

РАЗРАБОТКА ЦИФРОВОГО ДВОЙНИКА ТРЕХПЛУНЖЕРНОГО ГИДРАВЛИЧЕСКОГО НАСОСА

М. А. ЖУРАВКОВ¹⁾, М. А. НИКОЛАЙЧИК¹⁾, И. Н. КАРПОВИЧ¹⁾,
А. В. ПОЛЯГОШКО¹⁾, Ю ГУАНБИН²⁾, МА ВАНЬШЭН³⁾

¹⁾Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь

²⁾Харбинский политехнический университет, ул. Вест Дажу, 92, 150001, г. Харбин, Китай

³⁾Чунцинские насосные системы, ул. Цзинишэн роуд, 8, 400033, г. Чунцин, Китай

Аннотация. Представлен процесс разработки цифрового двойника триплексного плунжерного насоса. В его основу положена математическая модель, построенная на физических принципах работы рассматриваемого устройства. Модель учитывает рабочие параметры и взаимодействие компонентов внутри насоса. Для обучения системы прогнозирования использованы данные о трех ключевых типах неисправностей. Рассматриваемыми отказами являются утечки, закупорки входного канала и износ подшипника, а также их комбинации. Цифровой двойник позволяет выявлять характерные признаки зарождающихся отказов на ранней стадии и находить конкретную причину.

Ключевые слова: цифровой двойник; плунжерный гидравлический насос; моделирование неисправностей; прогнозная модель.

Образец цитирования:

Журавков МА, Николайчик МА, Карпович ИН, Полягошко АВ, Ю Гуанбин, Ма Ваньшэн. Разработка цифрового двойника трехплунжерного гидравлического насоса. *Журнал Белорусского государственного университета. Математика. Информатика.* 2026;2:62–76 (на англ.).
EDN: KLETWO

For citation:

Zhuravkov MA, Nikolaitchik MA, Karpovich IN, Poliahoshko AA, Yu Guangbin, Ma Wansheng. Development of digital twin of a three-plunger hydraulic pump. *Journal of the Belarusian State University. Mathematics and Informatics.* 2026; 2:62–76.
EDN: KLETWO

Авторы:

Михаил Анатольевич Журавков – доктор физико-математических наук, профессор; заведующий кафедрой теоретической и прикладной механики механико-математического факультета.

Михаил Александрович Николайчик – кандидат физико-математических наук, доцент; ведущий научный сотрудник кафедры теоретической и прикладной механики механико-математического факультета.

Иван Николаевич Карпович – младший научный сотрудник кафедры теоретической и прикладной механики механико-математического факультета.

Александр Владимирович Полягошко – студент механико-математического факультета. Научный руководитель – М. А. Николайчик.

Ю Гуанбин – доктор физико-математических наук; профессор школы мехатроники.

Ма Ваньшэн – руководитель исследовательского отдела.

Authors:

Michael A. Zhuravkov, doctor of science (physics and mathematics), full professor; head of the department of theoretical and applied mechanics, faculty of mechanics and mathematics. zhuravkov@bsu.by

<https://orcid.org/0000-0002-7420-5821>

Mikhail A. Nikolaitchik, PhD (physics and mathematics), docent; leading researcher at the department of theoretical and applied mechanics, faculty of mechanics and mathematics.

nikolaitchikma@bsu.by

<https://orcid.org/0000-0003-3733-1615>

Ivan N. Karpovich, junior researcher at the department of theoretical and applied mechanics, faculty of mechanics and mathematics.

karpovichin@bsu.by

<https://orcid.org/0009-0003-5056-3880>

Alexander A. Poliahoshko, student at the faculty of mechanics and mathematics.

polsaige@gmail.com

<https://orcid.org/0009-0006-3917-7075>

Yu Guangbin, doctor of science (physics and mathematics); professor at the school of mechatronics engineering.

johnnyzhyq@gmail.com

<https://orcid.org/0009-0007-8063-0051>

Ma Wansheng, head of the research department.

wanshengma@gmail.com

<https://orcid.org/0009-0009-8682-0475>

DEVELOPMENT OF DIGITAL TWIN OF A THREE-PLUNGER HYDRAULIC PUMP

*M. A. ZHURAVKOV^a, M. A. NIKOLAITCHIK^a, I. N. KARPOVICH^a,
A. A. POLIAHOSHKO^a, Yu GUANGBIN^b, MA WANSHENG^c*

^aBelarusian State University, 4 Niezaliezhnasci Avenue, Minsk 220030, Belarus

^bHarbin Institute of Technology, 92 West Dazhi Street, Harbin 150001, China

^cChongqing Pump, 8 Jingsheng Road, Chongqing 400033, China

Corresponding author: I. N. Karpovich (karpovichin@bsu.by)

Abstract. This paper presents the development and modelling of a digital twin of a triplex plunger pump. Its mathematical model is based on the physical process. The model takes into account the operating parameters and interaction of components inside the pump. Data on three key types of malfunctions are used to train the prediction system. The faults considered are leaks, inlet channel blockages, and bearing wear, as well as combinations thereof. The digital twin allows the characteristic signs of incipient failures to be identified at an early stage and the specific cause to be found.

Keywords: digital twin; plunger hydraulic pump; fault simulation; predictive model.

Introduction

Digital twin technology, a key enabler of digital transformation and data-driven decision-making across industries, has undergone substantial evolution over the past two decades. In the current era of digital transformation, digital twins have emerged as an essential technology, offering virtual counterparts of physical systems, processes, or products throughout their entire life cycle.

A digital twin is defined as a comprehensive set of virtual information constructs that fully describe a potential or actual physical object. Provided it is developed with sufficient fidelity, a digital twin can yield the same insights as direct examination of the manufactured physical product¹.

This technology is particularly valuable for modelling complex systems that interact with their environment in multiple ways, making long-term performance difficult to forecast. Examples include large-scale mining equipment and aircraft engines. When a physical asset is deployed, its digital counterpart can monitor and evaluate wear, fatigue, or specific stress patterns. The information generated can inform future design improvements and maintenance strategies [1].

Modelling and simulation are fundamental to digital twin development. They produce virtual representations of physical systems, enabling analysis, optimisation, and informed decision-making [2]. Techniques such as multiphysics modelling, finite element analysis, and computational fluid dynamics are commonly employed. Digital twin models generally fall into three categories: physics-based, data-driven, and hybrid models [1; 3].

Physics-based models use mathematical principles and physical laws to describe system behaviour and are widely applied in engineering disciplines like structural and fluid dynamics analysis. In contrast, data-driven models leverage machine learning and statistical methods to predict behaviour by analysing historical and real-time data. Hybrid models integrate both approaches, marrying the accuracy of physics-based modelling with the adaptability of data-driven techniques.

A core aspect of the digital twin concept is simulation that facilitates real-time, bidirectional interaction between the virtual model and the physical object. Unlike the conventional modelling, this approach incorporates real-time data gathered from the physical system via Internet of Things (IoT) technology. Parameters in the physical world can be estimated using image recognition and laser measurement, while control is managed through electrical, programmable, embedded, and networked systems [4].

Data forms the cornerstone of any digital twin. Implementing a digital twin requires the selection and integration of various data acquisition tools, including sensors, RFID tags and readers, cameras, and scanners. This data must be transmitted in real or near-real time. The significant volume, velocity, and variety of this data present challenges for the cloud transmission. Edge computing offers an effective solution by pre-processing data locally, reducing network load, and mitigating data leakage risks. Simultaneously, 5G technology facilitates an efficient real-time data transfer. Data mapping and fusion technologies, such as the commonly cited XML, are also crucial for interpreting the collected information.

¹Allen B. D. Digital twins and living models at NASA [Electronic resource]. URL: https://ntrs.nasa.gov/api/citations/20210023699/downloads/ASME%20Digital%20Twin%20Summit%20Keynote_final.pdf (date of access: 27.08.2025).

The IoT plays a pivotal role in advancing digital twins by supplying real-time sensor data from physical assets. IoT devices collect diverse parameters like temperature, pressure, vibration, and humidity, which are then transmitted to the digital twin platform for analysis [5]. This continuous data integration allows digital twins to deliver precise and timely insights into system performance.

As an example of industrial application, one can refer to the mining enterprise digital twin components previously developed by the authors [6; 7].

This study focuses on the development process of a digital twin for a three-plunger hydraulic pump.

Method modelling

A plunger pump is classified as a volume pump, where the working chamber's volume changes periodically due to the reciprocating motion of the plunger within the hydraulic cylinder. In terms of structure, the working chamber is isolated from the external environment by a sealing mechanism. During operation, the mechanical energy from the piston pump is converted directly into the pressure energy of the liquid [8; 9].

