ANÁLISE COMPARATIVA DOS MÉTODOS PARA SIMULAÇÃO DA OPERAÇÃO DE POÇOS COM INSTALAÇÕES DE BOMBA SUBMERSÍVEL ELÉTRICA

COMPARATIVE ANALYSIS OF METHODS FOR SIMULATING THE WELL OPERATION WITH ELECTRIC SUBMERSIBLE PUMP INSTALLATIONS

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МОДЕЛИРОВАНИЯ ПРОЦЕССА ЭКСПЛУАТАЦИИ СКВАЖИН С УСТАНОВКАМИ ЭЛЕКТРОЦЕНТРОБЕЖНЫХ НАСОСОВ

BUKHTOYAROV, Vladimir V.[1*]; TYNCHENKO, Vadim S.[2]; PETROVSKIY, Eduard A.[3]; Buryukin, Fedor A.[4]

[1,2,3] Department of Production Machinery and Equipment for Petroleum and Natural Gas Engineering, Siberian Federal University, Krasnoyarsk, Russia

[4]Department of Chemical Technology of Natural Energy Recourses and Natural Gas Engineering, Siberian Federal University, Krasnoyarsk, Russia

*Correspondence author
e-mail: bukhtoyarov.87@bk.ru

Received 12 June 2019; received in revised form 19 June 2019; accepted 19 June 2019

RESUMO

Este artigo apresenta os resultados da pesquisa de métodos paramétricos e não paramétricos de identificação dos modelos tecnológicos de operação de poços utilizando instalações de bombas submersíveis elétricas. Propõe-se a utilização de uma abordagem híbrida, combinando modelos paramétricos e não paramétricos para obter modelos precisos que permitam a previsão de parâmetros de desempenho de poços. Estudos de métodos de simulação sob condições de efeito de interferência de diferentes níveis, que são típicos para canais de sinalização de gerenciamento de dados reais, sistemas de controle e instrumentos de medição, foram conduzidos. Os modelos combinados propostos foram construídos com a ajuda da regressão não paramétrica de Rosenblatt-Parzen, modelos paramétricos com adaptação automática de parâmetros e redes neurais artificiais. Demonstrou-se que tais modelos combinados possuem possibilidades generalizadoras essenciais, permitindo a suavização de dados paramétricos e a restauração de dependências iniciais com um erro significativamente menor em relação à interferência perturbadora. Os métodos e modelos desenvolvidos foram implementados para fins de pesquisa no sistema de software, o que permite uma simulação complexa de mudanças nos parâmetros durante a operação do poço usando as instalações de bombas submersíveis elétricas. Para avaliar a significância estatística dos resultados, métodos de processamento estatístico foram aplicados usando ANOVA. Os resultados demonstram que para uma solução eficaz para o problema da simulação do processo de operação do poço e para garantir alta adaptabilidade dos modelos, a abordagem combinada é o método mais eficaz. Modelos com base em redes neurais artificiais após ajuste nos permitem melhorar a eficiência da solução para o problema de previsão e ao mesmo tempo ter flexibilidade necessária para adaptação da estrutura computacional sob as condições de mudança de parâmetros de desempenho. O bloco paramétrico de modelos nos permite utilizar informações a priori sobre dependências de parâmetros de desempenho e identificar razoavelmente precisas a deriva de parâmetros sob as condições de instabilidade do processo em estudo.

Palavras-chave: *regressão, rede neural, modelo combinado.*

ABSTRACT

This article presents the research results of parametric and non-parametric identification methods of the technological models of well operation using electric submersible pump installations. The use of a hybrid approach is proposed, combining parametric and non-parametric models to obtain accurate models that allow the prediction of well performance parameters. Studies of simulation methods under conditions of interference effect of different level, which are typical for signaling channels of real data management, control systems, and measuring instruments, have been conducted. The combined models proposed have been constructed with the help of the Rosenblatt–Parzen non-parametric regression, parametric models with automatic adaptation of parameters and artificial neural networks. Such combined models have been shown to possess essential generalizing possibilities, allowing for smoothing of parametrical data and the restoration of initial dependences with a significantly smaller error in relation to the disturbing interference. The developed methods and models were implemented for research purposes in the software system, which allows a complex simulation of changes in parameters during well operation using the electric submersible pump installations. To evaluate the results' statistical significance, methods of statistical processing have been applied using ANOVA. The results demonstrate that for an effective solution to the problem of the process simulation of well operation and to ensure high adaptability of the models, the combined approach is the most effective method. Models on the basis of artificial neural networks after adjustment allow us to improve the efficiency of the solution to the prediction problem and at the same time have necessary flexibility for adaptation of the computational structure under the conditions of changing performance parameters. The parametric block of models allows us to use a priori information about dependences of performance parameters and to identify reasonably accurate the drift of parameters under the conditions of instability of the process under study.

