength capacity F_y, then

$$F_y = \eta_1 f_{1y} A_{1e} = \eta_2 f_{2y} A_{2e} \tag{3.14}$$

By substituting Eq. (3.14) into Eqs. (3.12), the equilibrium condition of the normal force in the column node can be derived when the BRB reaches its yield point simultaneously

$$\frac{L_{1e}}{L_{2e}} = \frac{\beta}{K_r} \frac{L_B - 2L_{1j}\frac{A_1}{A_{1j}}}{L_B - 2L_{2j}\frac{A_2}{A_{2j}}} \tag{3.15}$$

Let β represent the yield strength ratio of the energy dissipation section steel of BRB_1 and BRB_2

$$\beta = \frac{\eta_2 f_{2y}}{\eta_1 f_{1y}} \tag{3.16}$$

Utilizing the steel strength data provided in Table 1, the yield strength ratio β of the energy dissipation section can be calculated using Eq. (3.16). The range of β has been demonstrated to be between 0.27 and 3.72. By selecting different types of steel to adjust the linear stiffness of both the core material energy dissipation section and the connecting section, it is possible to ensure that the yield forces of BRB_1 and BRB_2 are equal. Moreover, there is ample flexibility in terms of the matching space when different linear stiffness values are employed.

Equation (3.15) demonstrates that the design of the energy dissipation section of the core material in BRB_1 and BRB_2 products is mutually constrained under the double constraints of

maintaining the same optimal stiffness ratio in the first stage and the same yield force in the second stage. The length and linear stiffness of the energy dissipation section in either BRB_1 or BRB_2 product must correspond to the length and linear stiffness of the energy dissipation section in the other product. Additionally, this matching relationship is influenced by the material properties and the linear stiffness of the connection section selected for both products. When this criterion is satisfied, BRB_1 and BRB_2 will have the same yield force and can yield simultaneously, thereby achieving a balanced normal force at the column node after BRB yielding. In the project, the design of BRB_1 can be carried out initially based on the fundamental performance parameters of both BRB_1 and BRB_2, followed by the design of BRB_2 using the basic performance parameters of BRB_2 and Eq. (3.15). Conversely, BRB_2 can be designed first, and then BRB_1 can be designed according to Eq. (3.15).

4. The BRB products are reasonably matched with the performance objectives of the main structure

In the field of structural design, the storey displacement angle is a crucial performance indicator for a structure. Its limit value is dependent on the type of structure and the performance level of the structure. For structures of the same type, the inter-story displacement angle limit is a fixed value. When compared to commonly used layout forms such as a single diagonal brace, $\wedge$-shaped brace, and $\vee$-shaped brace, the diamond layout form BRB exhibits the minimum axial deformation. During an earthquake, BRB functions as an energy dissipation member and should enter the yielding energy dissipation state before the main structure, serving as the primary line of defense against a seismic activity. Additionally, BRB must possess sufficient ductility safety reserves, allowing it to operate without damage during rare earthquakes. In engineering design, the performance parameters of BRB products are established based on the performance objectives of the structure. The performance level at which BRB enters the yield energy dissipation state is determined by the compatibility between the product performance parameters and the performance objectives of the structure. This relationship directly impacts the rationality of the design results. This paper proposes a method for rational matching of product performance parameters and structure performance objectives, in accordance with the requirements of BRB product performance parameters for main structure performance objectives.

4.1. BRB as the control condition of the first seismic fortification line of the structure

The calculation diagram for lateral deformation of the DBRBF structural system when subjected to horizontal forces is shown in Fig. 8.

In Fig. 8, Δ represents the horizontal displacement between floors in the frame structure, while Δ_1 and Δ_2, respectively, denote the horizontal inter-story displacements of the upper and lower halves of the structure. θ represents the angular displacement of the structure between floors. When the structure experiences lateral deformation, the axial deformations δ_1 and δ_2 of BRB_1 and BRB_2 can be calculated using

$$\delta_1 = \frac{F}{EA_1} L_B \qquad \delta_2 = \frac{F}{EA_2} L_B \tag{4.1}$$

By taking into account the constraint of the optimal linear stiffness ratio, we can derive

$$\delta_1 = \frac{1}{K_r} \delta_2 \tag{4.2}$$

The horizontal displacement between the top and bottom of the column Δ can be defined as the difference in the horizontal displacement between the upper half layer Δ_1 and the lower half layer Δ_2. As θ is a small angle, so $\tan\theta \approx \theta$