Figure 1 shows the crank OA rotating with uniform angular velocity of ω , when the crank rotating angle $\varphi = 0 \sim \pi$, is the suction stroke, the crank rotating angle $\varphi = \pi \sim 2\pi$ is the discharge stroke. When cross head is at the left of crank, crank is rotating along clockwise.

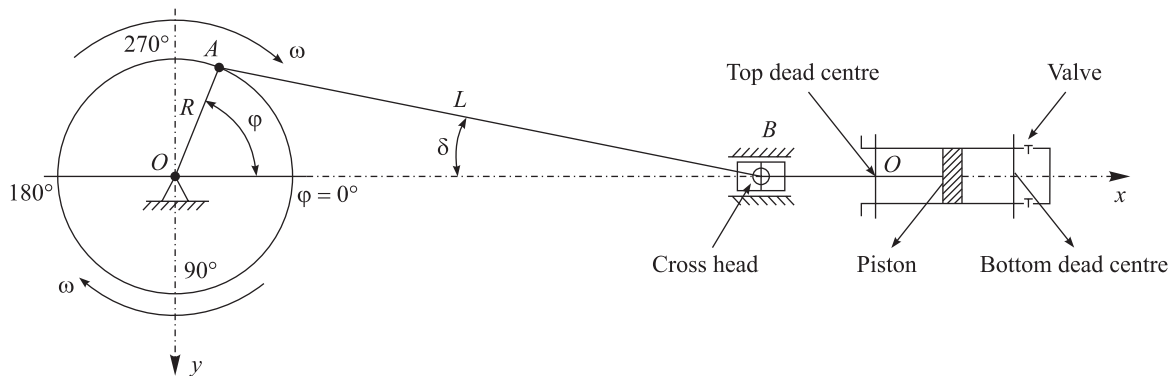


Fig. 1. Working principle of the plunger pump

Plunger displacement formula is stated as follows:

$$s \approx r \left(1 \mp \cos \varphi \frac{\lambda}{2} \sin^2 \varphi \right).$$

Plunger velocity formula is of the form

$$u \approx r\omega \left(\sin \varphi + \frac{\lambda}{2} \sin^2 \varphi \right).$$

The relation between the output pressure p and pump supply (productivity) Q can be stated as follows:

$$p(Q) = \begin{cases} p_{\max}, & \text{when } Q < Q_1, \\ \frac{A}{Q}, & \text{when } Q_1 \leq Q < Q_{\max}. \end{cases}$$

Here $p_{\max}$ is the maximum pressure, and it doesn't change with the increase in the supply Q up to some value Q_1 . Above that value, the output pressure declines non-linearly down to $Q_{\max}$. A corresponds to the maximum useful power of the plunger pump [10].

Specification of the model

The basis for the pump model was the stainless-steel plunger pump model 1051 from CAT pumps. The manufacturer provided open-source CAD models of the pump, which were imported into Simulink. This was utilised to create a mechanical model of the pump for 3D multibody simulation. To accurately represent the dynamic behaviour of the system, the pump was enhanced with hydraulic and electric components. Certain parameters required for developing a digital twin, such as bore, stroke, and shaft diameter, were available in the manufacturer's data sheet².

²Data sheet: 15 frame plunger pumps // Cat Pumps : website. URL: https://www.catpumps.com/sites/default/files/2020-09/1051_G_0.pdf (date of access: 27.08.2025).

Working principle of the triplex plunger pump digital twin

The principle of operation of the pump in question is as follows. The motor drive rotates the crankshaft, after which the motion is transmitted to three plungers that create pressure. Each plunger, as it moves backward, draws fluid from the container through the inlet valve. Before the liquid from the container enters the inlet valve, it passes through a 5 m long, 3 cm radius tube.

Hydraulic pump subsystem

The hydraulic pump subsystem contains three identical plunger models, the housing and the crank. Camshaft has the radius of 5 cm and width of 1.5 cm, while radius of crankshaft is 3 cm and its length is 4.6 cm. In the housing, the centre-to-centre distance between the plungers is 3.2 cm and the distance between the bearings is 20.75 cm [11].

First plunger subsystem

Consider the subsystem that represents the functioning of the first plunger. This subsystem consists of the plunger, the cylinder, the connecting rod, the hydraulic actuator, and the check valves that manage the flow to the intake and exhaust manifolds. The plunger, constructed from metal with the density of 7750 kg/m^3 , measures 7.5 cm in length and has the radius of 1.2 cm. The cylinder is 11.8 cm long, with the outer radius of 2.25 cm. The gap between the plunger and the cylinder wall is 0.5 mm. The piston stroke is 0.03 m. The connecting rod measures 7.25 cm in length. Its width at the pin is 8 mm, while at the crank it is 15 mm. The height of the piston along the radius is 1.14 cm.

A hydraulic actuator functions as a mechanical device that transforms the pressure of hydraulic fluid into a mechanical motion, which can be linear, rotary, or oscillatory. It operates using a piston inside a cylinder to apply force and generate movement. The model of the hydraulic actuator can be characterised by the following equations:

$$\begin{aligned} F &= A_A p_A - A_B p_B - F_c, \\ q_A &= A_A v, \quad q_B = A_B v, \\ \frac{dx}{dt} &= v, \\ v &= v_R - v_C, \\ F_C &= \begin{cases} (x - x_E) K_p \vartheta, & \text{if } x > x_E, v > 0, \\ (x - x_R) K_p \vartheta, & \text{if } x < x_R, v < 0, \\ 0 & \text{otherwise,} \end{cases} \\ x_E &= S - x_0, \quad x_R = -x_0, \end{aligned}$$

where F is the force developed by the cylinder; v is the cylinder rod velocity; v_R and v_C are absolute velocities of the cylinder rod and the cylinder case, respectively; A_A and A_B are piston areas at port A side and port B side; q_A and q_B are flow rates through port A into the cylinder and through port B from the cylinder; x is the piston position; x_0 is the initial distance between the piston and cap A ; F_C is hard stop force; x_E is the distance the piston can travel to fully extend from the initial position; x_R is the distance the piston can travel to fully retract from the initial position; S is the piston stroke.

The purpose of the check valve is to permit the flow in one direction and block it in the opposite direction.

$$\Delta p = p_A - p_B.$$

The valve stays closed as long as the pressure difference across it is below the cracking pressure (fig. 2). Once the cracking pressure is achieved, the valve control element (such as a spool, ball, or poppet) is pushed off its seat, creating a path between the inlet and outlet. If the flow rate is sufficiently high and the pressure keeps increasing, the passage area will expand further until the control element reaches its maximum position. At that point, the valve passage area is at its fullest.

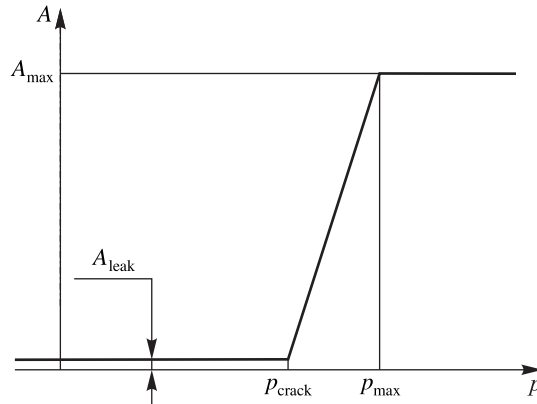


Fig. 2. Dependency between the valve passage area A and the pressure differential across

The opening area of the valve is set as a function of pressure

$$A = A(p).$$

The flow rate through the valve is determined according to the following equations:

$$q = C_D A \sqrt{\frac{2}{\rho} \frac{p}{(p^2 + p_{cr}^2)^{\frac{1}{4}}}},$$

$$A(p) = \begin{cases} A_{leak} & \text{for } p \leq p_{crack}, \\ A_{leak} + k(p - p_{crack}) & \text{for } p_{crack} < p < p_{max}, \\ A_{max} & \text{for } p \geq p_{max}, \end{cases}$$

$$k = \frac{A_{max} - A_{leak}}{p_{max} - p_{crack}},$$

$$\Delta p = p_A - p_B,$$

$$p_{cr} = \frac{\rho}{2} \left(\frac{Re_{cr} \vartheta}{C_D D_H} \right)^2,$$

$$D_H = \sqrt{\frac{4A}{\pi}}.$$

Here q is the flow rate, p_A and p_B are gauge pressures at the block terminals, C_D is the flow discharge coefficient, A is the instantaneous orifice passage area, $A(p)$ is the pressure-dependent orifice passage area, A_{max} is the fully open valve passage area, A_{leak} is the closed valve leakage area, p_{crack} is the valve cracking pressure, p_{max} is the pressure needed to fully open the valve, p_{cr} is the minimum pressure for turbulent flow, Re_{cr} is the critical Reynolds number, D_H is the instantaneous orifice hydraulic diameter, ρ is the fluid density, ϑ is the fluid kinematic viscosity.