Keywords: *regression, neural network, combined model*

АННОТАЦИЯ

В статье представлены результаты исследования методов параметрической и непараметрической идентификации моделей технологического процесса эксплуатации скважин с использованием установок электроцентробежных насосов. В статье предлагается использовать комбинированный подход, сочетающий параметрические и непараметрические модели для получения высокоточных моделей, позволяющих осуществить прогнозирование эксплуатационных параметров скважины. Выполнены исследования методов моделирования в условиях действия помех различного уровня, характерных для каналов передачи сигналов реальных информационно-управляющих систем и средств измерений. Предлагаемые для использования комбинированные модели построены с использованием непараметрической регрессии Розенблата-Парзена, параметрических моделей с автоматической адаптацией параметров и искусственны нейронных сетей. Показано, что такие комбинированные модели обладают существенными обобщающими возможностями, позволяющими осуществлять сглаживание параметрической информации и восстанавливать исходные зависимости с существенно меньшей по отношению к действующим помехам ошибкой. Разработанные методы и модели в целях исследования были реализованы в программной системе, позволяющей проводить комплексную симуляцию изменения параметров при эксплуатации с установками электроцентробежных насосов. Для оценки статистической значимости результатов были использованы методы статистической обработки с использованием метода ANOVA.Результаты показывают, что для эффективного решения задачи моделирования технологического процесса эксплуатации скважин и для обеспечения высокой адаптивности моделей, наиболее эффективным вариантом оказывается именно комбинированный подход. Модели на основе искусственных нейронных сетей после настройки позволяют повысить эффективность решения задачи прогнозирования и при этом обладают необходимой гибкостью для адаптации вычислительной структуры в условиях изменения эксплуатационных параметров. Параметрический блок моделей позволяет использовать априорные сведение о зависимостях эксплуатационных параметров и достаточно точно определять дрейф параметров в условиях нестационарности исследуемого процесса.

Ключевые слова: *регрессия, нейронная сеть, комбинированная модель*.

1. INTRODUCTION

The modern period of science and technology development, and specifically in the oil and gas industry, requires intensive use of simulation methods for solving the problems of decision-making support during the operation of equipment. The efficiency of such systems is obviously influenced by a significant number of factors, among which is the accuracy and efficiency of computational models. In combination with effective models and methods for generation of possible solutions based on estimates of the current and predicted values obtained with the help of calculation models, the model approach will significantly improve the effectiveness of many production problem-solving systems.

Developing an effective basis of systems based on technological models is an important trend of modern solutions in terms of management and optimization of technological processes in the oil and gas industry. In this direction, the critical task is to build effective process models of oil and gas well operation in different modes. Within the framework of the term 'effectiveness,' the set of characteristics of models that can be used within the framework of the mathematical basis for decision-making support systems during the operation of equipment is considered. Such characteristics, besides the accuracy of the obtained models, include, in particular, the possibility of correction (adaptation) of models in use, computational efficiency in terms of compactness, simplicity of models, and the possibility of structural and parametric synthesis of models in an automated mode.

These characteristics, to a significant extent, determine the possibility of using the models in real well operation conditions. It is understood that the accuracy of the mathematical model, which is one of the most important characteristics, provides an opportunity to compare and use the results obtained with the help of the calculation model in a real-life object. Taking into account the specifics of the simulated system and the tasks to be solved on the basis of simulation during well operation, significant deviations of the model results from the parameters observed in the real well can lead to a significant decrease in the efficiency of operation as a result of incorrect assessment of the current state of parameters or incorrect prediction of the development of the production facility parametric trends. It gives evidence of the necessity of using high-precision identification methods for simulation of the processes and objects during well operation. Such methods should provide an acceptable level of deviation in the estimation of the parameters of simulated objects in conditions that characterize the performance of the production facility 'production well – electric submersible pump (ESP) installation'.

Operational conditions of such production facilities are characterized by the presence of factor combinations that significantly complicate the use of methods of direct physicomathematical simulation in an explicit form. Such factors include significant variability of object parameters, even if they are operated within the same group or field, and, moreover, in situations of significant type difference and geographical diversity. This makes it necessary to clarify the situation and revise the parametric and structural components of such models, which for traditional physicomathematical models requires significant efforts of expert analytical orientation. Another factor is the instability of basic trends in technological parameters, which requires when using the method of direct physicomathematical simulation, the use of special procedures for parametric optimization and adaptation of models to obtain reasonably accurate model results.

Thus, the requirements for the model adequacy of the production facility in the case of the 'producing well – ESP installation' model consideration directly agrees and evolves into the requirement to ensure the adaptability of models, as well as into requirements to provide the possibility of structural and parametric synthesis and adaptation of models in the automated mode.

For this purpose, the paper considers a set of methods that allows the implementation of the model structural and parametric synthesis in an automated mode. Taking into account the specifics of the simulated production facilities, the requirement to obtain accurate model results under the conditions of different level of interference is also important today.

2. METHODS

Definition of effective methods for automated generation of mathematical computational models for facilities of the oil and gas industry such as 'production well – ESP installation' is one of the most important research trends in the field of simulation of oil and gas industry facilities.

The majority of the available works are focused on the model building with the help of the expert analytical approach, which allows us to obtain physicomathematical models of the 'particular' orientation (Ahmadi *et al*., 2015; Lu *et al*., 2018; Su and Oliver, 2010). In most cases, such models are obtained by researchers in the course of adaptation of 'well – producing formation' fundamental physicomathematical models for particular production facilities and have a high value within the framework of the considered type of wells, group of wells, field or production region (Clarkson and Pedersen, 2011; Ramirez-

Jaramillo *et al.*, 2010; Sanchez-Rivera *et al.*, 2015).

