

## IDENTIFICATION OF HYDRAULIC RESISTANCE PARAMETERS IN HYDRAULIC NETWORK MODEL

Vladimir Bukhtoyarov<sup>\*1</sup> Evgeny Agafonov<sup>1</sup> Nikita Antropov<sup>2</sup> Vadim Tynchenko<sup>1</sup> Valeriya Tynchenko<sup>1</sup>

<sup>1</sup>Siberian Federal University, Krasnoyarsk, Russian Federation

<sup>2</sup>Siberian State University of Science and Technology, Krasnoyarsk, Russian Federation

*The main purpose of the work is to develop and analyze new identification algorithm for multivariate multi-connected implicit system. Computational experiments and nonparametric statistics are implemented. The result of the work is a qualitatively new identification and output forecasting algorithm for multivariate systems. Based on the proposed algorithm a new class of generalized multivariate implicit nonparametric models is suggested. The significance of the research is subject to the fact that the majority of complex technical objects is usually described by large-scale nonlinear systems of equations up to input and output variables. Even if random effects in such models could be neglected a procedure of numerical solving the equation systems were either unstable or it would last too long to be implemented in forecasting of continuous process in the controlled object. Any delay in model calculation can also impact negatively on real-time control procedure. Using the proposed approach one can substitute solving the equation by estimation of the solution based on sample including measurements of input and output variables. Moreover, compared to counterparts, using the proposed approach it is possible to develop a single model for a range of input effects as it is presented in the research. Modelling and identification of multivariate systems is one of the most urgent tasks. Nowadays the development modelling methods does not keep up with the continuous and unlimited increase in the volume of information. To solve this problem, the authors proposed a new approach to constructing models of multidimensional systems. The scientific novelty of the article is as follows. A new modification of the nonparametric algorithm for identification of multidimensional systems has been proposed and tested. It makes possible to improve the accuracy and speed of computations.*

*Key words: Forecasting, Hydraulic network, Hydraulic resistance parameters, Identification, Multivariate system*

### INTRODUCTION

At present there are a great number of identification methods and algorithms applicable to complex system modelling. Applicability of the methods depends on type of the system, intended objective and problem statement, limitations, operating conditions, information on system's structure, effects and responses etc. These facts are reflected in preliminary (a priori) and current information of the system.

The most general identification problem statement is as follows. Let a system be influenced by both unobservable random effects having zero mathematical expectations and limited variance, and a set of observable and controllable effects. Responses of the system are synchronically measured together with the observable effects. Measurements are performed incrementally with a certain time interval. We assume additive nature of random effects in the measurements. An operator of the system establishing relationships between observable inputs and outputs should be estimated. The operator should satisfy certain quality criterion.

In addition, one can assume the case, when some output variables depend not only on input effects, but also on the rest of outputs. This is an equivalent for an implicit functional relation between corresponding variables. Such case is frequently encountered in practice and it

is common for continuous production systems with complex branching structure. For example, it is common with technological processes where outputs of some production steps are at the same time inputs for the others. To identify this kind of multi-connected systems it is necessary to use special implicit modelling methods. This fact confirms relevance of the research. Here we deal with an additional requirement: in order to develop a universal model one involves measured data for the whole range of input values. It means one-time training process should cover the whole range of values and provide good quality of predictive modelling.

The proposed algorithm is tested and approved for the problem of hydraulic network identification. Hydraulic network is a prominent example of multi-connected system of the above mentioned type. It usually consists of a number of interconnected pipes together with pumps and other equipment. The distinctive features of the system are large-scale nature, nonlinearity, multiconnectedness, complex topology, influence of diverse random factors. Control and technological parameters forecasting problems are essential for hydraulic networks and require explicit mathematical modelling. Developing and implementing effective modelling procedures can ensure more reliable control process over water and oil pipelines functioning. It could be an additional reason why the results of the work are relevant from the practical point of view.