For the model developed the following parameters were assumed: maximum passage area $A_{max} = 1 \cdot 10^{-4} \text{ m}^2$, cracking pressure $p_{crack} = 50\,000 \text{ Pa}$, maximum opening pressure $p_{max} = 140\,000 \text{ Pa}$, flow discharge coefficient $C_D = 0.7$, leakage area $A_{leak} = 1 \cdot 10^{-12} \text{ m}^2$. The block positive direction is from port A to port B . This means that the flow rate is positive if it flows from A to B , and the pressure differential is determined as $\Delta p = p_A - p_B$. The opening of the valve is directly proportional to the difference in pressure. Factors such as inertia, friction, springs, and other loads on the valve are not taken into account.

Model validation

Parameter estimation. As already mentioned, some of the parameters required to create a digital copy of the pump are known from the model description supplied by the manufacturer. However, some of the data may differ from model to model and may be specified as gaps and ranges. In this case, the digital model has to

be validated based on measured data from a real sample. For our model the unknown parameters are upper and lower pressures at which the three check valves feeding the outlet will open and close. Since these pressures depend on the temperature of the fluid transported, we don't have the exact values for them.

To start the validation of our digital twin we can roughly estimate the values of lower pressure boundary when the outlet is closed and the upper pressure boundary when the outlet is opened: $p_{\text{low}} = 50\,000\text{ Pa}$, $p_{\text{up}} = 140\,000\text{ Pa}$.

As we can see with these numbers being chosen as values of pressures, the simulated results of the model don't perfectly align with the measured ones (fig. 3). That's why we need to start the process of parameter estimation. We select parameter values, run a simulation, and calculate the difference between the simulated and the measured curves. Using this information, we adjust the parameter values and execute another simulation.

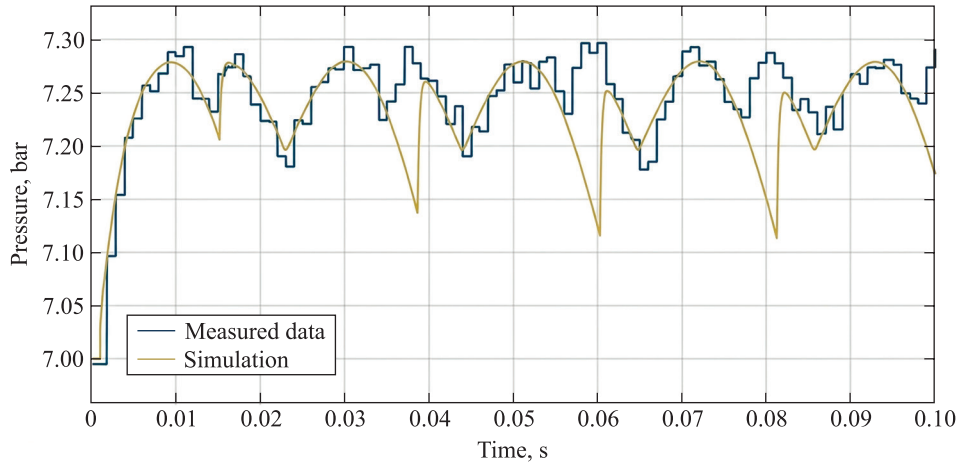


Fig. 3. Simulated output pressure compared with the measured pressure

Sum of square errors method. For the parameter estimation the sum of square errors method (SSE) is used. SSE is a part of one of the general methods most commonly used in the estimation of parameters – the least squares method. The method can be described as follows.

Consider the following variables and parameters: Y is the response or dependent variable, X is the auxiliary or independent variable, A and B are the parameters. The response variable is linear with the parameters $Y = A + BX$.

The objective of the method is to estimate the parameters of the model, based on the observed pairs of values and applying a certain criterium function (the observed pairs of values are constituted by selected values of the auxiliary variable and by the corresponding observed values of the response variable), that is observed values x_i and y_i for each pair I , where $i = 1, 2, \dots, i, \dots, n$. Values to be estimated are A and B , and $(Y_1, Y_2, \dots, Y_i, \dots, Y_n)$ for n observed pair of values.

Object function (or criterium function)

$$\Phi = \sum_{i=1}^n (y_i - Y_i)^2.$$

In the least squares method, the estimators are the values of $A(a)$ and $B(b)$ which minimise the object function. Thus, one has to calculate the derivatives $\frac{\partial \Phi}{\partial A}$, $\frac{\partial \Phi}{\partial B}$, equate them to zero and solve the system of equations in A and B .

The solution of the system can be presented as

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum x, \bar{y} = \frac{1}{n} \sum y, \\ S_{xx} &= \sum (x - \bar{x})(x - \bar{x}), S_{xy} = \sum (x - \bar{x})(y - \bar{y}), \\ b &= \frac{S_{xy}}{S_{xx}}, a = \bar{y} - b\bar{x}. \end{aligned}$$

Notice that the observed values of y , for the same set of selected values of X , depend on the collected sample. For this reason, the problem of the simple linear regression is usually presented in the form $y = A + BX + \varepsilon$, where ε is a random variable with expected value equal to zero and variance equal to σ^2 . So, the expected value of y will be Y or $A + BX$ and the variance of y will be equal to the variance of ε .

The terms deviation and residual will be used in the following ways. Deviation is the difference between y_{observed} and y_{mean} ($\bar{y}$), i. e., deviation = $(y - \bar{y})$.

To analyse the adjustment of the model to the observed data, it is necessary to consider the following characteristics: total sum of squares of the deviations of the observed values [12]

$$SQ = \sum (y - \bar{y})^2.$$

The less is the difference (i. e., error), the better the chosen values are the better the correspondence between the real model and the simulated one.

After several iterations the proper values of $p_{\text{low}} = 30\,000$ Pa, $p_{\text{up}} = 120\,000$ Pa can be obtained. Now the simulated model can reproduce results that are consistent with the measured ones (fig. 4).

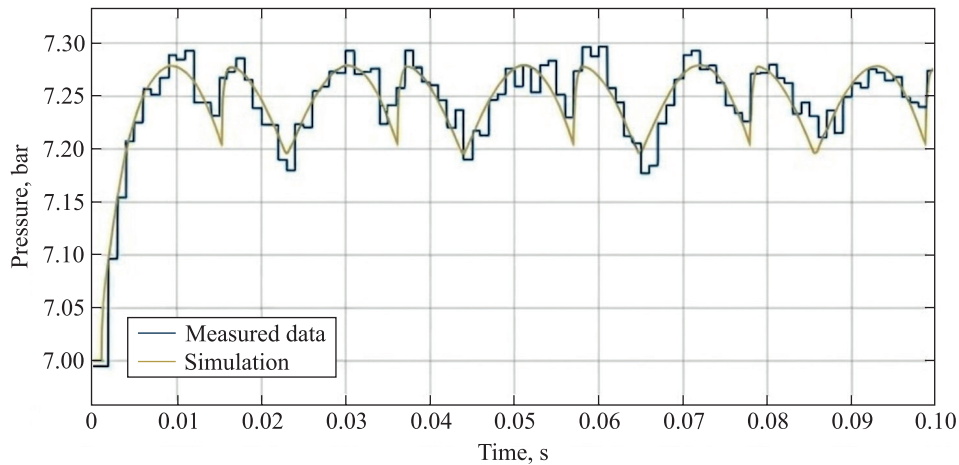


Fig. 4. Simulated output pressure compared with the measured pressure after parameter estimation

Thus, with a validated pump model, it is now possible to run simulations and obtain values for outlet volumetric flow rate, pump speed, and outlet pressure (fig. 5).