From the point of view of the system approach it appears that in order to ensure the efficiency of solving technological tasks, such physicomathematical models should be supplemented, if not replaced, with a set of methods and algorithms that allow us to identify the operational facility in an automated mode. Such methods for building models of technological processes of well operation can be used as part of the analytical simulating link of decision support systems. An important aspect also seems to be the possibility of their application as a 'computational measuring device' in the situational analysis during the performance of technological operations, when it is necessary to deal with significant interference in the measurement channels, or it is impossible to obtain correct data with the help of physical measuring devices. In such cases, the technological process control, parameter definition of the technological equipment operation is reasonable to carry out on the basis of design parameters of the operational facility. Obtaining high-precision design parameters before the recovery of correct readings or performance of physical measuring devices is one of the important factors to ensure reliable, safe, and efficient operation.

It is obvious that the accuracy of the parameters calculated on the models can be ensured in case of using only such models, which have a high degree of adequacy to the operational facility. It appears that the features of the operational facilities such as 'well – ESP installation' determine the need for the use of models which make it possible the adaptation of the models taking into account the instability of the processes in the 'well – producing formation' system and changes in the characteristics of the operational equipment. It is also important that adaptation, structural or parametric adjustment of models is also carried out in the automated mode, being the actual 'built-in' possibility of the simulation method. It will allow the formation of a relatively autonomous analytical and simulation computational module, focused on a specific operational facility and its operating conditions. This property seems to be important, because through the use of the direct method of generating physicomathematical models by the expert analytical method, the variability and instability of the parameters of the operational objects, as well as their significant number, will lead to the need to perform a significant amount of analytical work and, consequently, to the cost escalation.

In this connection, a set of methods has been defined that can be used to build models for operational parameters calculation of such facilities as 'production well – ESP installation'. The methods and their description are presented below.

Methods of model parametric optimization. Parametric identification assumes the use of methods that allow the evaluation of the model parameters, given with accuracy to a set of parameters on the basis of sample data. Let the model of the object be given, which is described by Equation 1.

$$\tilde{y}_t \quad A < x_t, \lambda > \qquad \text{(Eq. 1)}$$

where $\tilde{y}_t$ is the model output, x_t are object input variables (known), λ are parameters of the accepted structure of the object model, which require definition, *A* is the operator, which converts input variables to output ones (known). The optimality criterion for their selection is Equation 2

$$R(\lambda) \quad M\{(y(x) - \tilde{y}(x,\lambda))^2\} \to \min_{\lambda} \qquad \text{(Eq. 2)}$$

where $y(x)$ are the output values of the simulation object (measured values), and $\tilde{y}(x,\lambda)$ are the output values of the model.

It is necessary to determine the coefficients $\lambda = (\lambda^1, ..., \lambda^m)$ of the object model accepted structure according to the available input and output variables sampling $V \quad \{x_i, y_i\},\ i \quad \overline{1,s}$, (Eq. 3)

This condition can be written in the following form:

$$R(\lambda) \quad M\{(y(x) - \tilde{y}(x,\lambda))^2\} \quad M\{Q(x,\vec{\lambda})\} \to \min_{\lambda}, \qquad \text{(Eq.4)}$$

where $Q(x,\vec{\lambda})$ is some convex quality function.

The solution to this problem can be found with the help of the following recurrent procedure:

$$\lambda^i_s \quad \lambda^i_{s-1} - \gamma_s \nabla_{\lambda^i} Q(x_s, \lambda^i_{s-1}), \qquad \text{(Eq. 5)}$$

Where s is the number of the observation pair in the sample, $s \quad \overline{1,N}$ where N is the sample size which is selected randomly;

$\nabla_{\alpha^i} Q(x_s, \lambda^i_{s-1})$ is the gradient of the function $Q(x_s, \lambda^i_{s-1})$ by a parameter λ^i;

γ_s is some number that meets Robbins-Monroe conditions:

1. $\gamma_s > 0, \forall s$, (Eq. 6);
2. $\gamma_s \neq 0, \forall s$, (Eq. 7);
3. $\lim\limits_{s \to \infty} \gamma_s \quad 0.$, (Eq. 8);
4. $\sum\limits_{S \quad 1}^{\infty} \gamma_S$ is a divergent numeric sequence;
5. $\sum\limits_{S \quad 1}^{\infty} \gamma_S^2$ is a convergent numeric sequence.

Let us note that $\vec{\lambda}^*$ is the exact solution, i.e., that $R(\vec{\lambda}^*) \quad \min\limits_{\vec{\lambda}} R(\vec{\lambda})$ exists, but it is never known in real problems.

For a recurrent procedure, the convergence theorem is proved:

$$\lim_{s \to \infty} M\{(\vec{\lambda}_s - \vec{\lambda}^*)^2\} = 0, \qquad \text{(Eq. 9)}$$

There are several algorithms which allow us to improve the efficiency of the solution to the problem of model parameters setting with the help of the recurrence procedure described above. The higher efficiency is understood as the finding of parameters closer to the optimal ones, and the higher degree of parameter approximation to the optimal parameters.