\* Siberian Federal University, 82 Svobodny Ave., bdg 6, Krasnoyarsk, 660041, Russia, bukhtoyarov.87@bk.ru

## LITERATURE REVIEW

There are three basic approaches to identification. The first one is parametric approach [01]. The substance of the approach is that preliminarily some assumptions about the object structure based on the available information research should be made. The assumptions determine structure of the model up to parameters. This step is followed by parameters tuning within the accepted structure [02]. Parametric models are convenient and, in most cases, easy to tune. Reasonable disadvantage of the approach is complexity of model structure defining, which cannot be explicitly formalized.

The second approach is nonparametric. It involves usage of kernel estimation [03] of functional from random values. A plenty of multivariate regression estimates have been already developed [04]. Nonparametric approach does not require any assumptions about the model structure. Modelling in this case is entirely based on information contained in data sets. One disadvantage of the approach is significant reliance on amount of data in training samples. Lack of data can cause imprecise and biased estimation, and, as a result, limited quality of the corresponding models. Nevertheless the nonparametric algorithms are capable to deal with various problems in many spheres, for example economics [05], medicine [06], biology [07] etc.

The third approach is called hybrid or combined, and it incorporates modelling by using combinations of both parametric and nonparametric methods [08, 09]. This approach inherits advantages and disadvantages of the approaches mentioned above. The ratio between pros and cons depends of contribution of each component, namely, parametric and nonparametric. Hybrid models are rather hard to be constructed. Nevertheless, they represent a flexible solution exceeding capabilities of the rest two methods in terms of modelling accuracy.

Nowadays a pressing problem is identification of multivariate systems, displaying situation when input and output variables are interconnected functionally or stochastically. This case is very likely to be encountered in practice [10-11]. Model in the case of interconnected variables can be represented by an implicit system of equations, which is often appearing to be undefined. The processes in such systems have been recently named T-processes [12, 13]. Modelling and control in systems of the described nature are non-trivial problems, and in order to be solved they require properly designed methods and algorithms.

Construction of parametric models for the T-processes is a task with great computational difficulties: to identify output for each new input it is necessary to solve a system of implicit equations. This imposes a significant constraints on their practical application in decision-making and control tasks. Control systems for processes with implicit functional and stochastic relations between variables could only be Model-Based Control (MBC) systems providing model to transfers the control effects to

the process itself when the necessary output is attained [02]. In this regard, the T-process models are research subject to high demands, both accuracy and speed of computation terms.

In this situation, nonparametric methods for modelling of multivariate systems have become widespread and progressive [14-16]. One well-developed method of the interconnected systems modelling is the response prognosis algorithm [13] making possible identification of multivariate systems represented by the set of implicit operators under combined a priori information. Further development of the nonparametric approach to the identification of the T-processes is related to the identification problem under nonparametric uncertainty, when information about the operator structure is unknown [17]. Except nonparametric methods, in many cases sufficiently qualitative models are constructed by ensemble methods [18] and decision-making trees [19] in connection with their adaptive and flexible quantities.

Hydraulic network is a remarkable example of multivariate multi-connected systems. Scientific and specialized literature proposes various models of hydraulic networks. Some of the sources are devoted to methodology of developing dynamic models, for example [20, 21]. In most cases models are within the concept of the Computational Fluid Dynamics (CFD). Equations expressed in the form of partial derivatives, partial volumes or finite elements are widely used for distributed systems description [22-24]. The CFD approach requires quite complete information on the physical characteristics of the pumped liquid, nature of its flow, internal profile and geometric configuration of pipeline, and explicit information on functioning of pumps. In most cases, such models are computationally complex and they require unknown, often immeasurable or unpredictable in time and space variables.

For processes in complex hydraulic networks the CFD models are rather inapplicable. In this regard, steady state stochastic multivariate models with improved features are still of interest when dealing with hydraulic network modelling.

## MATERIALS AND METHODS

Denote by  $x = (x_1, \dots, x_m)$  the vector of input effects of the system,  $y = (y_1, \dots, y_l)$  - the vector of output responses,  $\{x[t], y[t], t = \overline{1, s}\}$  - the data set of statistically independent observations of system's state  $\{x, y\}$  at discrete time moments  $t = \overline{1, s}$ , where  $s$  is sample size;  $F(x, y, \alpha) = f_j(x^{(j)}, y^{(j)}, \alpha_{1j}, \dots, \alpha_{pj}), j = \overline{1, l}$  - the implicit equation system describing topology of the multivariate multi-connected object under research, where  $f_j$  are functions known up to their parameters;  $\alpha_j$  - the vectors of the reduced size with components of  $\alpha_j$  appearing in  $j$ -th equation;  $\alpha_j$  - the vector of random uncontrollable effects influencing respectively input and output variables of the system.