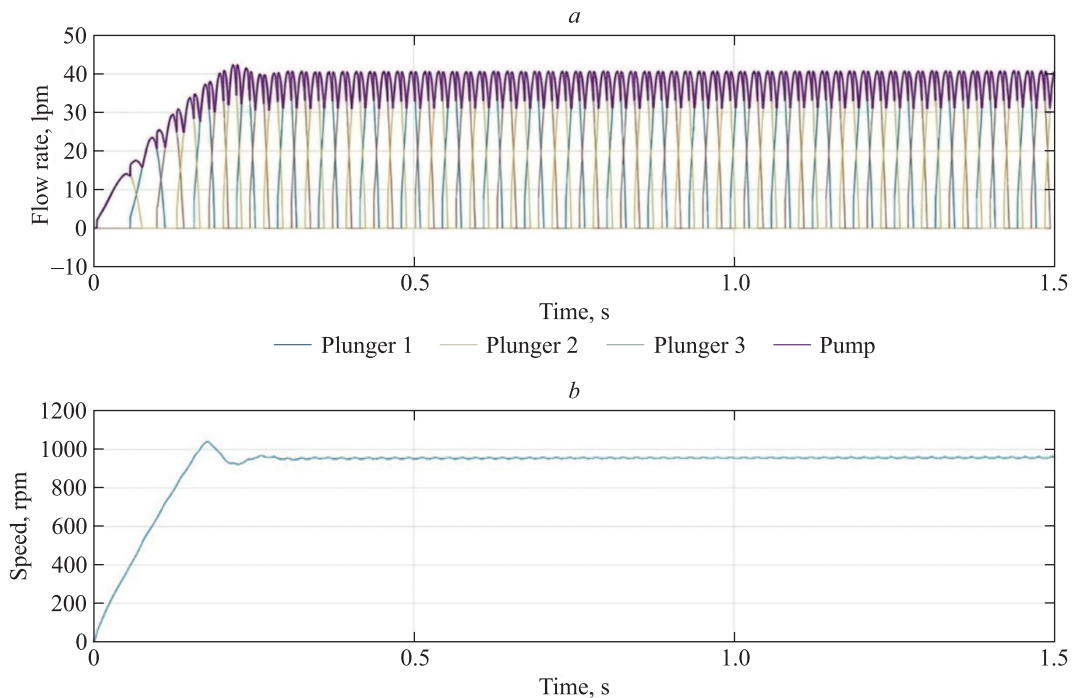


Fig. 5. Volumetric flow rate at output (a) and pump speed (b) of the pump with default configuration

The faults considered in this paper are categorised into three types: seal leakage, blocked inlet, and worn bearing.

Fault modelling. Seal leakage. The modelling of typical pump faults, such as seal leakage, is a critical step towards the predictive maintenance, which has been actively researched for hydraulic systems [13; 14]. Consider the first case. Seal leakage – fluid escapes through the seal gap between plunger and cylinder. The seal leakage

is defined by a certain parameter (hereinafter referred to as the leakage parameter). The leakage is modelled by changing this parameter in one of the pump cylinders. The leakage parameter is represented as a fraction of the maximum cylinder area. In other words, to simulate leakage, we reduce the area of the cylinder bore through which the liquid passes by taking away a fraction of it – the leakage parameter.

The formula for the leakage flow rate is written as follows:

$$Q_{\text{leak}} = C_d A_{\text{leak}} \sqrt{\frac{2\Delta P}{\rho}},$$

$$A_{\text{leak}} = A_{\text{max}} n_{\text{leak}},$$

where Q_{leak} is the leakage flow rate, m^3/s ; C_d is the discharge coefficient; A_{leak} is the leakage orifice area, m^2 ; $\Delta P = P_{\text{in}} - P_{\text{out}}$ is the pressure difference in the cylinder and in the leakage area; A_{max} is the maximum orifice area without leakage; n_{leak} is the leakage fraction in inlet cylinder.

After running the simulation with a leak in the seal, the following results were obtained. From the graphs above, it can be seen that the volume flow rate at the outlet drops compared to the model without any faults (fig. 6 and 7). Also, the cylinder pressure drops in the system and the piston stroke becomes insufficient.

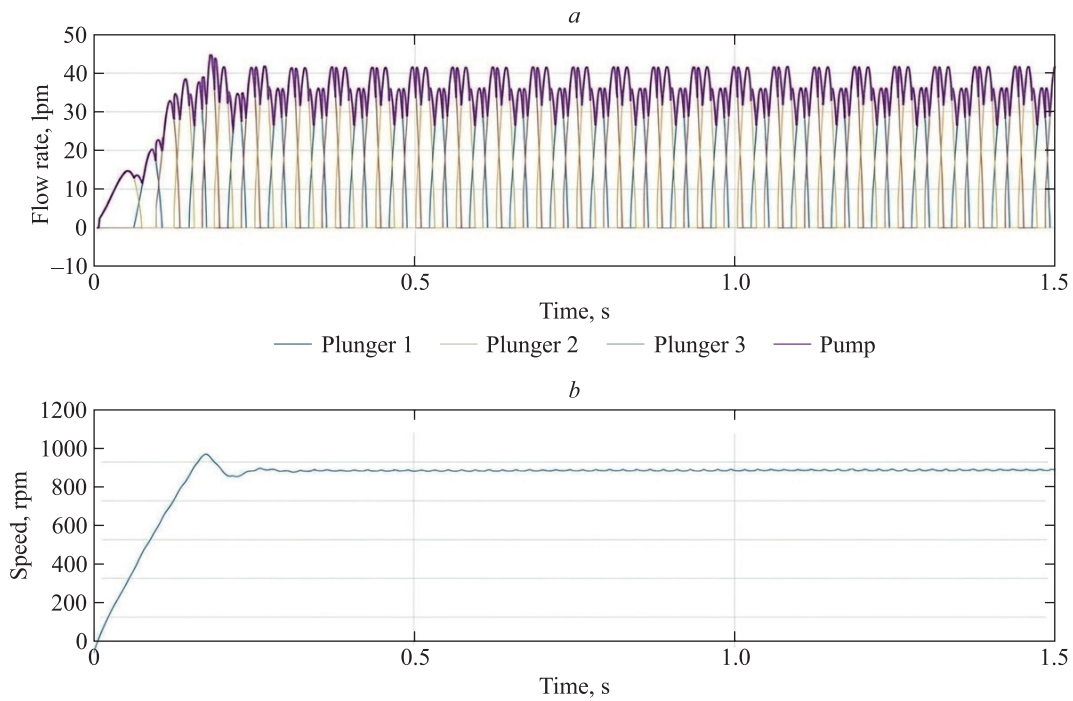


Fig. 6. Volumetric flow rate at output (a) and pump speed (b) of the pump with seal leakage

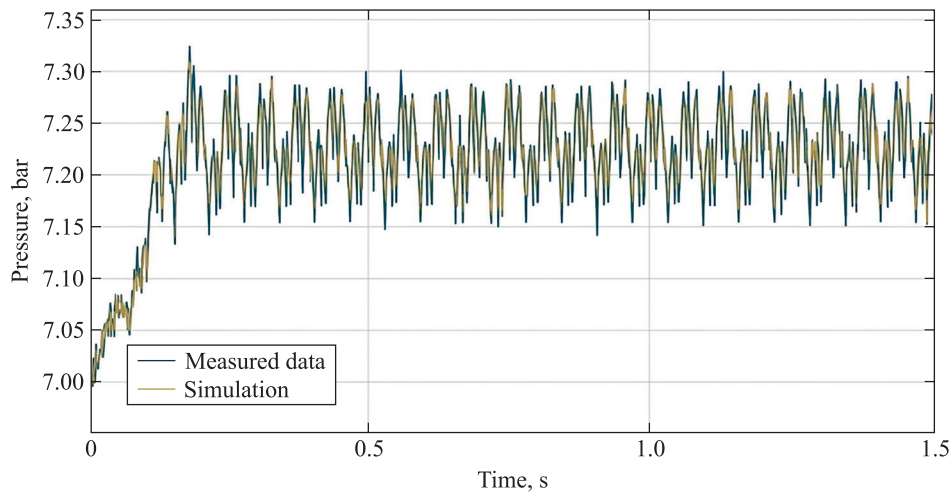


Fig. 7. Output pressure of the pump with seal leakage

Fault modelling. Blocked inlet. A blocked inlet is an inlet restriction where fluid flow is impeded due to a gap, cavitation, or a blocked valve. The case of blocked inlet in this model is realised by reducing the maximum area of the inlet valve with a coefficient of n_{block} .

$$A_{\text{leak}} = A_{\text{max}} n_{\text{leak}}.$$

From these graphs it is easy to see that when the inlet is blocked, the flow and pressure in the pump drop dramatically (fig. 8 and 9). The pump itself becomes intermittent and there is a lack of fluid in the cylinder.