It is proposed to use the following parametric identification algorithms as optional ones in the operated experimental sample of the expert system:

Litvakov's algorithm (Rosenblatt, 1956):

1. Let us select some initial approximation $\vec{\lambda}_0$.

2. For t iterations, we get some estimation $\vec{\lambda}_t$, where the number of t is equal to the number of observations in the sample.

3. Let us start the recurrent procedure again, but for the initial approximation $\vec{\lambda}_0$ we take $\vec{\lambda}_t$ obtained at the previous step, and we get the estimation $\vec{\lambda}_{2t}$

M such cycles.

Usually, in practice, the number is $M \quad 3 \div 5$.

The theorem is proved for Litvakov's algorithm:

$$\lim_{M \to \infty} \vec{\lambda}_{Mt} \to \vec{\lambda}_t^{\text{îïò}}$$

Application of Litvakov's algorithm allows us to increase the accuracy (from the point of view of proximity to $\vec{\lambda}^{opt}$) of the obtained estimation $\vec{\lambda}_t$.

Kesten's algorithm (Parzen, 1962).

It is assumed that the initial approximation is taken relatively far from the minimum point $\vec{\lambda}^*$. As long as the gradient does not change the sign (moving in the required direction) we don't change γ^{λ}, and when the sign of the gradient changes, i.e., there will be a jump over the minimum point, we change $\gamma_s^{\lambda_i}$. And again, we leave it unchanged until the jump-over occurs again. This algorithm allows us to increase the speed of finding the minimum point because the speed of movement to it does not constantly decrease with the increase of the iteration number.

Litvakov's algorithm and Kesten's algorithm belong to one-step recurrent algorithms. These two algorithms can be used together.

Methods of nonparametric simulation. The regression is some average quantitative dependence between the object input and output (Ahmadi *et al.*, 2015). The problem of regression analysis can be formulated as follows. Suppose there is some object. There is an observational sampling of the object inputs and outputs $V \quad \{x_i, y_i\}, i = \overline{1,n}$. The function describing the dependence between the object input and output is unknown. It is necessary to restore the estimation of the function describing the dependence between the input and output of the object according to available observations of the object's input and output values. The nonparametric regression estimate (taking into account the estimate of Parzen–Rosenblatt probability density) is calculated as follows (Davis *et al.*, 2011; Demuth *et al.*, 2014; Szegedy *et al.*, 2013):

$$M\{Y \mid x\} \equiv \eta(x) \quad \sum_{i\ \ 1}^{n} K_N\left(\frac{x - x_i}{h}\right) y_i$$

,

$$K_N\left(\frac{x - x_i}{h}\right) \quad K\left(\frac{x - x_i}{h}\right) \Big/ \sum_{j\ \ 1}^{n} K\left(\frac{x - x_j}{h}\right);$$

where $K\left(\frac{x - x_i}{h}\right)$ is the truncated bell-shaped function.

Thus, it is possible to restore the dependence between the input and output of the object, using a training sample of the object observations for the construction of a non-parametric estimator.

When using the quadratic criterion of the best correspondence between the estimation and the true dependence, it was found that the optimal form of the bell-shaped kernel is the parabolic kernel:

$$K(z) \quad \begin{cases} 0.75(1 - z^2), & |z| \le 1 \\ 0, & 1 \le |z| \end{cases},$$

The optimal value of the blur parameter $h(n)$ is found from the ratio $h(n) \quad cn^{-1/5}$, where n is the sample volume and c is the positive constant. Exactly the constant c has the greatest influence on the quality function and determines the blur coefficient. It is calculated on the basis of the sample by minimizing the quality indicators that characterize the best smoothing of experimental data.

Apart from the use of the parabolic kernel, it is possible to use the triangular kernel:

$$K(z) \quad \begin{cases} (1 - |z|), & |z| \le 1 \\ 0, & 1 \le |z| \end{cases}, \qquad \text{(Eq. 10)}$$

A quality criterion based on the use of a sliding examination' is used to adjust the blur parameter $h(n)$:

$$I_{2n} \quad \frac{1}{n}\sum_{i\ \ 1}^{n}(y_i - \overline{\eta}_n(x_i))^2 \quad \min, \qquad \overline{\eta}_n(x_i) \quad \sum_{\substack{k\ \ 1 \\ k \ne i}}^{n} K_N\left(\frac{x - x_i}{h}\right), \qquad \text{(Eq. 11)}$$

The peculiarity of this criterion is that the examination point (x_i, y_i) is not involved in the construction (training) of the model $\overline{\eta}(x_i)$.

As was already mentioned above, the optimal, in terms of minimizing the quality indicator, the value of the blur parameter is determined by the following formula

$$h(n) \quad cn^{-1/5}, \qquad \text{(Eq. 12)}$$

The choice of the optimal value of the constant c is significant complexity.

The optimal value of the constant can be found by the method of the golden section. The boundaries of the search interval $[a;b]$ were defined as follows:

$$a \quad \Delta \cdot n^{1/5},$$
$$\Delta \quad x_{i+1} - x_i, (i \quad \overline{1, n-1}), \qquad \text{(Eq. 13)}$$

$b \quad 4$ is the empirical value (determined by multiple observation of the optimal value of the constant c at various values of b).