Let identification problem statement be as follows. A multivariate inertia-free system is influenced by observable input effects  $x$  and unobservable random effects  $\xi[t]$  having zero mathematical expectation  $M\{\xi[t]\} = 0$  and limited variance  $D\{\xi[t]\} < const.$  Observations of input and output variables  $\{x[t], y[t], t = \overline{1, s}\}$  are assumed to be available. Functional dependence of input and output variables has an implicit form which is common for description of multi-connected system, and it is described by the equation system  $F(x[t], y[t], \alpha)$  up to parameters  $\alpha_{1j}, \dots, \alpha_{pj}$ . Based on the available data set with observations of effects and responses  $\{x[t], y[t], t = \overline{1, s}\}$  and the implicit model with the known structure  $F(x[t], y[t], \alpha) = 0$  one should develop model of the system in the form  $y[t] = A(x[t], \xi[t], t)$  where  $A$  is an operator describing relations between inputs and outputs in an explicit way. Structure of the operator  $A$  is unknown. This operator should estimate output response  $y$  for an arbitrary input effect  $x$ , i.e. should represent a tool for forecasting of system's response.

**ALGORITHM OF MULTIVARIATE SYSTEM IDENTIFICATION**

The proposed algorithm starts with a preliminary step, at which parameters  $\alpha_{1j}, \dots, \alpha_{pj}$  of the system  $F(x[t], y[t], \alpha) = 0$  are found using the training data set  $\{x[t], y[t], t = \overline{1, s}\}$ . To adjust the parameters one can apply any suitable procedure of parametric identification, for instance

$$\begin{aligned} \alpha_j[t] &= \alpha_j[t-1] - \gamma_i^{(j)}[t] \cdot f_j(x^{(j)}[t]), \\ \gamma_i^{(j)}[t], \alpha^{(j)}[t-1] &_{j = \overline{1, l}, i = \overline{1, p}, t = \overline{1, s}} \end{aligned} \tag{1}$$

where  $\gamma$  is a coefficient satisfying Robbins-Monroe requirements [19].

Further, we assume three main steps. At the first step residual matrix is calculated:

$$\begin{aligned} \varepsilon_j[k, t] &= f_j(x^{(j)}[k], y^{(j)}[t]), \\ j &= \overline{1, l}, k, t = \overline{1, s} \end{aligned} \tag{2}$$

At the second step for each t-th observation of residual matrix solution of the system is estimated:

$$\begin{aligned} y_j^s[t] &= \sum_{t=1}^s y_j[t] \prod_{j \in N_\varepsilon} \Phi \left( \frac{0 - \varepsilon_j[k, t]}{c_{jk}^\varepsilon(s)} \right) / \\ & / \sum_{t=1}^s \prod_{j \in N_\varepsilon} \Phi \left( \frac{0 - \varepsilon_j[k, t]}{c_{jk}^\varepsilon(s)} \right), j = \overline{1, l}, k, t = \overline{1, s} \end{aligned} \tag{3}$$

where bandwidth parameters  $c_{jk}^\varepsilon(s)$  are optimized according to the cross-validation procedure using residuals sample  $\{\varepsilon[k, t], y[t], k, t = \overline{1, s}\}$ :

$$\begin{aligned} c_{jk}^\varepsilon(s) &= \frac{1}{s} \sum_{t=1}^s \left[ y_j[t] - \frac{\sum_{v \neq t} y_j[v] \prod_{j \in N_\varepsilon} \Phi \left( \frac{\varepsilon_j[k, t] - \varepsilon_j[k, v]}{c(s)} \right)}{\sum_{v \neq t} \prod_{j \in N_\varepsilon} \Phi \left( \frac{\varepsilon_j[k, t] - \varepsilon_j[k, v]}{c(s)} \right)} \right]^2 \\ & \rightarrow \min_{c(s)}, j = \overline{1, l}, k = \overline{1, s} \end{aligned} \tag{4}$$