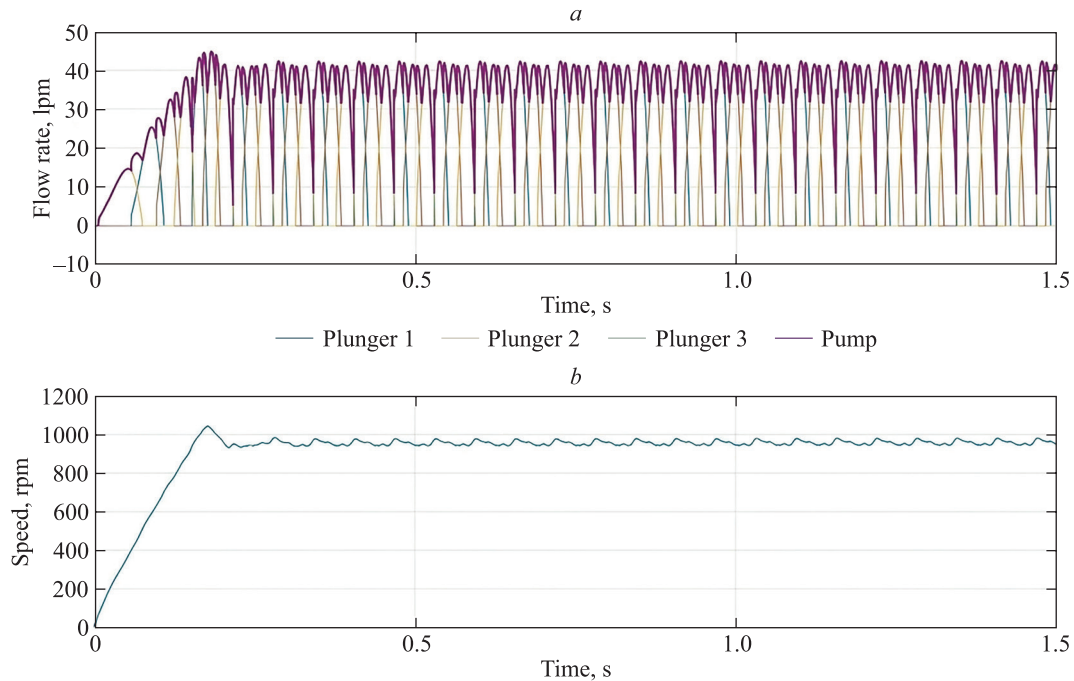


Fig. 8. Volumetric flow rate at output (a) and pump speed (b) of the pump with blocked inlet

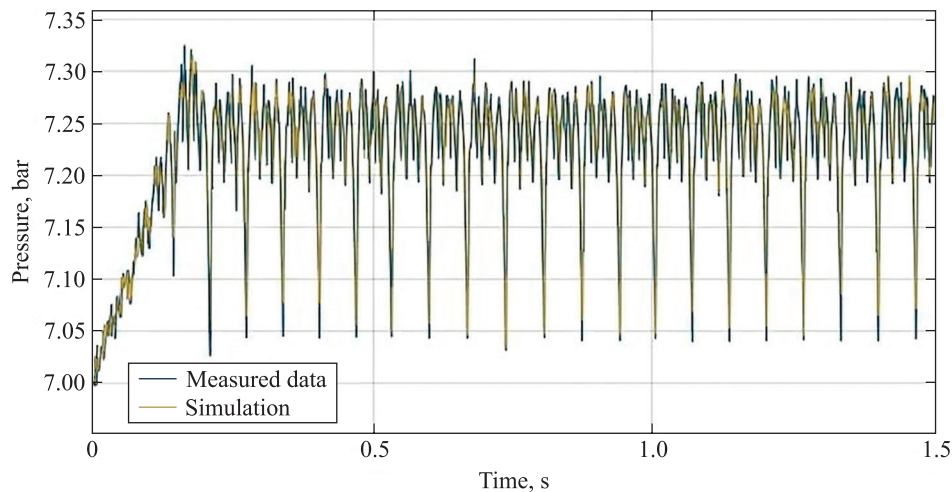


Fig. 9. Output pressure of the pump with blocked inlet

Fault modelling. Worn bearing. A worn bearing is modelled through the coefficient of viscous friction. Rotational friction can be described by the following law:

$$\tau = b\omega,$$

where τ is the resisting torque; b is the viscous friction coefficient; ω is the angular speed of shaft rotation. If the bearing is worn, the coefficient of viscous friction increases. This in turn requires more torque to maintain the same speed. If the shaft rotates slower, this will cause the pistons to pump less fluid. Irregularities in the shaft rotation can also result in changes in the torque transmitted, changes in the hydraulic system pressure, and vibration of the housing or rotor (fig. 10).

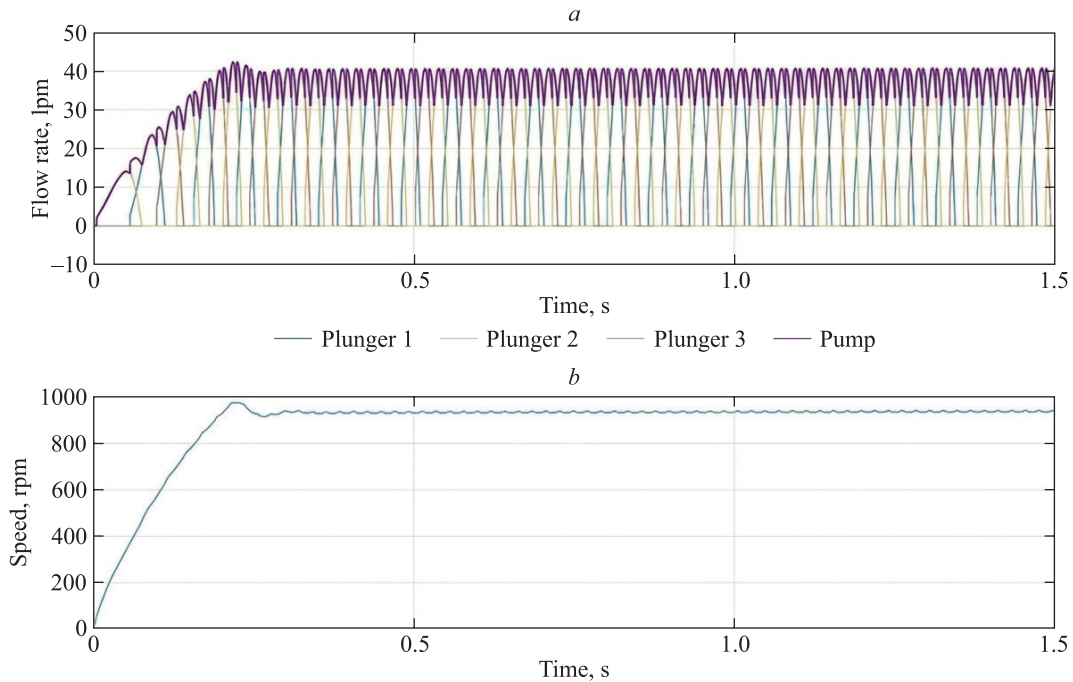


Fig. 10. Volumetric flow rate at output (a) and pump speed (b) of the pump with worn bearing. The speed has increased

Creating a predictive model

Design of experiments. Conducting these simulations is a part of design of experiments. The design of experiments is a special technique for planning, conducting and analysing experiments, using which we have the opportunity to study the influence of various factors (variables) on the result (output characteristic) as efficiently as possible. In other words, the design of experiments is a systematic way of testing in which we decide in advance what to change and how to do it. From this we can subsequently understand which parameters affect the system and estimate how much they do. This will allow us to build a mathematical model of the behaviour of the system and find the optimal settings for it [15; 16].

In our case, the parameters used in the design of experiments are the cylinder leakage fraction, inlet blockage area, and viscous friction coefficient. By utilising the simulation data to create pressure data for the pump across all potential fault conditions, a large number of scenarios can be created for the digital twin. To avoid random errors in the sensor, each scenario must be simulated a large number of times. To model a large number of simulations, an approach can be used in which the simulation takes some time before reaching a steady state. The part where the system reaches steady state is simulated only once, after which each subsequent simulation starts from the point where the system reached steady state. In this way, up to two-thirds of the computation time can be saved (fig. 11).