Models based on artificial neural networks. Artificial neural networks are computational structures that simulate biological processes, usually those taking place in the human brain (Gurney, 2014). They represent distributed and parallel systems capable of adaptive learning. An artificial neuron, named by the analogy with a biological prototype, is used as an elementary transducer in such networks.

The neural networks have a number of advantages over other methods of data simulation and analysis due to their ability to generalize, i.e., produce effective results based on data that was not used during training. They are also highly efficient as a result of internal data parallelization, nonlinearity, and adaptability (Curteanu *et al.*, 2011; Osovsky, 2020). In sum, artificial neural networks are a powerful simulation and data analysis tool and can be used to solve complex, large-scale problems that are difficult or impossible to solve by other methods.

While there are many topologies of the artificial neural networks, about 80% of all real-life applications are based on multilayer fully connected networks of direct distribution — multilayer perceptrons (Bukhtoyarov *et al.*, 2010;, Hansen *et al.*, 1990). This network type was also adopted in the course of the studies described in this article. For the structural synthesis of neural network models, the probabilistic evolutionary method was used (Gutta, 1996).

The ability of neural networks to learn from environmental data is perhaps their most important and useful feature. Artificial neural networks learning is the process during which free parameters characterizing the network are adjusted by simulating the environment the network aims to represent. The learning type is determined by the way these parameters are adjusted (Hansen *et al*., 1990). Typically, neural network training involves an iterative process as a part of which weight coefficients are determined. The evolutionary genetic algorithm, which was proven particularly efficient in solving such problems, was used in the present study to adjust the weight coefficients.

Combined models. Combined models represent a defined set of calculation models of different (sometimes the same) types, which are used together to solve one problem. For the first time, the idea of combining separate computational models (such a combination can be called an 'ensemble of models' or a 'team') was proposed in work (Folino *et al*., 2016). The combined model approach was subsequently developed further and was successfully used to solve a wide range of practical problems, such as recognition tasks (Yu *et al*., 2011; Akhand *et al*., 2012), medical diagnostics (Breiman, 1996), classification of seismic signals (Schapire, 1990) and many others.

Structural and parametric synthesis methods, which are commonly adopted when developing individual models, are supplemented in a combined approach by specialized procedures that allow the efficiency of solving a number of problems to be improved.

The most common approaches are bagging and boosting. Bagging or bootstrap aggregating was first proposed by Briemann and is based on random sampling with replacement (bootstrap) (Mishchenko *et al*., 2011). In this context, the term 'bagging' pertains to the fact that each model is built using a separate sample, denoted as a booster sample, drawn from the initial set of observations.

Boosting, developed by R. Shapire (Maltsev, 2013), involves forming each subsequent model in the set using a sample that includes those examples on which previous models gave significantly different results from the target one. Due to the sequential nature of the processing, the task for each subsequent model becomes more difficult. In the present study, the combined models (neural network model – nonparametric model – parametric model) were developed using this approach.

Numerical studies. To carry out numerical studies, a software system "Program for simulation of oil production conditions by electric submersible pump installations" was developed. In this software system, physicomathematical calculation models of 'production well − ESP installation' object parameters have been implemented. The physical and mathematical models incorporated in the program were verified on the basis of operational data obtained from 10 oil and gas production facilities owned by one of the oil and gas companies. The physicomathematical models implemented in the software system are based on the solution obtained by taking into account pertinent results reported in extant literature (Asuncion *et al.*, 2007). It is important to note that all appropriate models were developed using an expert analytical approach without the possibility of automating the relevant procedures, which limits the potential for their generalization to other simulated facilities.

The verified model solutions adopted in this work were applied to obtain the initial dataset for numerical studies of the simulation methods considered in the article. In order to simulate real systems, in which the measuring channels are affected by different interference levels, initial datasets with different levels of overlaid interference on the recorded parameters pertaining to the operational facilities were formed. For this purpose, additive interference dependent on the signal level in the simulated measuring channel was considered. The analyzed levels of interference had values of 0% (no interference), 5%, 10% and 25%.

In order to study the structural synthesis methods, and parametric parameter, adjustment approaches, observation samples of the 'production well − ESP installation' object operational parameters were obtained. In order to study the possibility of adapting the models, multiplicative superimposition of trends (increasing and decreasing) sourced from the ControlChartDataSeries set was used and was allocated in the dataset repository for data analysis algorithms. This set is widely used to evaluate methods and algorithms for the ability to capture the drift of parameters and determine the instability of objects.

As the main criterion of efficiency the estimation of mathematical expectation and estimation of the error dispersion of the simulation, calculated on the basis of the data obtained during 50 independent startups of the algorithms, were used. Thefollowingformulawasusedtocalculatetheapproximationerrorineachstartup:

$$Error = \frac{100\%}{s\left(y^{\max}-y^{\min}\right)}\sum_{i=1}^{s}\left|o_i-y_i\right| , \qquad \text{(Eq. 14)}$$

Here i is the number of the record in the sample, o_i is the calculated value obtained on the model, y_i is the value of the output variable in the sample, $y^{\max}$ and $y^{\min}$ are the maximum and minimum values of the output variable, s is the number of elements in the sample.