The third step involves nonparametric regression estimation using the data set  $\{x[t], y^s[t], t = \overline{1, s}\}$ :

$$\begin{aligned} y_j[x] &= \sum_{t=1}^s y_j^s[t] \prod_{i \in N_x} \Phi \left( \frac{x - x_i[t]}{c_i^x(s)} \right) / \\ & / \sum_{t=1}^s \prod_{i \in N_x} \Phi \left( \frac{x - x_i[t]}{c_i^x(s)} \right), j = \overline{1, l}, i = \overline{1, k}, t = \overline{1, s} \end{aligned} \tag{5}$$

where bandwidth parameters  $c_i^x(s)$  are optimized according to the cross-validation procedure using the sample  $\{x[t], y^s[t], t = \overline{1, s}\}$ :

$$\begin{aligned} c_i^x(s) &= \frac{1}{s} \sum_{t=1}^s \left[ y_j^s[t] - \frac{\sum_{v \neq t} y_j^s[v] \prod_{i \in N_x} \Phi \left( \frac{x_i[t] - x_i[v]}{c(s)} \right)}{\sum_{v \neq t} \prod_{i \in N_x} \Phi \left( \frac{x_i[t] - x_i[v]}{c(s)} \right)} \right]^2 \\ & \rightarrow \min_{c(s)}, j = \overline{1, l}, i = \overline{1, k} \end{aligned} \tag{6}$$

For the all estimates above we assume that kernels  $\Phi(\cdot)$  and bandwidths  $c_{jk}^\varepsilon(s)$  and  $c_i^x(s)$  satisfy convergence criteria [04]. It must be pointed out that in estimates (3)-(6)  $N_\varepsilon, N_x$  denote sets of selected components of the residual vector  $\varepsilon$  and the input vector  $x$ , associated with  $\varepsilon_j$  and  $x_i$ . Estimation of each j-th component of the estimate  $y^s$  as well as the output vector  $y$  is performed by taking into account only those dimensions of  $\varepsilon$  and  $x$  that have influence on the mentioned component.

The suggested algorithm has some unique features if compared to counterparts [07, 10]. Firstly, an improved procedure of residuals calculation is introduced. While calculating residuals, one creates matrix, where residuals are placed according to corresponding input value leading to response change. Such matrix is able to be used when forecasting system's outputs with every fixed input configuration. In most cases, to detect how response depends on any conceivable input pattern is impossible. Therefore, estimation of such relation is essential, especially for systems, which cannot be explored by arranging experiments.

Secondly, new way of model construction is proposed. Estimation of output variables forecasting involves only those inputs that directly influence the particular output

component. This leads to significant reducing of dimensionality of corresponding nonparametric estimates and removal of irrelevant inputs having no contribution to the result but producing extra noise and offsets. Thus, the measure results in considerable improvement of modelling accuracy and calculation speed.

Thirdly, the suggested algorithm is universal, and it can easily accept further modifications of estimates configuration and ways of residuals calculation. One can implement the approach for any comprehensiveness of available information on functional relations between variables, namely, for parametric, nonparametric and combined cases. In particular, this will affect residuals calculation according to the formula (2). Right part of the equality can, therefore, differ from the parametric form represented in (2).

### EXPERIMENTAL RESEARCH

Simulation and computational approval of the suggested algorithm is performed while modelling flow distribution problem [25] in hydraulic network. The three-loop network designed for computational experiment is depicted in Figure 1.

