Research of methods for pumping technological equipment condition predictive monitoring

V V Bukhtoyarov\textsuperscript{1,2}, V S Tynchenko\textsuperscript{1,2}, K A Bashmur\textsuperscript{1}, E A Petrovskiy\textsuperscript{1}, V V Kukartsev\textsuperscript{1,2} and N A Bukhtoyarova\textsuperscript{1}

\textsuperscript{1}Siberian Federal University, 79, Svobodny pr., Krasnoyarsk, 660041, Russian Federation
\textsuperscript{2}Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation

E-mail: vladber@list.ru, vadimond@mail.ru

Abstract. The article discusses the need to develop a diagnostic prognostic unit for certain types of technological equipment. Several well-known methods for predicting time series are considered: a method based on artificial neural couples, a method for analyzing a singular spectrum, an exponential smoothing method. A brief description of the methods under consideration is given. The parameters of the numerical study of the methods are described. The results of a numerical study are presented, the advantage of the neural network approach and the method of singular spectral analysis are shown.

1. Introduction

Currently, high demands are placed on production systems in terms of reliability, safety and operational efficiency. These properties are largely determined by the ability to predict failures for individual elements of such production systems. In this regard, it seems relevant to study the construction of high-precision diagnostic systems for individual elements of production systems. This plays an important role in the implementation of various production processes. First of all, for those elements and for the equipment that plays an important role in the implementation of various production processes.

Technological pumping units are one of the most common types of equipment in many industries, also in the oil and gas industry. The continuity and production efficiency of the technological chains both in the production and transportation of oil and gas and in oil refining largely depend on the efficiency and reliability of their functioning. Experience in the implementation of processes in the relevant industries shows that up to a 20\% decrease in efficiency and production line shutdowns due to a technical reason is associated with the influence of unstable operation of technological pumping units. Moreover, a significant proportion of such cases is defined as a sudden failure of the pumping unit during operation. However, experience shows that in the course of further analysis, cause-effect chains that are quite distributed in time are determined, leading to shutdowns of pumping units [1-3].

Therefore, the development and study of methods for monitoring the technical condition of such equipment in operation is an urgent area. It seems that the strengthening of the monitoring component is possible due to the construction of procedures for forecasting the technical condition based on the predictive model of technological equipment trends of parameters. Such a model, in addition to
analyzing real-time data, will allow integrating predicted values into the decision-making procedure on the state of a pumping unit.

In the aspect of the current direction of development of production systems within the framework of the Industry 4.0 concept, the construction of technical state forecasting systems should be carried out taking into account the properties of self-diagnosis, self-monitoring and machine-to-machine interaction [4-6]. This suggests that, in fact, an effective predictive diagnostic module should be created. Such a module, based on information from a specific pump installation, and, if there are similar installations in the equipment fleet, based on information from them, provide forecasting of technical condition parameters. To ensure the efficiency of such a module, it must be implemented using high-precision models for predicting the parameters of technological equipment (pumping units). Given the concept of intellectualization of production systems and the need to build accurate forecasts based on a significant amount of information, it is proposed to build such predictive models based on the most modern approaches to data mining. To determine the options for constructing such a predictive model system, we consider a scheme for using a technical condition monitoring module with a predictive model. Methods of statistical processing of data series and data mining methods that can be used as analytical support for forecasting and modeling for such a module are considered [7-9]. The following are conditions and discusses the results of numerical studies of the proposed methods for constructing prognostic models for the installation of process pumps.

2. Description of subject methods
Taking into account a significant number of types of technological equipment, in addition to pumping equipment, the general scheme of the application of prognostic models of monitoring and diagnostics is considered. It is assumed that, having been worked out for certain types of equipment, such a universal model-diagnostic module can be created for an extensive number of technological equipment. In the future, the integration of such modules into a hierarchical system will allow to create flexible platform solutions for a large number of diverse technological systems. In this regard, it seems important to develop and study the algorithmic base - predictive models for assessing the technical state, since their effectiveness will largely determine the effectiveness of the integrated system of monitoring and diagnostics.

By forecasting is meant the prediction of future values of a time series by its current and past values. Let us consider briefly the statement of the problem of forecasting the values of time series. Suppose that at discrete equidistant times, observations $x(1), x(2), ..., x(t)$ are available for the object. Let a function $\hat{x}(l), l = 1,2, ...$, be a function that gives at time $t$ a forecast for all future times $t + l$ (i.e., with lead $l$). Such a function is called a predictive function. It is necessary to obtain a predictive function $\hat{x}(l)$ for which the deviation of predicted value from the real value would be the smallest for each lead $l$. The square deviation of predicted values from real values is usually minimized.

2.1. Artificial neural networks
An artificial neural network (ANN) is called computational structures that simulate biological processes that are usually associated with the processes of the human brain [10-12]. They are distributed and parallel systems capable of adaptive learning. An artificial neuron, called by analogy with a biological prototype, is used as an elementary converter in such networks. Artificial neural networks are widely used in various fields in solving problems associated with forecasting and diagnostics [13-17].