For all methods, in order to obtain correct results of the numerical experiments, the same amount of computational resources was used. The volume of observation sampling used for building and adjustment of models was defined equal to 500, 5000, and 100000 parameter measurements for each of the approaches. The results averaged over such samples are given in the results section.

ANOVA methods were used to study the significance of differences in the effectiveness of the approaches used. A pair-wise comparison of the methods studied was performed in order to identify the statistical significance in the distinguishability of the results obtained in the course of the numerical study at significance level $\alpha = 0{,}05$.

3. RESULTS AND DISCUSSION:

The following are the results of the method study. The table presents the evaluation results of the models for determining the following parameters of the operational facility, which were calculated with the help of computational models at different levels of interference: pressure P, level H in the well, flow rate Q.

The obtained results give evidence of the applicability of the considered methods for building effective computational models for estimation of operational parameters of such objects as ‘production well – ESP installation’. Let us analyze the obtained results in relation to the specified criteria for the efficiency of the models for the use in real well operation supporting systems with electric submersible pumps installations. Such criteria, which have been proved earlier, include:

- adequacy of the model characterizing the degree of closeness of the model results and the results of the real production facility;

- possibility of adaptation (structural or parametric) of the obtained computational model;

- stability of the model to the presence of interference in the measuring channels in forming the initial information for the procedures of structural and parametric synthesis.

Generally, the methods of parametric adaptation considered in the article have shown sufficiently high efficiency in the cases of a model building under the action of small interference in measuring channels. Adequacy of such models, for the most part, is determined by the initially defined structure of models of analyzed dependences. If there is a sufficiently large amount of a priori information about the simulated operational facility and its features, the structure can be synthesized by an expert analytical method. However, in practice, this is an extremely difficult task, usually solved with a high degree of generalization by research teams. Such an approach has a significantly limited applicability in real conditions when the stock of operational facilities reaches hundreds and thousands of ‘producing formation –well – pump’ systems. Taking into account the significant differences in the characteristics of such systems, the generalized structure of the model may be unreliable for a number of simulated facilities, and, consequently, the methods of parametric identification will not be able to achieve the required accuracy indicators. A separate factor complicating the application of such an approach is the need to determine a totality of algorithm settings for adjustment of parameters, which have a significant impact on their effectiveness.

Determination of such parameters in the course of the study was performed at a preliminary stage during test startups, which in the simulation conditions of real objects have limited effectiveness due to, for example, the presence of errors in the data measurement for adjustment. Based on the results of the conducted research, it is also possible to judge about the significant sensitivity of parametric adaptation methods to the increase in the level of interference overlaid on the simulated parameters. This makes it difficult to use such methods in the building models for use as a ‘computational measuring device’. Thus, this method obviously appears to be effective in case of a successful choice of the initial structure, but in real production situations, taking into account the variability of simulated objects, such a choice is significantly complicated. The flexibility is significantly restricted by the choice of the initial structure, which has a negative impact on the results of model adaptation in case of incorrect definition of such a structure, or in case of structural evolution of the simulated system, for example, in case of qualitative changes in well operation conditions or change in equipment composition.

The method appears to be sensitive to the interference, taking into account the need for the correct choice of parameters for the corresponding algorithms of adjusting the model's coefficients, which requires expert participation and complicates the possibility of its automated use.

In contrast to the methods of the parametric setting of regression models, the method of nonparametric regression does not require the presence of a priori assumptions about the structure of simulated dependencies for the operational facility. This makes it possible to build such models for a wide range of facilities using automated computational procedures based on a totality of observation sets. In this connection, such models can be used to build high-precision models of such objects as well – ESP installation’, as evidenced by the results obtained. A significant positive aspect of using such regression models is their resistance to interference in the samples of measurements (observations) in case of correct adjustment of the blur parameter, which determines the degree of smoothing through the use of bell-shaped functions. On the other hand, exactly the necessity for correct adjustment of the blur parameter is one of the factors making it difficult to use such models in practice. It is obvious that in case of restoring multidimensional dependences, the blur parameter should be defined and optimized with respect to each input variable because the range of variation of input variables and discretization of measurements can be significantly different. This requires the use of special optimization procedures, the implementation of which is an additional factor complicating the use of such an approach. In addition, it is obvious that through the use of nonparametric regression models it is necessary to ensure a compromise between the capability of the model to smooth out perturbations, the accuracy in the reproduction of simulated dependencies and the possibility of effective adaptation of the model .In the case of the algorithmic solution of such a problem, nonparametric models appear to be quite effective according to the three criteria indicated above. It is important to note that their value is somewhat reduced by their relatively high computational complexity, which is conditioned by the need for calculations using a complete sample of observations. Besides that, such models are actually ‘black’ box models that don't allow us to receive analytical data about the restored dependences. As can be seen from the above, it is expedient to use such approach in the absence of the a priori information about the dependencies, interference effect and provision of computational capabilities allowing to process data of considerable volumes.