In Figure 1, arcs denote single pipeline segments; block  $H$  - an active head being a model of a pump. The amount of arcs  $n = 10$ , and of knots  $k = 8$ , the dimension of the corresponding system of equations  $l = 10$ , input vector dimension  $m = 7$ . At some knots outflows  $q_i, i = \overline{1, k}$  are implemented. Every segment of the network has its hydraulic resistance  $s_j, j = \overline{1, l}$ . The arrows in contours correspond to selected directions of traversal. In accordance with Kirchhoff's law for pipeline networks the equation system is as follows:

$$F(x, y, \alpha) = \begin{cases} -y_1[t] + y_8[t] = x_1[t] \\ y_1[t] + y_2[t] - y_4[t] = x_2[t] \\ -y_2[t] + y_3[t] = x_3[t] \\ -y_5[t] + y_9[t] = x_4[t] \\ -y_3[t] + y_4[t] + y_7[t] - y_9[t] = 0 \\ -y_7[t] + y_{10}[t] = x_5[t] \\ y_6[t] - y_{10}[t] = x_6[t] \\ \alpha_1|y_1[t]|y_1[t] + \alpha_3|y_3[t]|y_3[t] + \alpha_4|y_4[t]|y_4[t] + \alpha_8|y_8[t]|y_8[t] = x_7[t] \\ \alpha_2|y_2[t]|y_2[t] + \alpha_4|y_4[t]|y_4[t] + \alpha_5|y_5[t]|y_5[t] + \alpha_9|y_9[t]|y_9[t] = 0 \\ \alpha_3|y_3[t]|y_3[t] + \alpha_6|y_6[t]|y_6[t] + \alpha_7|y_7[t]|y_7[t] + \alpha_{10}|y_{10}[t]|y_{10}[t] = 0 \end{cases} \quad 7)$$

In the expression (7) we denote input effects, namely  $x_i = q_i, i = \overline{1, m-1}$  outflows in knots of the network, and the active hydraulic head  $x_m = h$ . Outputs  $y_j, j = \overline{1, l}$  are flows in the corresponding arcs. Hydraulic resistances are represented by the parameters  $\alpha_j = s_j, j = \overline{1, l}$  of the equation system (7). The system of equations (7) has a single solution.

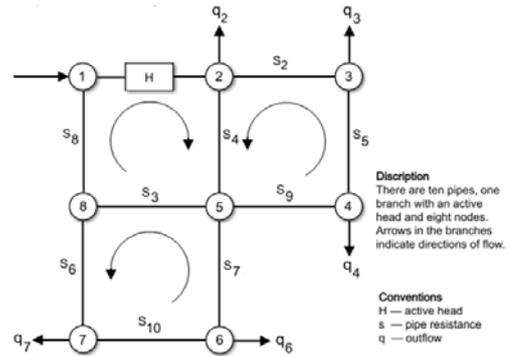


Figure 1: Schematic representation of hydraulic network

Data sets are generated according the following procedure. Input vector sample is as follows:

$$\tilde{x}[t] = x + 2 \cdot p \cdot (\mu[t] - q), t = \overline{1, s} \quad 8)$$

where  $x$  is an initial value,  $p$  - variation from the initial approximation,  $\mu[t], t = \overline{1, s}$  - random sequence evenly distributed within the interval  $[0, 1]$ ,  $q$  - shift parameter.

Each value in (8) should be subject to random noise influence. Random noise is added with the procedure similar to (8):

$$x[t] = \tilde{x}[t] + 2 \cdot \xi \cdot (r[t] - q), t = \overline{1, s} \quad 9)$$

where  $\xi$  is random noise amplitude,  $r[t], t = \overline{1, s}$  - random sequence evenly distributed within the interval  $[0, 1]$ ,  $q$  - shift parameter.

Output values  $y[t] = (y_1[t], \dots, y_l[t])$  for given inputs  $x[t] = (x_1[t], \dots, x_k[t])$  are taken as numerical solutions of the system (4) according to the Newton's method.

Assume  $\{x[t], y[t], t = \overline{1, s}$  an initial training sample of the state vector of the hydraulic network.