In this paper, neural networks are considered in the aspect of solving the forecasting problem and ANN in accordance with the classical approach is considered as a “black box”. Accordingly, the goal of solving problems is to obtain an effective computational procedure, and not to extract knowledge and hidden patterns in an explicit form.

2.2. Exponential smoothing
Exponential smoothing is one of the simplest and most common techniques for processing of a time series [18,19]. Exponential smoothing can be represented as a filter, at the input of which the members of the original series successively arrive, and at the output the current values of the exponential average are formed.

2.3. Singular spectrum analysis method
The singular spectrum analysis (SSA) method is a time series analysis method that uses the main component analysis method as a basic approach. The SSA method involves the conversion of a one-dimensional time series into a multidimensional one and the subsequent application of the principal component method to it [20].

This method is widely used for the analysis of time series, including such areas as the analysis of climatological information, forecasting in financial analysis, geophysics and many other areas [21-23]. Significantly less scientific research is associated with the use of this approach for the analysis of time series describing technical systems. Given the rather high efficiency of applying methods in other directions, it seems important to study it as part of the work on predicting the state parameters of technological pumping units. The basic version of the SSA method consists of four steps, a detailed description of which is given in [20].

At present, a significant number of modifications of this method have been proposed, which allow one to obtain more efficient solutions for some applications [24]. In our work, the Caterpillar SSA 3.40 program was used to calculate the estimates using this method.

3. Experimental study
As test forecasting tasks, we used data samples generated from the Synthetic Control Chart Time Series Data Set dataset obtained from the Machine Learning Repository [25]. This dataset is a synthetic test for prediction algorithms. The data in it is an imitation of control chart data compiled from observations of technological processes. The data set includes examples specific to different classes of time series. The following are the names of the classes of time series used for testing methods, and drawings with images of a typical type of series for each of the classes. For clarity, the points of the row are connected by lines:

- Time series without trend, cyclicity and bias.
- Cycle time series.
- Time series with an increasing trend.
- Time series with a decreasing trend.

The use of different types of time series in test problems makes it possible to evaluate forecasting algorithms well in terms of their ability to predict time series with different characteristics. Preliminary processing of data by smoothing methods was not carried out.

An important part of the test data used was the data obtained from the technological pumps of the refinery of a Russian oil company. In accordance with the policy of the oil company, its name and the name of the diagnostic object cannot be disclosed. The data were normalized using the Statistica applied statistical analysis package. The number of observations in each time series was reduced to 60 by averaging neighbouring points.

As the main criterion of effectiveness, the mathematical expectation was used, which was calculated according to the results of the studied methods in 20 independent runs. In each of the launches, a unique time series of a specific type (1-4) was used, generated from the data set under consideration. There are 60 values in each time series. In this study, the following parameters were used to model the predictive function:

- Number of previous measurements taken into account when calculating the forecast: 3.
- Number of ticks for which the forecast is calculated: 1.
To calculate the forecast error, we used a formula:

\[
Error = \frac{100\%}{s(y_{\text{max}} - y_{\text{min}})} \sum_{i=1}^{s}|o_i - y_i| \tag{1}
\]

Here \( i \) is the point in time, \( o_i \) is the forecast obtained using the collective of neural networks, \( y_i \) is the true value of the time series at time moment \( i \), \( y_{\text{max}} \) and \( y_{\text{min}} \) is the maximum and minimum value in the time series, \( s \) is the total number of predicted values.

For all methods, in order to obtain the correct results of numerical experiments, the same restriction was used on the maximum number of available computing resources in one run of the corresponding algorithm.

A pairwise comparison of the studied methods to identify statistical significance in the distinguishability of the results was carried out by ANOVA methods at a significance level 0.05. The results of the research methods on the task of forecasting time series are given in the table 1.

**Table 1. The results of problem solving.**

| Controller / Signal source | Artificial Neural Networks | Exponential smoothing | SSA |
|---------------------------|---------------------------|----------------------|-----|
| Time series without trend, cyclicity and bias | 8 | 15.9 | 7.5 |
| Cycle time series | 6.9 | 23.5 | 6.7 |
| A time series with an increasing trend | 8.4 | 16.4 | 8.1 |
| The time series with a decreasing trend | 7.1 | 15.6 | 7.3 |
| Trends of refinery technological pumping units | 8.3 | 21.6 | 8.8 |

The results demonstrate the advantages of the neural network methods and the SSA method over the method of exponential smoothing. In general, the artificial neural network method and the SSA method demonstrate similar results in many of the problems under consideration. Statistical analysis shows the absence of a significant difference in the results obtained by these two methods. Therefore, it is precisely these two methods that are preferably used in the construction of highly accurate prognostic modules. The differences in resource allocation for such methods seem important. A neural network requires resource-intensive training, but further training and forecasting requires relatively lower computational costs.

### 4. Conclusion

The article considers the problem of choosing algorithmic support for solving the problem of monitoring and predicting the state of a technological pumping unit. A scheme is proposed for using a unit for predicting the technical condition of a pumping unit under conditions of monitoring parameters carried
out by standard means for installations of this type. Important is the choice of forecasting methods for the algorithmic core of such diagnostic forecasting blocks.