In the case of using artificial neural networks as a simulation method, the necessity of constant computational processing of initial data, typical for the nonparametric approach, disappears. The neural network model can be built using computer-aided procedures on the basis of available data about the parameters of the operational facility ‘well – ESP installation’, or on the basis of data obtained during the operation of a similar facility. It seems important that the models based on the neural networks allow us to organize the procedure of additional training (additional adjustment), which will allow the initial model for the specific simulated object to be adapted. Thus, the models based on the artificial neural networks allow us to obtain sufficiently effective computing procedures for simulation and forecasting of the technological process parameters of well operation with the help of ESP installations. As the analysis of the results of the numerical studies indicates, the neural network model makes it possible to ensure the efficiency compared with the efficiency of nonparametric models, and specifically, in conditions of interference with the values of simulated parameters of the operational facility. At the same time, the resulting computational procedure appears to be more computationally effective, because in order to obtain individual predicted values it is not necessary to carry out the computational operation fully involving the available samples of observations. Prime computational costs in case of using neural networks are typical for the stage of the model building, namely the synthesis of the models structure and setting of parameters (of training). This task is usually solved at the preliminary stage, and then the neural network provides fast enough calculations in the mode of real operation. If additional training is necessary, only the task of parametric adaptation is usually solved, which, in the typical case of small amounts of new data on the functioning of the simulation object is not computationally capacious. Methods of structural synthesis of the neural network models and their settings are well developed and have a high degree of automation. In this work, a probabilistic approach to structure generation was used to construct neural network structures, and an effective genetic algorithm was used to train neural networks [Hansen, 1990; Gutta, 1996]. The disadvantage of the neural network models, in general, is their non-transparency that means the impossibility of restoring the formula and structure of the model in the form which is explicit and convenient for further analysis. Nevertheless, taking into account the computational efficiency, accommodation capability and stability of solutions in relation to interference in the samples of parameters, along with a high degree of automation of all stages of building and use, based on the research findings, the neural network models can be considered as one of the most effective solutions for simulating the operational parameters of the facilities like 'production well – ESP installation'.

A combined approach can be considered as an even more effective solution in conjunction with the three indicated performance criteria. The combined model allows the synergetic interaction of models of certain types to be provided. Previously, the effectiveness of such a solution was repeatedly shown through the use of the models of the same type, for example, the neural network models. In this study, we have considered an approach that uses different types of models. The results of the numerical studies demonstrate that the accuracy of the obtained solutions of the combined model appears to be statistically higher under different conditions of experiments (different levels of interference and types of trends that determine the drift of the object parameters). From the point of view of ensuring the approach correctness, as one of the criteria, ensuring equality of conditions for comparing separate methods and combined approaches, the principle of equivalence of used computing resources (processor time consumption) in the data processing of a specified volume was implemented through the studies. Corresponding parameters were selected in advance in the course of preliminary startups of the algorithms and methods. For the combined approach, restriction to the total amount of computing resources available to all combined approaches were formed.

As can be seen from the above, the study has identified the characteristics of the simulation methods, their advantages, and disadvantages. It has been shown that on retention of the computational resources equivalency, the combined approach was the most effective.

4. CONCLUSIONS:

The article considers the problem-solving method of structural and parametric identification of models that allow the estimation of the current values and predict operational parameters of such facilities as 'production well – ESP installation'. It is proposed to use a set of simulation methods based on parametric optimization methods, the nonparametric regression method, and the artificial neural networks method to solve such a problem. The corresponding algorithms have been implemented in the software system, providing the opportunity to study their properties on the simulation data of real wells, operated with the use of ESP installations.

Numerical studies of the considered methods in the conditions of variation of well operational parameters, parameters of overlaid interference under the conditions of physical measuring instruments operation and various trends of the drift of simulated parameters were carried out. It has been shown that taking into account the criteria for such models, it is reasonable to use combined models built on the basis of a model symbiosis of the types under consideration. Efficiency, particularly in terms of the accuracy criterion under different conditions of the experiment, of such combined models built with the use of the bagging approach has appeared to be higher than for models built using each of the methods separately. Thus, the results of the numerical studies demonstrate that for the effective solution to the problem of the process simulation of well operation and to ensure high adaptability of models, the most effective option appears to be exactly the combined approach. In this case, the non-parametric component of these combined models ensures high accuracy of the results in the mode of computational ‘measuring device’. Models on the basis of the artificial neural networks after adjustment allow us to improve the efficiency of the solution to the prediction problem and at the same time have the necessary flexibility for adaptation of the computational structure under the conditions of changing performance parameters.The parametric block of models allows the use a priori information about dependences of performance parameters and to identify reasonably accurate the drift of parameters under the conditions of instability of the process under study.

In future, the methods and approaches considered in the article are supposed to be used to build models of operational oil and gas production facilities, defined by the system ‘production well – autonomous packer – ESP installation’. Such a system appears to be more complex for simulation and, obviously, will require further development of direct physical and mathematical simulation approaches and simulation methods on the basis of the combined models studied in this paper.

5. ACKNOWLEDGMENTS

The work was completed with financial support from the Ministry of Education and Science of the Russian Federation within the framework of the Federal Target Program "Research and development in priority areas of scientific and technological complex of Russia for 2014 – 2020", action 1.3. Unique identifier of applied researches (project) RFMEFI57817X0236.