Let an identification quality criterion be mean squared error (MSE):

$$MSE = \frac{1}{s} \sum_{t=1}^s (y[t] - y^*[t])^2 \quad 10)$$

where  $y$  is forecast of the output,  $y^*$  is actual output. Computational research of the identification algorithm was performed by repeated starting of the process for the following parameters:

- Number of repetitions  $N = 100$ ;
- Spread of the sample from the initial state  $p = 25\%$ ;
- Noise levels  $\xi = 5\%$ ,  $\xi = 10\%$ ,  $\xi = 25\%$ ;
- Shift parameter  $q = 0.5$ ;
- Parameter  $\gamma = d/t$ , where  $d = 10^{-3}$ ;
- Training sample size  $s = 100, 150, 300$ ;

- Parameters of the model  
 $\alpha = (1.47 \ 23 \ 431 \ 1.74 \ 522 \ 5.14 \ 5.14 \ 1.74 \ 1.74 \ 522) \times 10^3$ ;
- Vector of input effects  
 $x = (-89 \ 16.5 \ 19.2 \ 23.1 \ 17.4 \ 12.8 \ 12.2)$ .

Results of computational research of the proposed algorithm compared to the basic multivariate identification algorithm [07] are represented in Table 1 showing criteria for the comparison are accuracy *MSE* and average operation time *T* for different noise levels  $\xi$  and amounts of measurements *s* in the sample. Averaging was performed over *N* computational experiments.

Further, the authors illustrate how the proposed identification algorithm reacts on the varying input vector. Apply following vector  $\tilde{x}[t], t = \overline{1, s}$  as an initial one:

$$\tilde{x}[t] = (-89 \ 16.5 \ 19.2 \ 23.1 \ 17.4 \ 12.8 \ 12.2) + d\tilde{x}[t], t = \overline{1, s}$$

where sample size *s* = 200 and  $d\tilde{x}[t], t = \overline{1, s}$  is the function in Figure 2.

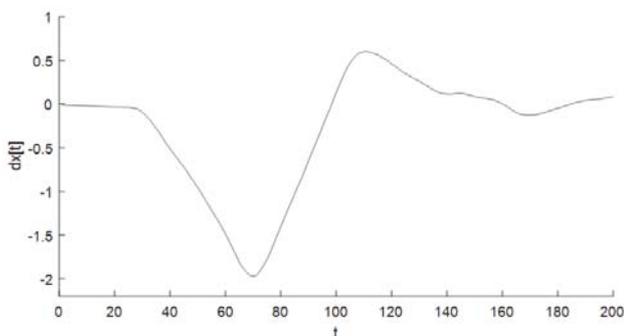


Figure 2: Function  $d\tilde{x}[t]$  for sample size *s* = 200

In accordance with (9), impose noise  $\xi = 10\%$  on the vector  $\tilde{x}[t], t = \overline{1, s}$  at  $q = 0.5$ . For the accepted input vector  $x[t], t = \overline{1, s}$ , generate the output vector  $y[t], t = \overline{1, s}$  in the course of solving the system (7) by the Newton's method. The resulting sample  $\{x[t], y[t]\}, t = \overline{1, s}$  is taken as training sample for the next experiment.

According to the procedure (1), results of parameter estimation are represented in Figure 3.

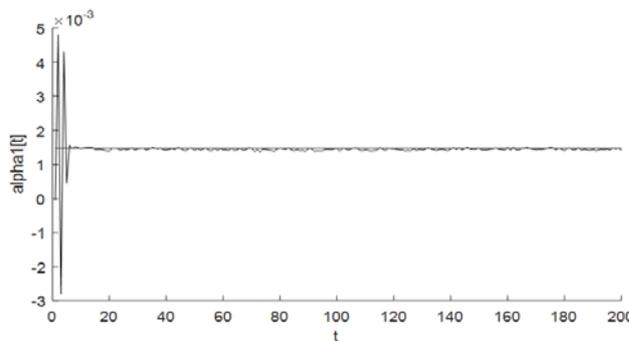


Figure 3: An example of the parameter estimation  $\alpha_1[t]$  compared to the actual value for sample size *s* = 200, and noise level  $\xi = 10\%$

Figure 4 depicts results of the output forecasting compared to the actual output of the system (7). One experiment for sample size *s* = 200, additive noise level  $\xi = 10\%$ , and for the first parameter  $\alpha_1$  and the first component of the output vector *y*<sub>1</sub> is illustrated here.