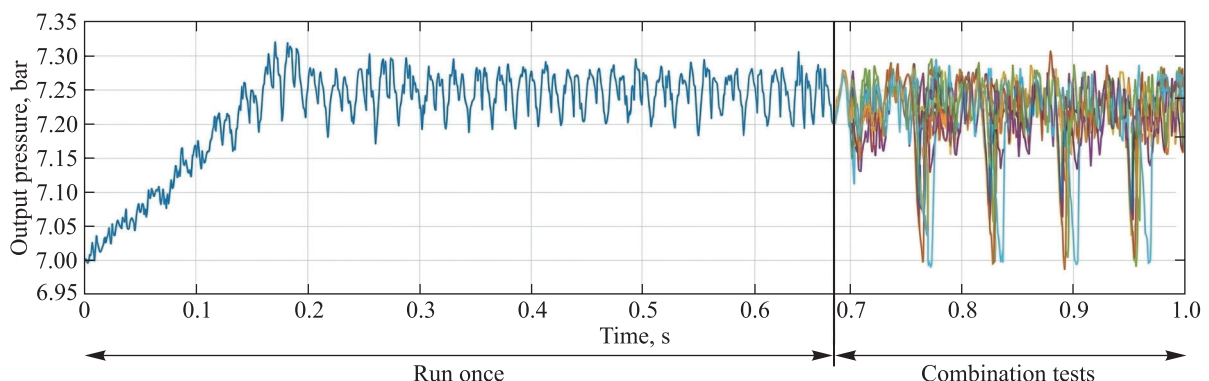


Fig. 11. Sweep test of faults (individual and combination)

Data pre-processing. The large number of simulations performed allows the creation of a large data set, which is one of the key factors for creating an adequate digital twin model, since it is on the basis of this data that the predictive model is trained. The more data on the flow rate, pressure, and pump outlet speed are obtained in different configurations of sample behaviour, the better the model will be able to predict the mechanism behaviour. However, it is important to keep in mind that when collecting data from sensors, or modelling them, the phenomenon of noise – short-term spikes that are atypical for the model in question and beyond the sensitivity of the sensor – can often appear. Therefore, before loading data into the model for its subsequent training, it is important to process the information received from the sensors, to get rid of noise and gaps. After this procedure, the processed data can be used to train the predictive model.

Having collected the leakage output pressure data at the first plunger, we load this data into the *MATLAB* environment for further processing.

As shown in fig. 12, the following problems can be seen in the raw data. There is noise in the graph, which shows up as spikes that exceed the maximum sensor sensitivity of 9 bar. There are gaps in the data (points where the output pressure has not been measured). The data is time-shifted – even though the traces are the same duration, the traces are preceded by different measurement delays.

Now it is necessary to process the information obtained: we get rid of outliers (values at the sensor limit) and fill in the gaps where the data has not been measured. To do this, we remove all areas where the output pressure exceeds the maximum value of 9 bar. We fill in the gaps using linear interpolation

$$y = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)}(x - x_1),$$

where time is used as the independent variable x and the dependent parameter y is the output pressure (see fig. 12).

One can see that the noise and gaps are now absent from the processed data plot. Having tested the data processing steps on a single trace, we will apply these operations to all collected data. The graphs below show all the output pressure data before and after processing (fig. 13).

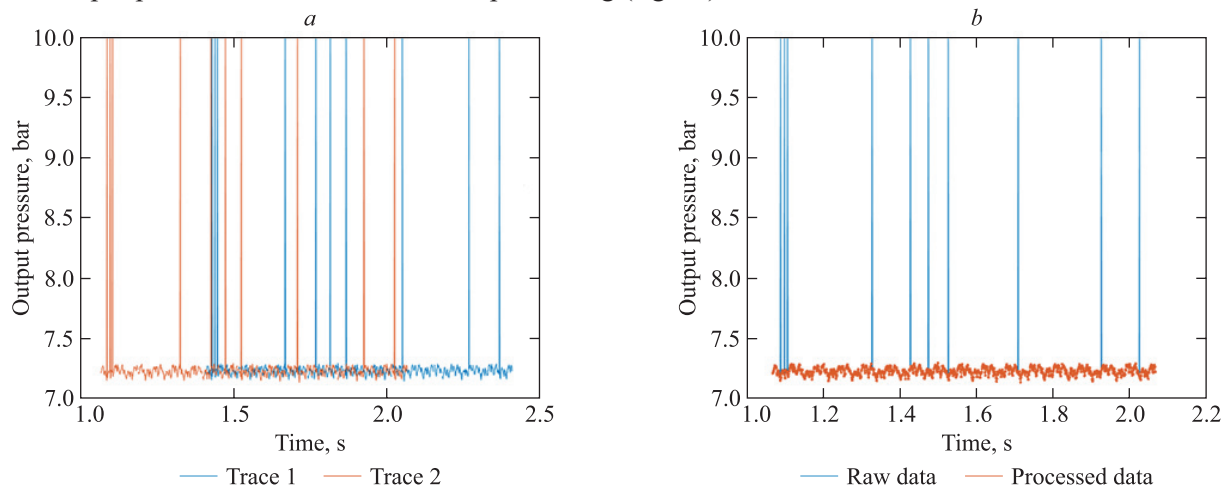


Fig. 12. Output pressure measurements depending on processing:
a – data is not processed yet; b – raw data is compared with processed ones (remove noise, fill gaps)

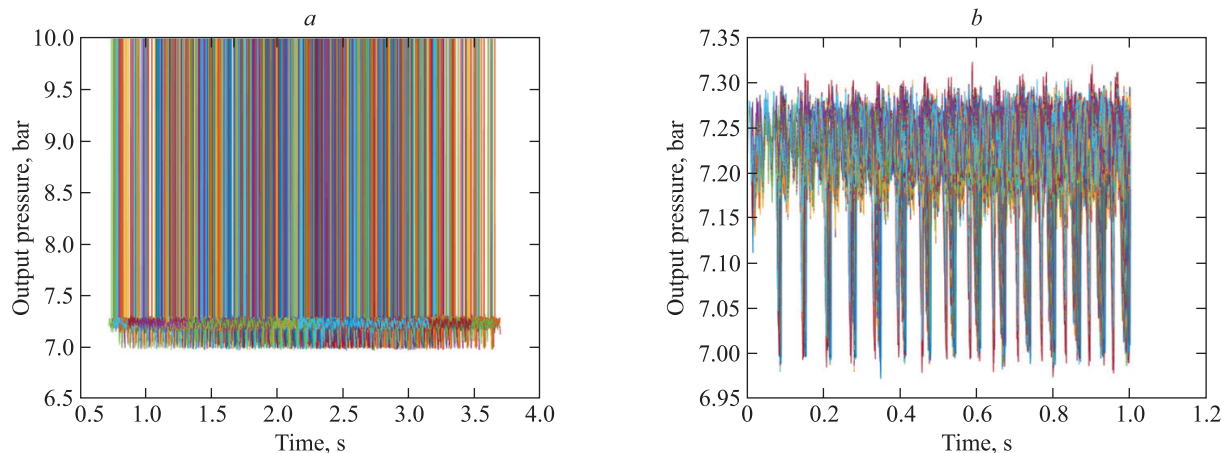


Fig. 13. Output pressure measurements for multiple simulations depending on processing:
a – data is not processed; b – data is processed

Display the sensor data for different pump faults and combinations of faults. We can compare how the behaviour of a triplex plunger pump changes as the parameters in the pump degrade. The cyclic nature of the time domain signal provides unique signatures for each fault and their combinations (fig. 14).