The article analyzes the methods of forecasting time series that can be used to solve the problems under consideration. A description of the methods is given and the results of the study of methods on a set of test diagnostic data and data from real pumping units of an oil refinery are given. It is shown that methods based on neural networks and the method of analyzing the singular spectrum are much more effective than the simple method of exponential smoothing. However, the use of such methods is more resource-intensive and requires special calculations, in particular, training for neural networks and the principal component method for SSA. In this regard, in the future it is planned to develop an adaptive prognostic module that allows flexible determination of the used algorithmic basis at the stage of system initialization, taking into account the features of the diagnosed equipment.

Acknowledgments
The reported study was partially funded Scholarship of the President of the Russian Federation for young scientists and graduate students SP.869.2019.5.

References
[1] Oscar S and Anvar V 2017 The monitoring system of an actual technical condition for pumping units with frequency analysis Procedia Engineering 176 144-9
[2] Dong S 2012 Research on the Technology of Remote Intelligent Fault Diagnosis Based on Electrical Parameters of Pumping Unit Process Automation Instrumentation 5 182-92
[3] Attivissimo F, Di Febo M, Paganini P and di Mango G 2014 Model based control and diagnostic system for centrifugal pumps 2014 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems Proceedings 1-6
[4] Niggemann O, Biswas G, Kinnebrew J S, Khorasgani H, Volgmann S and Bunte A 2015 Data-Driven Monitoring of Cyber-Physical Systems Leveraging on Big Data and the Internet-of-Things for Diagnosis and Control DX@ Safeprocess 185-92
[5] Lee J and Bagheri B 2015 Cyber-physical systems in future maintenance 9th WCEAM Research Papers 299-305
[6] Wang L, Tömgren M and Onori M 2015 Current status and advancement of cyber-physical systems in manufacturing Journal of Manufacturing Systems 37 517-27
[7] Harirchi F and Ozay N 2018 Guaranteed model-based fault detection in cyber–physical systems: A model invalidation approach Automatica 93 476-88
[8] Bunte A, Stein B and Niggemann O 2019 Model-Based Diagnosis for Cyber-Physical Production Systems Based on Machine Learning and Residual-Based Diagnosis Models Thirty-Third AAAI Conference on Artificial Intelligence 1-8
[9] Gorunescu F 2011 Data Mining: Concepts, models and techniques Springer Science & Business Media 12 211-26
[10] Daniel G 2013 Principles of artificial neural networks World Scientific 7 82-9
[11] Dagli C H and Poshyananda P 2012 Basic artificial neural network Artificial Neural Networks for Intelligent Manufacturing 39 29-59
[12] Suykens J A, Vandewalle J P and de Moor B L. 2012 Artificial neural networks for modelling and control of non-linear systems (Springer Science & Business Media)
[13] Yang S, Li W and Wang C 2008 The intelligent fault diagnosis of wind turbine gearbox based on artificial neural network 2008 International Conference on Condition Monitoring and Diagnosis 1327-1330
[14] Hammer M, Kozlovsky T, Szabo R and Svoboda J 2004 The use of neural networks for the life prediction of insulating material of electric rotating machines Proceedings of the 2004 IEEE International Conference on Solid Dielectrics 2 546-9
[15] Fink O, Zio E and Weidmann U 2014 Predicting component reliability and level of degradation with complex-valued neural networks Reliability Engineering & System Safety 121 198-206
[16] Bukhtoyarov V V, Tynchenko V S and Petrovsky E A 2019 Multi-Stage Intelligent System for Diagnostics of Pumping Equipment for Oil and Gas Industries *IOP Conference Intelligent Systems 272*(3) 032030

[17] Bukhtoyarov V, Semenkin E and Shabalov 2012 A Neural networks ensembles approach for simulation of solar arrays degradation process *International Conference on Hybrid Artificial Intelligence Systems* 186-95

[18] De Livera A M, Hyndman R J and Snyder R D 2011 Forecasting time series with complex seasonal patterns using exponential smoothing *Journal of the American Statistical Association* 106 (496) 1513-27

[19] Huang J, Li C and Yu J Resource prediction based on double exponential smoothing in cloud computing *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)* 2056-60

[20] Golyandina N and Zhigljavsky A 2013 *Singular Spectrum Analysis for time* (Springer Science & Business Media)

[21] Golyandina N, Korobeynikov A 2014 Basic singular spectrum analysis and forecasting with R. *Computational Statistics & Data Analysis* 71 934-54

[22] Hassani H, Soofi A S and Zhigljavsky A A 2010 Predicting daily exchange rate with singular spectrum analysis *Nonlinear Analysis: Real World Applications* 11(3) 2023-34

[23] Rocco S C M 2013 Singular spectrum analysis and forecasting of failure time series *Reliability Engineering and System Safety* 114 126-36

[24] Ma X, Jin Y and Dong Q 2017 A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting *Applied Soft Computing* 54 296-312

[25] Asuncion A and Newman D 2007 *UCI machine learning repository* (Irvine, CA, University of California)