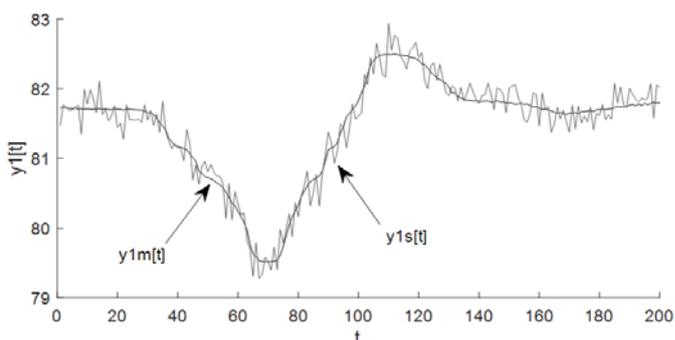


Figure 4: An example of the model output  $y_{1m}[t]$  compared to the actual response  $y_{1s}[t]$  of the system for sample size *s* = 200, and noise level  $\xi = 10\%$

Table 1: Results of Computational Experiments on the Proposed Algorithm Compared To the Basic One

		Basic algorithm			Proposed algorithm		
s		100	150	300	100	150	300
$\xi = 5\%$							
MSE		$1.4789 \cdot 10^{-4}$	$8.8372 \cdot 10^{-5}$	$4.7507 \cdot 10^{-5}$	$1.4522 \cdot 10^{-4}$	$8.6311 \cdot 10^{-5}$	$4.4636 \cdot 10^{-5}$
T		0.0425 sec	0.0691 sec	0.1862 sec	0.0230 sec	0.0293 sec	0.0708 sec
$\xi = 10\%$							
MSE		$2.2526 \cdot 10^{-4}$	$1.5019 \cdot 10^{-4}$	$8.3073 \cdot 10^{-5}$	$2.2061 \cdot 10^{-4}$	$1.4882 \cdot 10^{-4}$	$7.9786 \cdot 10^{-5}$
T		0.0373 sec	0.0635 sec	0.2101 sec	0.0202 sec	0.0296 sec	0.0823 sec
$\xi = 25\%$							
MSE		$9.0610 \cdot 10^{-4}$	$7.7164 \cdot 10^{-4}$	$4.5179 \cdot 10^{-4}$	$8.9630 \cdot 10^{-4}$	$7.4954 \cdot 10^{-4}$	$4.3147 \cdot 10^{-4}$
T		0.0408 sec	0.0591 sec	0.2227 sec	0.0257 sec	0.0297 sec	0.1018 sec

## DISCUSSION

Results of computational experiments confirm convergence of the proposed algorithm. As expected, accuracy of modelling improves for greater sizes of training sample. Noise causes negative effect on identification process. The proposed algorithm provides more accurate results if compared to the basic algorithm. Moreover, it is at least two times faster than its predecessor.

The improvement of basic features of the suggested algorithm if compared to the basic one is due to its selectivity. The algorithm involves in estimation of system solution as well as in response estimation only those components of input effect that have direct influence on it. In other words, a priori information is used more efficiently in the algorithm. Usage of the approach to system's response estimation provides reducing of dimensionality of the estimates, which is additionally resulted in the increased calculation rate.

## CONCLUSION

In the paper a new approach to identification and forecasting of multivariate systems response was proposed. This approach was submitted as an improved algorithm able to distinguish better accuracy and faster calculation speed due to optimal design of nonparametric response estimation procedure. The main idea was to reduce complexity by selecting relevant input components directly influencing the corresponding output response. This approach can be considered as substantial step in developing novice methods of multivariate identification.

The proposed algorithm makes it possible to forecast the system output for arbitrary input effects by one-time training. Compared to counterpart algorithms, there is no more need to solve system of equation numerically or to train the model for each new input. Prediction of output variables is performed by imputation of an input to the nonparametric estimates within the modelling algorithm.

An advantage of the approach is simplified training of the model. Training procedure can be performed at once for all the possible inputs. A new way of residual calculation was implemented in the algorithm. Residuals were referenced to certain combination of input values. That means residuals serve as intermediate source of information about system's behavior enabling fast and accurate forecasting of output variables.