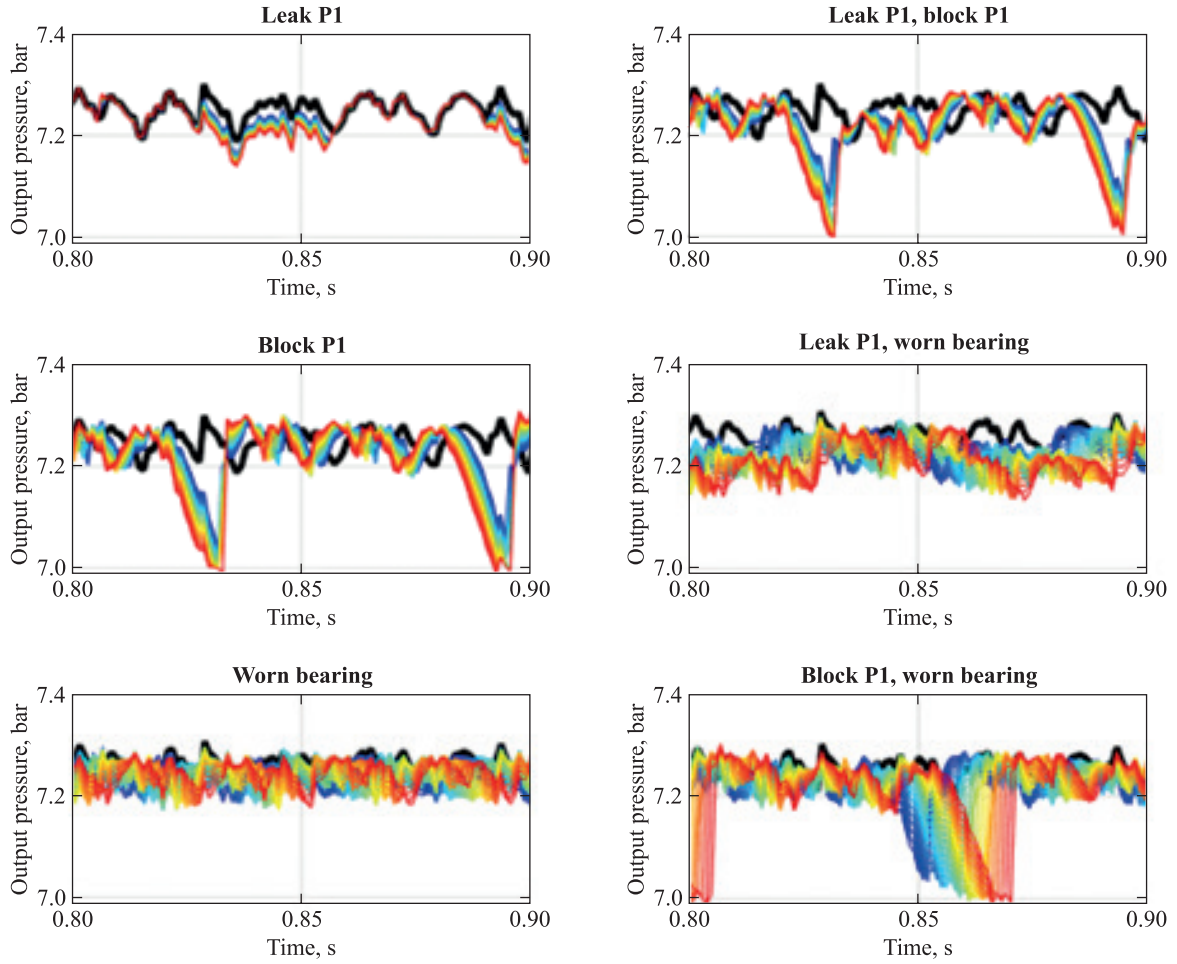


Fig. 14. Output pressure measurements for multiple simulations depending on processing

Machine learning algorithm

Fourier transforms. Once the sensor data has been processed, the patterns and distinguishing features of each fault need to be extracted from it. This information will be used as the input data for the machine learning algorithm. The signal under consideration is periodic, so we will apply Fourier transforms [17; 18].

Signal $y(t_i)$, given at discrete points in time t_i , is translated to a uniform time grid

$$t_k = k\Delta t,$$

where $k = 0, 1, \dots, N-1$; $\Delta t = \frac{t_{\text{end}}}{N-1}$,

$$y_{\text{intp}}(t_k) = \text{interp1}(t_i, y(t_i), t_k).$$

To reduce spectrum leakage, Hamming window function is applied [19]

$$y_w(t_k) = y_{\text{intp}}(t_k)w_k, \quad w_k = 0.54 - 0.46\cos\left(\frac{2\pi k}{N-1}\right),$$

where w_k are the coefficients of Hamming window.

Then the discrete Fourier transform is performed

$$Y_n = \frac{1}{N} \sum_{k=0}^{N-1} y_w(t_k) e^{-j\frac{2\pi nk}{N}}, \quad n = 0, 1, \dots, N-1.$$

Spectrum modulus (amplitude at each frequency)

$$P_n = |Y_n| = \sqrt{\operatorname{Re}(Y_n)^2 + \operatorname{Im}(Y_n)^2}.$$

Corresponding angular frequency

$$\omega_n = \frac{2\pi n}{N\Delta t}, \quad n = 0, 1, \dots, \frac{N}{2}.$$

In our case we consider only frequencies from the range $\omega_n \in (20, 500)$ rad/s. For these frequencies, the signal amplitudes at these frequencies are returned. By computing Fourier transforms for all fault simulations, we can observe how the faults progress in the frequency space (fig. 15).

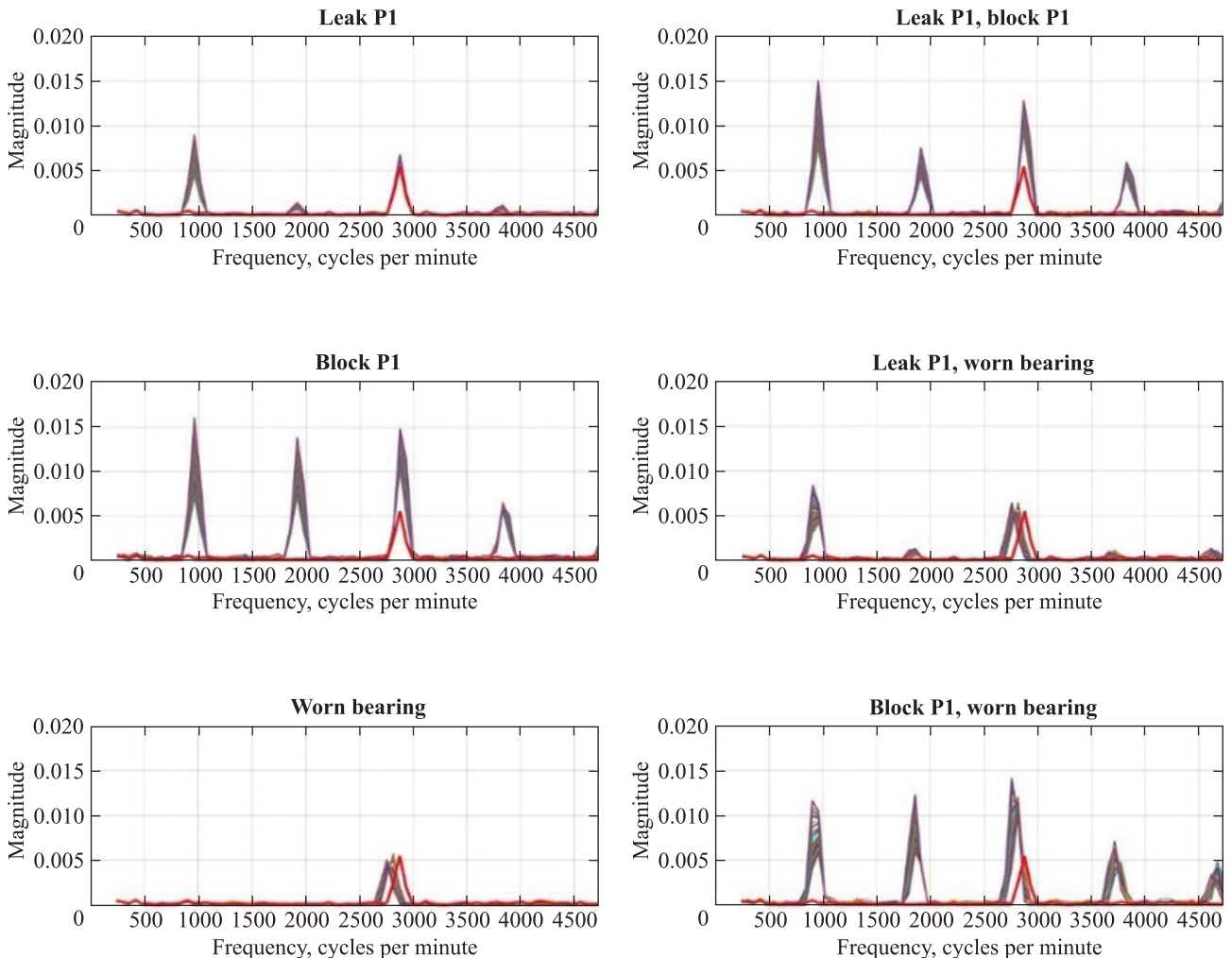


Fig. 15. Frequencies for different fault scenarios

The FFT results for every fault scenario are compiled into a table that includes the faults introduced, as well as the recorded frequencies and signal magnitudes. Consequently, the number of parameters to take into account is fairly limited. Now we can use *Machine Learning Toolbox* to import there all the necessary data for training the fault detection algorithm. We will use a subset of the generated data to validate the trained algorithm.

Extraction algorithm. The algorithm for extracting features from the pressure signals obtained from the pump model for subsequent training of the machine learning model can be described as follows. By performing Fourier transforms, the algorithm extracts the main peaks of the amplitude spectrum and also reads the mean values of the signals [20].