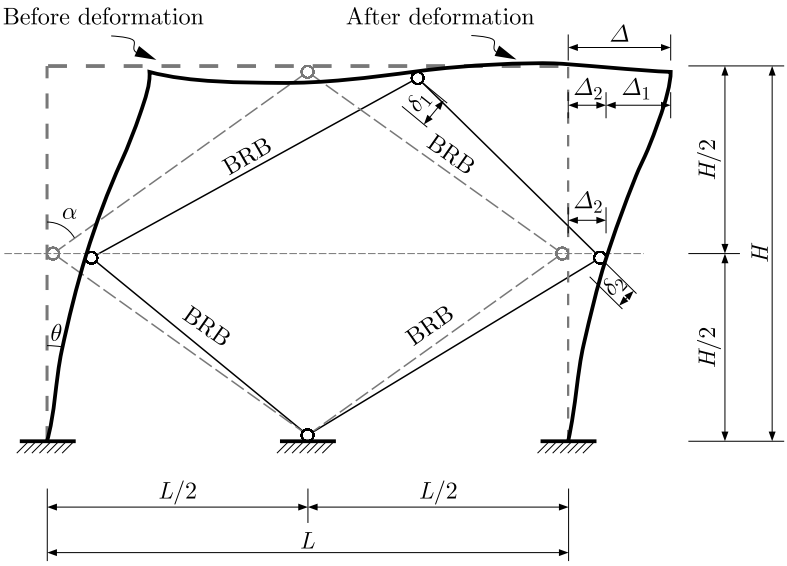


Fig. 8. Schematic of DBRBF structure deformation calculation

$$\Delta = \Delta_1 + \Delta_2 = \frac{\delta_1}{\sin\alpha} + \frac{\delta_2}{\sin\alpha} = H\tan\theta = H\theta \tag{4.3}$$

By substituting $L = H/2\cos\alpha$, Eq. (3.11), Eq. (4.1), and Eq. (4.2) into Eq. (4.3), we can derive

$$\theta = \frac{K_r + 1}{\sin 2\alpha}\frac{A_{1e}}{A_1}\varepsilon_{1e} = \frac{K_r + 1}{K_r \sin 2\alpha}\frac{A_{2e}}{A_2}\varepsilon_{2e} \tag{4.4}$$

Let ε_{1e}, ε_{2e} be the strain of the core material energy dissipation sections of BRB_1 and BRB_2, respectively. Upon yielding of BRB, the strain $\vee$ of the energy dissipation section is obtained, and the corresponding interlayer displacement angle of the DBRBF structure when BRB yields is given by

$$\theta_{By} = \frac{K_r + 1}{\sin 2\alpha}\frac{A_{1e}}{A_1}\varepsilon_{1ey} = \frac{K_r + 1}{K_r \sin 2\alpha}\frac{A_{2e}}{A_2}\varepsilon_{2ey} \tag{4.5}$$

To activate the energy dissipation of BRB before the main frame in the DBRBF structure, it is required that the inter-story displacement angle θ_{By} is smaller than the inter-story displacement angle θ_y when the structure reaches its yield point

$$\theta_{By} < \theta_y \tag{4.6}$$

4.2. BRB matching design under different performance levels

According to literature (Li *et al.*, 2010), BRB can be categorized into three types in engineering applications based on different stages of energy dissipation: damper type, energy dissipation type, and bearing type. The damper type BRB functions as a damper, dissipating energy during frequent earthquakes. The energy dissipating BRB maintains elasticity and only offers stiffness under frequent earthquakes, but yields to dissipate energy during fortification earthquakes or rare earthquakes. The load-bearing BRB remains elastic at all levels and solely provides stiffness. It has been demonstrated that the level at which BRB enters the energy dissipation state significantly impacts the structural design. Therefore, the selection of material yield strength and adjustment of the energy dissipation section length L_e of BRB to align with the performance target of the structure under varying levels of the seismic action is a crucial aspect in the design of the DBRBF structural system. Currently, the availability of steel grades for BRB products is

limited, making it important to determine the appropriate material yield strength and L_e length for achieving the desired working performance target.