5. REFERENCES:

1. Ahmadi, M. A.; Soleimani, R.; Lee, M., Kashiwao, T.; Bahadori, A.; Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool, *Petroleum*, **2015**, 1(2), 118-132.
2. Lu, J.; Shi, S.; Rahman, M. M.; New Mathematical Models for Production Performance of a Well Producing at Constant Bottomhole Pressure. *Special Topics & Reviewsin Porous Media: AnInternational Journal*, **2018**, 9(3).
3. Su, H. J.; Oliver, D. S.; Smart well production optimization using an ensemble-based method. *SPE Reservoir Evaluation & Engineering*, **2010**, 13(06), 884-892.
4. Clarkson, C. R.; Pedersen, P. K.; Production analysis of Western Canadian unconventional light oil plays. In Canadian Unconventional Resources Conference. Society of Petroleum Engineers, **2011**, doi: 10.2118/149005-MS.
5. Ramirez-Jaramillo, E.; del Rio, J. M.; Manero, O.; Lira-Galeana, C.; Effect of deposition geometry on multiphase flow of wells producing asphaltenic and waxy oil mixtures. *Industrial & Engineering Chemistry Research*, **2010**, 49(7), 3391-3402.
6. Sanchez-Rivera, D.; Mohanty, K.; Balhoff, M.; Reservoir simulation and optimization of Huff-and-Puff operations in the Bakken Shale. *Fuel*, **2015**, 147, 82-94.
7. Rosenblatt, M.; Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, **1956**, 832-837.
8. Parzen, E.; On estimation of a probability density function and mode, *The Annals of Mathematical Statistics*, **1962**, 33(3), 1065-1076.
9. Davis, R. A.; Lii, K. S.; Politis, D. N.; Remarks on some nonparametric estimates of a density function. *In Selected Works of Murray Rosenblatt . Springer, New York, NY*, **2011**, 95-100
10. Demuth, H. B.; Beale, M. H.; De Jess, O.; Hagan, M. T.; *Neural network design*, **2014**, 15-28.
11. Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R.; *Intriguing properties of neural networks*, **2013**, arXivpreprint arXiv:1312.6199.
12. Gurney, K.; *An introduction to neural networks*. CRC press, 2014, 317.
13. Curteanu, S.; Cartwright, H.,; Neural networks applied in chemistry. I. Determination of the optimal topology of multilayer perceptron neural

networks. *Journal of Chemometrics*, 2011, 25(10), 527-549.
14. Osovsky S.; Neural networks for information data processing, *Finance and Statistics*; **2002**, 344.
15. Bukhtoyarov, V.V.; Semenkina, O.E.; 2010. Comprehensive evolutionary approach for neural network ensemble automatic design, *IEEE*, 2010, doi: 10.1109/CEC.2010.5586516, 1–6.
16. Hansen, L.K.; Neural network ensembles, *IEEE Trans. Pattern Analysis and Machine Intelligence,* **1990**, 12 (10), 993-1001.
17. Gutta, S.; Face recognition using hybrid classifier systems, *IEEE Computer Society Press*, **1996**, 1017-1022.
18. Folino, G.; &Sabatino, P.; Ensemble based collaborative and distributed intrusion detection systems: A survey. *Journal of Network and Computer Applications*, **2016**, 66, 1-16.
19. Yu, Z., Wongb, H. S., You, J., Yang, Q., & Liao, H.; Knowledge-based cluster ensemble for cancer discovery from biomolecular data, *IEEE Transactions on Nanobioscience*, **2011**, 10(2), 76-85.
20. Akhand, M. A. H.; & Murase, K.; Ensembles of neural networks based on the alteration of input feature values, *International Journal of Neural Systems*, **2012**, 22(01), 77-87.
21. Breiman, L.; Bagging predictors, *Machine Learning,* **1996**, 24 (2), 123-140.
22. Schapire, R.E.; The strength of weak learnability, *Machine Learning,* **1990**, 5, 197-227.
23. Mishchenko I.T.; Leonov I.V.; Fundamentals of physical and mathematical model of the system ‘production facility – production well – ESP installation’. *Bulletin of the Association of Drilling Contractors*, **2011,** (3), 36-40.
24. Maltsev N.V.; Development of a physical and mathematical model of the well development process with the help of ESP installations, Doctor thesis, 2013.
25. Asuncion, A.; Newman D.J.; UCI Machine Learning Repository, *Irvine, CA: University of California, School of Information and Computer Science,* **2007**, http://www.ics.uci.edu/~mlearn/MLRepository.html, accessed May 2019

Table 1*. Evaluation results of the models*

Simulation method	**Interference level, %**											
	0			**5**			**10**			**25**		
	P	**H**	**Q**	**P**	**H**	**Q**	**P**	**H**	**Q**	**P**	**H**	**Q**
	Average simulation error, %											
Model with parametric optimization	5.5	4.4	5.2	8.1	7.4	8.8	11.2	11.9	12.2	19.7	19.4	19.8
Non-parametric regression model	6.1	6.8	6.5	7.7	8.1	8.6	8.6	9.6	8.7	10.4	10.3	10.7
Neural network model	6.7	7.1	6.9	7.6	8.0	7.6	9.1	9.7	9.7	12.7	12.5	11.5
Combined models	3.9	2.7	4.4	4.9	5.7	5.1	6.2	7.2	6.9	6.2	8.9	6.9