Definitions:

- N_p – number of points to analyse (time window);
- $T_j(t)$ – time axis for the j^{th} simulation;
- $P_j(t)$ – output pressure in the j^{th} simulation;

- $S_j(f)$ – signal amplitude spectrum $P_j(t)$;
- f_j – frequency vector corresponding to $S_j(f)$;
- $f_{nom} = 958 \cdot 3$ – nominal speed multiplied by 3 steps.

1. Converts a signal in the time domain to the frequency domain. For each test j [21]

$$\hat{P}_j(f) = \left| \mathcal{F} \{ P_j(t) w(t) \} \right|,$$

where $\mathcal{F}\{\cdot\}$ – discrete Fourier transform; $w(t)$ – Hamming window. Speed standardisation

$$f_j^{norm} = \frac{f_j}{f_{nom}}.$$

2. Extraction of peaks in the spectrum. From the spectrum $\hat{P}_j(f)$ peaks with a threshold of prominence (height relative to neighbours) are selected

$$\left\{ (f_{j,1}, A_{j,1}), \dots, (f_{j,K}, A_{j,K}) \right\} = \text{findpeaks}(\hat{P}_j(f), \text{prominence} > 0.0005),$$

where K – number of peaks found. The found peaks are converted to relative frequencies

$$\delta f_k = \left| \frac{f_{j,k}}{f_{nom}} - 1 \right|$$

and sorted by proximity to the fundamental frequency f_{nom} :

$$\text{sort}(\delta f_k) \Rightarrow 5 \text{ closest peaks are selected.}$$

3. Calculation of average pressure values:

$$\bar{P}_j = \frac{1}{N_p} \sum_{n=1}^{N_p} P_j(t_n).$$

As a result, a string of features is formed for each test j

$$x_j = [\delta f_{j,1}, A_{j,1}, \dots, \delta f_{j,5}, A_{j,5}, \bar{P}_j],$$

where $\delta f_{j,k}$ is the frequency of the k^{th} peak relative to the main peak; $A_{j,k}$ is the k^{th} peak amplitude; $\bar{P}_j$ is the average pressure.

Subsequently, we import the trained algorithm that has achieved maximum accuracy in detecting pump faults based on measured data into a digital twin. Validation using seven test cases shows that the algorithm is able to reliably detect these seven scenarios.

Conclusions

The article focuses on uncovering the significance of digital twins in various areas of human endeavour. Thus, this work has successfully developed a digital twin of a triplex plunger pump with an integrated predictive model capable of diagnosing and predicting failures based on the analysis of system input and output data. The model was trained on three separate fault cases: leakage, closed inlet, and worn bearing, as well as various combinations of them. This allowed the model to effectively recognise the characteristic signs of failures, forming a stable knowledge base for accurate prediction of the equipment's technical condition.

The resulting system demonstrates high application relevance for practical implementation in predictive maintenance strategies. Predictive maintenance reduces downtime and prevents equipment failures by enabling maintenance planning based on actual demand rather than a predefined schedule. By automating data analysis and the model's ability to adapt, the digital twin can respond quickly to changes in the operating environment. This approach forms the basis for intelligent condition management and contributes to more sustainable and reliable operation of engineering systems.

The novelties of this investigation are:

1. A system stabilisation approach was used to accelerate computation. The simulation took some time to reach a steady state. This part of the system was simulated only once, after which each subsequent simulation started from the point at which the system reached steady state.

2. Fourier transforms were used to extract the key data needed to train the predictive model. The signal was transformed from the time domain to the frequency domain, after which peaks – maximum values – were extracted from the frequency spectrum.

References

1. Parrott A, Warshaw L. *Industry 4.0 and the digital twin: manufacturing meets its match*. [S. l.]: Deloitte University Press; 2017. 18 p. (A Deloitte series on Industry 4.0, digital manufacturing enterprises, and digital supply networks).
2. Liu M, Fang S, Dong H, Xu C. Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*. 2021;58(part B):346–361. DOI: 10.1016/j.jmsy.2020.06.017.
3. Hananto AL, Tirta A, Herawan SG, Idris M, Soudagar MEM, Djamari DW, et al. Digital twin and 3D digital twin: concepts, applications, and challenges in Industry 4.0 for digital twin. *Computers*. 2024;13(4):100. DOI: 10.3390/computers13040100.
4. Wang Z, Gupta R, Han K, Wang H, Ganlath A, Ammar N. Mobility digital twin: concept, architecture, case study, and future challenges. *IEEE Internet of Things Journal*. 2022;9(18):17452–17467. DOI: 10.1109/JIOT.2022.3156028.
5. Bhatti G, Mohan H, Singh RR. Towards the future of smart electric vehicles: digital twin technology. *Renewable and Sustainable Energy Reviews*. 2021;141:110801. DOI: 10.1016/j.rser.2021.110801.
6. Zhuravkov MA, Nikolaitchik MA. Construction of mine elements digital twins. *Mining Mechanical Engineering and Machine-Building*. 2023;2:5–14. Russian. EDN: ARDPNA.
7. Zhuravkov MA, Nikolaitchik MA, Wang Z, Yu G. Fatigue life prediction of hoist drum-rail contact pair based on digital twin. *Results in Engineering*. 2025;26:105271. DOI: 10.1016/j.rineng.2025.105271.
8. Andenyangtso NE, El-Suleiman A, Chollom TD, Yekini B. Development of a test stand for determining the technical condition of a plunger pump. *FUW Trends in Science & Technology Journal*. 2024;9(1):410–414.
9. Lian Z, Liu Y, Lin T, Li L, Lei Z, Han C. Study on the fluid flow rule of five-cylinder plunger pump hydraulic end. *Petroleum*. 2018;4(4):457–465. DOI: 10.1016/j.petlm.2017.11.001.
10. Velikanov NL, Naumov VA, Koryagin SI. Characteristics of plunger pumps. *Russian Engineering Research*. 2018;38(6):428–430. DOI: 10.3103/S1068798X18060175.
11. Karassik IJ, Messina JP, Cooper P, Heald CC. *Pump handbook*. 4th edition. New York: McGraw-Hill; 2007. 1632 p.
12. Cadima EL. *Fish stock assessment manual*. Rome: FAO; 2003. XX, 161 p. (FAO fisheries technical paper; No. 393).
13. Rivera DL, Scholz MR, Fritscher M, Krauss M, Schilling K. Towards a predictive maintenance system of a hydraulic pump. *IFAC-Papers Online*. 2018;51(11):447–452. DOI: 10.1016/j.ifacol.2018.08.346.
14. Green E, Taylor L. Development of a predictive maintenance framework for hydraulic systems using IoT and machine learning. *International Journal of Mechanical Industrial and Control Systems Engineering*. 2024;1(2):48–53. DOI: 10.61132/ijmicse.v1i2.83.
15. Box GEP, Hunter JS, Hunter WG. *Statistics for experimenters: design, innovation, and discovery*. 2nd edition. Hoboken: Wiley-Interscience; 2005. XVII, 633 p. (Wiley series in probability and statistics).
16. Montgomery DC. *Design and analysis of experiments*. 9th edition. Hoboken: John Wiley & Sons; 2017. XV, 734 p.
17. van Drongelen W. *Signal processing for neuroscientists: an introduction to the analysis of physiological signals*. Amsterdam: Academic Press; 2007. Chapter 6, Continuous, discrete, and fast Fourier transform; p. 91–105. DOI: 10.1016/B978-012370867-0/50006-1.
18. Ma L, Jiang A, Jiang W. The intelligent diagnosis of a hydraulic plunger pump based on the MIGLCC-DLSTM method using sound signals. *Machines*. 2024;12(12):869. DOI: 10.3390/machines12120869.
19. Patseika PG, Rouba YA, Smatrytski KA. On the approximation of conjugate functions and their derivatives on the segment by partial sums of Fourier – Chebyshev series. *Journal of the Belarusian State University. Mathematics and Informatics*. 2024;2:6–18. EDN: ZITNRS.
20. Lughofer E, Sayed-Mouchaweh M, editors. *Predictive maintenance in dynamic systems: advanced methods, decision support tools and real-world applications*. Cham: Springer; 2019. XIII, 567 p. DOI: 10.1007/978-3-030-05645-2.
21. Patseika PG, Rouba YA. Fejer means of rational Fourier – Chebyshev series and approximation of function $|x|^s$. *Journal of the Belarusian State University. Mathematics and Informatics*. 2019;3:18–34. Russian. DOI: 10.33581/2520-6508-2019-3-18-34.

Received 19.12.2025 / revised 10.02.2026 / accepted 17.03.2026.