As depicted in Fig. 7 above, the axial deformation of the Buckling-Restrained Brace (BRB) is characterized by deformation of the energy dissipation section δ_e and deformation of the connecting section δ_j

$$\delta = \delta_e + \delta_j = \varepsilon_e L_e + 2\varepsilon_j L_j \tag{4.7}$$

Let ε_e and ε_j denote the strain of the energy dissipation section and the connecting section, respectively. A previous research (Cai *et al.*, 2005) indicates that the ratio of the strain in the connecting section to the total axial strain of the BRB is less than 6%. This is primarily due to the requirement in BRB product design that the connecting section should remain elastic even when the yield section of the BRB experiences strain strengthening or failure caused by earthquakes at different levels. Consequently, the deformation of the connecting section can be disregarded, and the strain in the energy dissipation section can be approximately considered as the axial strain

$$\delta_1 = \varepsilon_{1e} L_{1e} \qquad \delta_2 = \varepsilon_{2e} L_{2e} \tag{4.8}$$

The determination of the BRB deformation and structural displacement angle θ in any state can be achieved by substituting Eqs. (4.8) and (4.2) into Eq. (4.3)

$$\theta = \frac{(1+K_r)\varepsilon_{1e} L_{1e}}{H \sin\alpha} = \frac{(1+K_r)\varepsilon_{2e} L_{2e}}{K_r H \sin\alpha} \tag{4.9}$$

Definition λ represents the ratio of length of the energy dissipation section of the BRB to overall length of its axis, denoted as $\lambda = L_e/L_B$. It is evident that λ is less than 1. Eq. (4.10) is derived from Eq. (4.9)

$$\theta = \frac{(1+K_r)\varepsilon_{1e} \lambda_1}{\sin 2\alpha} = \frac{(1+K_r)\varepsilon_{2e} \lambda_2}{K_r \sin 2\alpha} \tag{4.10}$$

The response of BRB_1 and BRB_2 subjected to the specific horizontal seismic loading is represented by

$$\varepsilon_{1e} = \frac{\theta_i \sin 2\alpha}{(1+K_r)\lambda_1} \qquad \varepsilon_{2e} = \frac{\theta_i K_r \sin 2\alpha}{(1+K_r)\lambda_2} \tag{4.11}$$

θ_i $(i = 1, 2, 3)$, representing frequent earthquake, fortification earthquake, and rare earthquake, respectively, refers to the angular displacement of the structural storey caused by earthquakes of varying magnitudes. Equation (4.12) describes the conditions for yield energy dissipation of BRB_1 and BRB_2 when subjected to a specific level of earthquake

$$\varepsilon_{1e} = \frac{\theta_i \sin 2\alpha}{(1+K_r)\lambda_1} \geqslant \varepsilon_{1ey} = \frac{\sigma_{1y}}{E} \qquad \varepsilon_{2e} = \frac{\theta_i K_r \sin 2\alpha}{(1+K_r)\lambda_2} \leqslant \varepsilon_{2ey} = \frac{\sigma_{2y}}{E} \tag{4.12}$$

The lengths of the energy dissipation sections of BRB_1 and BRB_2 should satisfy the following conditions when BRB yield energy is consumed at various levels of seismic activity

$$\lambda_1 \leqslant [\lambda_{1i}] = \frac{\theta_i E \sin 2\alpha}{(1+K_r)\sigma_{1y}} \qquad \lambda_2 \leqslant [\lambda_{2i}] = \frac{K_r \theta_i E \sin 2\alpha}{(1+K_r)\sigma_{2y}} \tag{4.13}$$

where $[\lambda_{1i}]$ and $[\lambda_{2i}]$ represent the permissible values for the respective lengths of the energy dissipation sections under the influence of BRB_1 and BRB_2 earthquakes at different intensities. They also denote the maximum lengths of the energy dissipation sections when BRB_1 and BRB_2

reach the state of energy consumption during the occurrence of earthquakes of corresponding intensities.

From Eq. (4.12), it has been demonstrated that an increase in the permissible value of the energy dissipation section length leads to enhanced yield energy dissipation of the BRB under the influence of earthquakes at corresponding levels. Conversely, a decrease in the permissible value of the energy dissipation section length results in shorter energy dissipation section length, which may lead to inadequate ductility reserve of the BRB during rare earthquake events, making it highly susceptible to reaching the limit deformation and experiencing fracture failure. It is also evident that the permissible percentage of BRB energy dissipation section length $[\lambda_i]$ and the material yield strength σ_y are dependent on the performance level target established for the structure. Moreover, the optimal stiffness ratio K_r and the BRB are influenced by the column angle α. Based on the above analysis, it can be concluded that the optimal linear stiffness ratio K_r is associated with the bending stiffness E_1I_1, E_2I_2 and the included angle α of the beam and column making it analogous to the allowable proportion of the energy dissipation section $[\lambda_i]$. This value is solely determined by the four parameters σ_y, E_1I_1, E_2I_2 and α.

In the design of BRB, the proportion of energy dissipation section length should be as large as possible. Too short energy dissipation section length will limit the energy consuming capacity of BRB on the one hand, and deformation of the energy dissipation section will be too large under a rare earthquake on the other hand. When the axial strain of BRB energy dissipation section is less than 3%, BRB can maintain stable mechanical behavior. When the axial strain exceeds 3%, the amplitude of friction on the compression side increases rapidly, and the compression side of the hysteresis curve is prone to instability and buckling failure (Iwata and Murai, 2006). Therefore, the reference (T/CECS817-2021, 2021) stipulates that the length of the energy dissipation section shall not be less than 60% of the total length of the BRB, and the axial strain of the energy dissipation section under the design displacement shall not exceed 3%. As the limit value of structural displacement angle under the action of a fortification earthquake is not clearly specified in the Chinese Code for Seismic Design of Buildings (GB50011--2010, 2010), designers often pay more attention to structural deformation characteristics under frequent and rare earthquakes. Taking the reinforced concrete frame structure as an example, the yield energy dissipation of BRB_1 and BRB_2 under frequent earthquakes and their ductility requirements under rare earthquakes are discussed below.

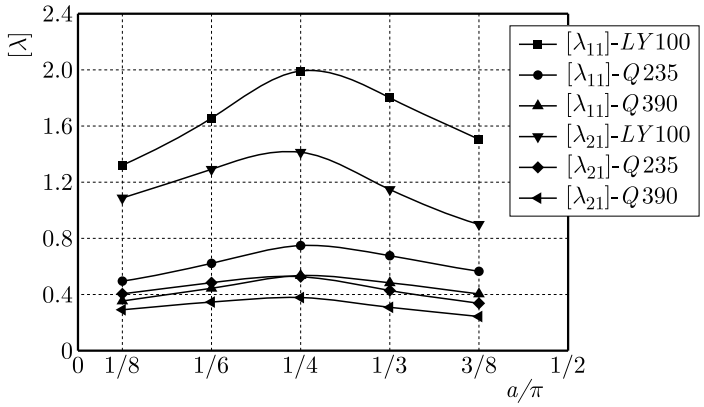


Fig. 9. Curve of $[\lambda]$ for different angles α

In Fig. 9, the fixed $E_1I_1/E_2I_2 = 1.0$ DBRBF concrete frame structure is shown under frequent earthquakes ($\theta 1 = 1/550$). The allowable values for the energy dissipation section percentages of phase BRB_1 and BRB_2, denoted as $[\lambda_{11}]$ and $[\lambda_{21}]$, are displayed in the figure. It can be observed that when other conditions are held constant and the included angle between the BRB and the column is $\alpha = \pi/4$, both BRB_1 and BRB_2 are most likely to experience yield

energy dissipation during frequent earthquakes. Under the same included angle condition, BRB_1 exhibits a higher likelihood of yielding and dissipating energy compared to BRB_2 when utilizing steel with the same yield point in the energy dissipation section.

Figure 10 illustrates the permissible values of length ratios $[\lambda_{11}]$ and $[\lambda_{21}]$ for the energy dissipation sections of BRB_1 and BRB_2 in frequent earthquakes at the optimal angle of $\alpha = \pi/4$. As λ is always less than 1, it can be inferred that the energy consumption requirements of frequent earthquakes are naturally fulfilled when $\lambda > 1$. It is evident from Fig. 10 that only LY100 steel satisfies the yield energy consumption performance target for BRB_1 and BRB_2 under frequent earthquakes. Furthermore, the figure indicates that in the design of BRB_1 and BRB_2 products, choosing steel with a lower yield point for the energy dissipation section facilitates easier yielding and energy consumption in frequent earthquakes. Conversely, higher yield point steel poses challenges in achieving the BRB energy consumption target in frequent earthquakes, with potential difficulty in meeting the requirements stated in document (T/CECS817-2021, 2021). As a result, low yield point steel should be preferred in the product design. Additionally, under the same yield strength of steel, a higher bending stiffness ratio E_1I_1/E_2I_2 between the beam and column leads to a greater proportion of energy dissipation section length for BRB_1 and a smaller limit value of $[\lambda_{11}]$. Consequently, this unfavorably impacts energy consumption of BRB_1 in frequent earthquakes. Conversely, BRB_2 exhibits the opposite trend. Thus, when selecting the structure, a comprehensive consideration of these factors is necessary.

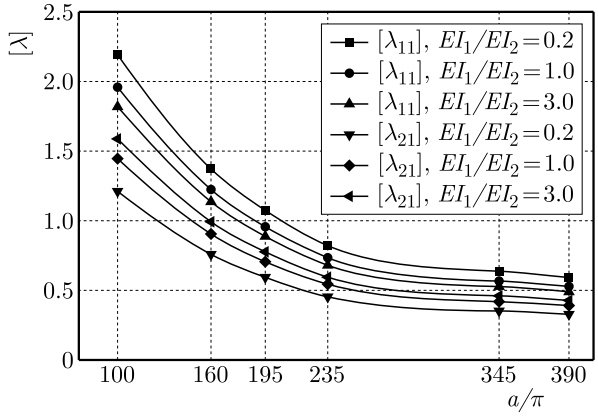


Fig. 10. Curve of $[\lambda]$ for different yield strengths of steel

The analysis above is based on the performance target of frequent earthquakes. Likewise, when determining the design for the DBRBF structure type selection under the influence of fortification earthquakes and rare earthquakes, the allowable values $[\lambda_{1i}]$ and $[\lambda_{2i}]$ for the proportion of the BRB_1 and BRB_2 energy dissipation sections lengths can be obtained by substituting the displacement angle limit value under the corresponding level earthquake into Eq. (4.13). This allows for selection of appropriate steel to achieve the desired yield energy consumption during the fortification earthquake or rare earthquake stage. It has been demonstrated that the DBRBF structural system initiates the yield energy dissipation of the BRB at a specific performance target level. The design of the energy dissipation section is crucial. By considering the structural performance target, geometric characteristics of the structure, and the bending stiffness of the beam and column, the selection of suitable steel should be made to ensure the desired performance of the BRB_1 and BRB_2 products.

4.3. BRB ductility guarantee conditions

The BRB should have a sufficient safety reserve of ductility during rare earthquakes, and also provide enough ductility allowance under the maximum limit of structural elastoplastic

deformation. The previous analysis has shown that length of the energy dissipation section in a diamond-shaped layout for BRBs is shorter compared to the traditional layout. However, a too short energy dissipation section during rare earthquakes can result in a insufficient safety reserve for ductility and potential fracture failure. According to FEMA450 regulation (FEMA450, 2000), the design deformation of BRB products should be determined as 1.5 times the maximum interlayer design deformation. In this study, the axial strain, Eq. (4.14) of the energy dissipation section should not exceed 3% to ensure the ductility of the BRB, as stated in Eq. (4.11)

$$\varepsilon_{1e} = \frac{1.5\theta_3 \sin 2\alpha}{(1+K_r)\lambda_1} \leqslant 3\% \qquad \varepsilon_{2e} = \frac{1.5\theta_3 K_r \sin 2\alpha}{(1+K_r)\lambda_2} \leqslant 3\% \tag{4.14}$$

Further

$$\lambda_1 \geqslant [\xi_1] = \frac{50\theta_3 \sin 2\alpha}{1+K_r} \qquad \lambda_2 \leqslant [\xi_2] = \frac{50\theta_3 K_r \sin 2\alpha}{1+K_r} \tag{4.15}$$

According to Eq. (4.15), the minimum limit proportion of energy dissipation sections BRB_1 and BRB_2, denoted as $[\xi_1]$ and $[\xi_2]$, respectively, is required to meet the ductility requirements in rare earthquakes. The formula shows that the minimum value of the energy dissipation section proportion is only influenced by three parameters E_1I_1, E_2I_2 and α. This dependency is due to K_r being related solely to these three parameters, and not affected by the yield strength of steel used in the energy dissipation section. This differs from the starting condition of BRB yield energy dissipation. The main reason is that the axial strain of the energy dissipation section is set as 3% under rare earthquake conditions. Essentially, the minimum limit of the length proportion of the energy dissipation section is determined by the type of structure, with the minimum length of the BRB energy dissipation section meeting the ductility requirement under rare earthquake action being fixed.

The change curve of the minimum limit $[\xi_1]$ and $[\xi_2]$ of the proportion of energy dissipation section length of BRB_1 and BRB_2 in the DBRBF reinforced concrete frame structure is shown in Fig. 11 for rare earthquakes $\theta_3 = 1/50$. It can be observed from Fig. 11 that when the included angle between the BRB and column is $\alpha = \pi/4$, the proportion of BRB_1 or BRB_2 to the energy dissipation section under a rare earthquake action has the most stringent minimum limit λ. In terms of the proportion of energy dissipation section length under the same included angle,

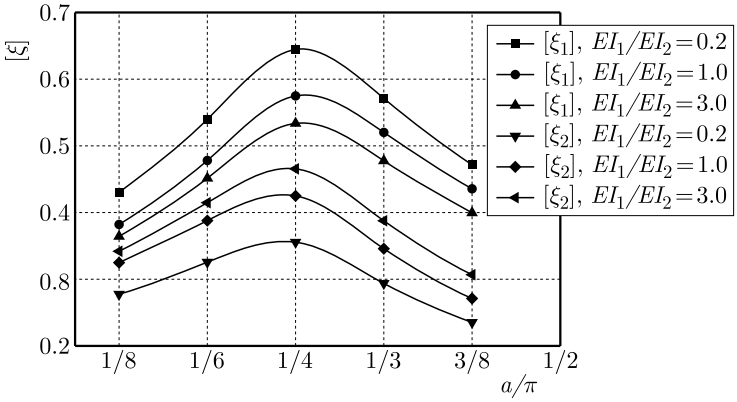


Fig. 11. Curve of $[\xi]$ for different angles α

BRB_1 has higher requirements than BRB_2. Comparing the proportion of energy dissipation section in frequent earthquakes, it can be seen from the maximum limit of λ that the proportion of BRB energy dissipation section under frequent earthquakes has a greater allowable value, indicating a stricter minimum length limit of the energy dissipation section length under the rare earthquake action. This is mainly because the BRB enters the yield energy dissipation state

earlier under rare earthquakes, leading to greater plastic deformation and requiring longer length of the energy dissipation section to ensure ductility.

The previous analysis demonstrates that BRB_1 and BRB_2 products should be designed to both absorb and dissipate energy during corresponding-level earthquakes. They should also ensure that no fractures occur during rare earthquakes, while maintaining a certain level of ductility safety reserve. Therefore, based on the aforementioned analysis, it can be concluded that Eq. (4.16) determines the range of percentage for the energy dissipation sections of BRB_1 and BRB_2 in DBRBF structures

$$[\xi_1] \leqslant \lambda_1 \leqslant [\lambda_1] \qquad [\xi_2] \leqslant \lambda_1 \leqslant [\lambda_1] \tag{4.16}$$

After comparing the calculation formulas on each side of the inequality, it has been demonstrated that the left-hand side of Eqs. (4.16) is a constant value for the given structural type, while the right-hand side is influenced by factors such as material properties, structural selection, and performance objectives. Among these factors, the strength of the core material has a significant impact. As the strength of the core material increases, the calculated value on the right-hand side of Eqs. (4.16) gradually decreases, sometimes even becoming smaller than the left-hand side. As shown in Figs. 10 and 11, BRB starts to consume energy during frequent earthquakes. In the case of DBRBF reinforced frame structures, when LY100 steel is used as the core material, Eqs. (4.16) can be satisfied under any conditions. However, when high-strength steel Q390 is utilized, neither BRB_1 nor BRB_2 can meet the requirements of Eqs. (4.16). This is primarily due to the limitations imposed by λ in completing the product design, making it impossible to achieve energy dissipation during frequent earthquakes while ensuring the ductility safety reserve demand of BRB under rare earthquakes. Thus, it is evident that Eqs. (4.16) is not universally applicable, indicating that BRB_1 and BRB_2 can only fulfill either the energy dissipation condition or the ductility guarantee condition. The interplay between the performance of BRB_1 and BRB_2 leads to a more complex process of core material matching and selection during product design.

The project adopts the DBRBF structural design method, which involves selecting the appropriate structure and calculating the optimal stiffness ratio of BRB_1 and BRB_2. After completing the structural analysis, the performance parameters of BRB_1 and BRB_2 products are obtained. The product design is then carried out based on the requirements of these performance parameters. When designing BRB_1 and BRB_2 products, several conditions need to be considered, including product stiffness matching, post yield strength, deformation matching, as well as startup energy consumption and ductility guarantee conditions, which act as the primary control factors. If the BRB product design fails to meet the performance target requirements, it is necessary to re-optimize the structural design.

The above design methods can be applied to various types of structures, including steel structures, reinforced concrete frame structures with seismic walls, slab column seismic walls, frame core tubes, tubes in tubes, and other structural forms.

5. Conclusion

By adjusting the stiffness ratio of BRB_1 and BRB_2, the adverse effects of brace arrangement on columns can be eliminated, thus achieving a normal force balance of column joints in the elastic stage in the design of the DBRBF structure.

In the design process of BRB_1 and BRB_2 products, achieving a normal force balance at any level stage of column joints relies on parameter matching between the two products, in addition to the conventional design method of BRB product parameter equivalence.

Compared to the traditional BRB layout, the diamond layout exhibits reduced axial deformation and strain in the energy dissipation section of the BRB. Therefore, it is not suitable for

initiating energy consumption. In a determined structural system, the utilization of a lower steel yield point in the energy dissipation section enables the BRB to achieve yield energy dissipation at an earlier stage, thereby serving as a more effective first line of defense before the main frame yields.

The relationship between length of the BRB energy dissipation section and structure type, structure selection, and material properties determine the maximum proportion. By controlling the length proportion of the energy dissipation section during product design, the desired yield energy consumption can be achieved under different levels of earthquakes. Additionally, the design should meet the ductility safety reserve requirements during rare earthquakes.

References

1. ANSI/AISC 341-02, 2002, *Seismic Provisions for Structural Steel Buildings*, American Institute of Steel Construction, Inc., Chicago, IL
2. ANSI/AISC 341-10, 2010, *Seismic Provisions for Structural Steel Buildings*, American Institute of Steel Construction, Inc., Chicago, IL
3. Benavent-Climent A., Oliver-Saiz E., Donaire-Avila J., 2015, New connection between reinforced concrete building frames and concentric braces: Shaking table tests, *Engineering Structures*, **96**, 7-21
4. Berman J.W., Bruneau M., 2009, Cyclic testing of a buckling restrained braced frame with unconstrained gusset connections, *Journal of Structural Engineering*, **135**, 12, 1499-1510
5. Black C., Makris N., Aiken I., 2002, Component testing, stability analysis and characterization of buckling-restrained unbonded braces, *PEER Report 2002/08*, Pacific Earthquake Engineering Research Center, University of California, Berkeley
6. Cai K.-Q., Huang Y.-Z., Wong C.-X., 2005, Seismic performance and applications of double-tube buckling-restrained braces (in Chinese), *Progress in Steel Building Structures*, **7**, 3, 1-8
7. Di Sarno L., Manfredi G., 2010, Seismic retrofitting with buckling restrained braces: Application to an existing non-ductile RC framed building, *Soil Dynamics and Earthquake Engineering*, **30**, 11, 1279-1297
8. Fan C., Bao X., Hao J., Zhang H., Zhong W., Zhang K., 2021, Performance-based aseismic design method for diamond grid braced frame structure (in Chinese), *Journal of Vibration and Shock*, **40**, 23, 143-151
9. FEMA356, 2000, FEMA, *Prestandard and Commmentary for the Seismic Rehabilitation of Buildings*, Federal Emergency Management Agency, Washington, D.C.
10. GB50011-2010, 2010, *Code for Seismic Design of Buildings*, Ministry of Housing and Urban-Rural Development of the People's Republic of China, Beijing
11. Iwata M., Murai M., 2006, Buckling-restrained brace using steel mortar planks; performance evaluation as a hysteretic damper, *Earthquake Engineering and Structural Dynamics*, **35**, 14, 1807-1826
12. Khampanit A., Leelataviwat S., Kochanin J., Warnitchai P., 2014, Energy-based seismic strengthening design of non-ductile reinforced concrete frames using buckling-restrained braces, *Engineering Structures*, **81**, 110-122
13. Kurata M., Sato M., Zhang L., Lavan O., Becker T., Nakashima M., 2016, Minimal-disturbance seismic rehabilitation of steel moment-resisting frames using light-weight steel elements, *Earthquake Engineering and Structural Dynamics*, **45**, 3, 383-400
14. Li G., Sun F., Zhang Y., 2010, Application classification standard and performance standard of buckling restrained brace (in Chinese), *Journal of Civil, Architectural and Environmental Engineering*, **33**, 2, 391-396

15. Lin P.C., Tsai K.C., Wu A.C., Chuang M.C., 2014, Seismic design and test of gusset connections for buckling-restrained braced frames, *Earthquake Engineering and Structural Dynamics*, **43**, 4, 565-587

16. Maheri M.R., Ghaffarzadeh H., 2008, Connection overstrength in steel-braced RC frames, *Engineering Structures*, **30**, 7, 1938-1948

17. Qu Z., Kishiki S., Maida Y., Sakata H., Wada A., 2015, Seismic responses of reinforced concrete frames with buckling restrained braces in zigzag configuration, *Engineering Structures*, **105**, 12-21

18. Qu Z., Kishiki S., Sakata H., Wada A., Maida Y., 2013, Subassemblage cyclic loading test of RC frame with buckling restrained braces in zigzag configuration, *Earthquake Engineering and Structural Dynamics*, **42**, 7, 1087-1102

19. Qu Z., Xie J., Wang T., Kishiki S., 2017, Cyclic loading test of double K-braced reinforced concrete frame subassemblies with buckling restrained braces, *Engineering Structures*, **139**, 1-14

20. Sabelli R., Roeder C.W., Hajjar J.F., 2013, Seismic design of steel special concentrically braced frame systems: A guide for practising engineers, *NEHRP Seismic Design Technical Brief*, **8**, National Institute of Standards and Technology, USA

21. T/CECS 817-2021, 2021, *Technical Specification for Application of Buckling Restrained Brace*, China Association for Engineering Construction Standardization, Beijing

22. Wakabayashi M., Nakamura T., Katagihara A., 1973, Experimental study on the elasto-plastic behavior of braces enclosed by precast concrete panels under horizontal cyclic loading, *Technical Papers of Annual Meeting*, Architectural Institute of Japan, Tokyo, 1041-1044

23. Wu K., Tao Z., Yu W., Lan X., Zhang L., 2021, Discussion on the parameters matching of a buckling restrained brace in structural design (in Chinese), *Journal of Vibration and Shock*, **40**, 14, 117-124

24. Xie J., 2016, *Seismic Performance Research of Double K-braced Reinforced Concrete Frame with Buckling Restrained Braces*, Master's Thesis, Institute of Engineering Mechanics, China Earthquake Administration, Harbin

25. Yoshino T., Karino Y., 1971, Experimental study on shear wall with braces, Part 2, *Technical Papers of Annual Meeting*, Architectural Institute of Japan, Tokyo, 403-404

Manuscript received February 1, 2024; accepted for publication October 7, 2024

INFORMATION FOR AUTHORS

Journal of Theoretical and Applied Mechanics (***JTAM***) is devoted to all aspects of solid mechanics, fluid mechanics, thermodynamics and applied problems of structural mechanics, mechatronics, biomechanics and robotics. Both theoretical and experimental papers as well as survey papers can be proposed.

JTAM accepts full-text articles as well as short communications.

We accept articles in English only. The contents of a paper submitted to *JTAM* should not exceed 12 pages (short communication: 4 pages) of standard A4 format (11-point type size, standard margins – 2.5 cm, single line spacing) including abstract, figures, tables and references.

The material for publication should be sent to the Editorial Office via electronic journal system: http://www.editorialsystem.com/jtam

Papers are accepted for publication after the review process. Blind review model is applied, which means that the reviewers' names are kept confidential to the authors. Reviewer(s) declare that there is no interpersonal relation with the author(s) that would affect the opinion and recommendation of the article for publication in *JTAM*. The final decision on paper acceptance belongs to the Editorial Board.

Publication of a paper in the Journal is subject to fee. The fee is EUR 700 for a full length paper and EUR 350 for a short communication. The fee is due upon acceptance of the paper for publication by the Editorial Board.

After qualifying your paper for publication we will require LaTeX or TeX or Word document file and figures. The best preferred form of figures are files obtained by making use of editorial environments employing vector graphics.

Requirements for paper preparation

Contents of the manuscripts should appear in the following order:

- Title of the paper.
- Authors' full name, affiliation and e-mail.
- Short abstract (maximum 100 words) and 3-5 key words (1 line).
- Article text (equations should be numbered separately in each section; each reference should be cited in the text by the last name(s) of the author(s) and the publication year).
- References (maximum 25) in alphabetical order.
- Titles of references originally published not in English, should be translated into English.

All the data should be reported in SI units.

Contents