One fundamental problem when identifying multivariate system is amount of available information of the system under research. Relationships between variables are often pure qualitative, i.e. one can only determine some subsets of interconnected variables without any parametric structure within. In this case, the proposed algorithm is applicable as well. Residuals calculation step (2) remains the same with one remark that the right side of the expression will be used to calculate pre-trained nonparametric regression up to all variables in the  $j$ -th mixture.

To sum up, we can claim that the proposed approach and corresponding computational algorithm is universal and efficient instrument of solving multivariate identification problem for diverse statements and different combinations of available information of the system under research.

One of the most demanding problems of the modern theory of identification and modelling is identification of multivariate dynamic systems [25]. The direction of future research is related to the development of algorithms discussed in the article for dynamic case. At present, there exist reliable algorithms for linear dynamical systems only [26]. The development of algorithms dealing with dynamic processes is associated with noticeable difficulties such as lack of suitable mathematical methods and extremely large computational resources requirements. In order to overcome these difficulties new methods are required. Development of such methods is the subject of future research.

## ACKNOWLEDGMENTS

The study has been undertaken as part of the research into the subject MK-1574.2017.8 "Designing the expert system of the analysis and control of reliability, risks and emergencies in support of the operation of petroleum refinery equipment" funded by the Grant Advisory Board for the President of the Russian Federation in a bid to provide governmental support to young Russian scientists.

## REFERENCES

1. Tsympkin, Y.Z. (1995). Information Theory of Identification. Science.
2. Eykhoff, P. (1975). Fundamentals of Identification of Control Systems. Mir.
3. Abdushukurov, A.A., Norboev, F.S. (2017). Kernel estimation of characteristic function. Journal of Mathematical Sciences, vol. 220, 643-649.
4. Hardle, W. (1993). Applied Nonparametric Regression. Mir.
5. Du, Z. (2016). Nonparametric bootstrap tests for independence of generalized errors. The Econometrics Journal. vol. 19, 55-83.
6. Friedrich, S., Beyersmann, J., Allignol, A., Winterfeld, U., Schumacher, M. (2017). Nonparametric estimation of pregnancy outcome probabilities. Annals of Applied Statistics. vol. 11, 840-867.
7. Hozé, N., Holcman, D. (2017). Statistical methods for large ensembles of super-resolution stochastic single particle trajectories in cell biology. Annual Review of Statistics and Its Application, vol. 4, 189-223.
8. Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

9. Yaping, X., Pengfei, C. (2015). The application of a presented hybrid BFGS-based method for data analysis in automation system. Jabalpur: International Conference on Computational Intelligence and Communication Networks.
10. Tsai, Y. (2017). The multivariate climatic and anthropogenic elasticity of streamflow in the eastern united states. *Journal of Hydrology: Regional Studies*, vol. 9. 199-215.
11. Fan, Y., de Micheaux, P.L., Penev, S., Salopek, D. (2017). Multivariate nonparametric test of independence, *Journal of Multivariate Analysis*. vol. 153, 189-210.
12. Medvedev, A.V. (2015). *Fundamentals of the Theory of Adaptive Systems*. Siberian State Aerospace University.
13. Krasnoshtanov, A.P. (2001). *Combined Multiply Connected Systems*. Science.
14. Markovich, N.M., Ryzhov, M., Krieger, U.R. (2017). Nonparametric analysis of extremes on web graphs: pagerank versus max-linear model. *Communications in Computer and Information Science*, vol. 700, 13-26.
15. Karpenko, S.S., Zybin, E.Yu., Kosyanchuk, V.V. (2018). Nonparametric method for failures detection and localization in the actuating subsystem of aircraft control system. *IOP Conference Series: Materials Science and Engineering Workshop on Materials and Engineering in Aeronautics*, 10-12.
16. Cai, C., You, H., Chen, N. (2018). Nonparametric estimation for a spectrally negative lévy risk process based on low-frequency observation. *Journal of Computational and Applied Mathematics*. vol. 328, 432-442.
17. Medvedev, A.V., Nizameev, A.R. (2006). *Nonparametric Modelling and Control of Multi Connected Systems*. *Siberian Journal of Science and Technology*, vol. 8, 44-47.
18. Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. *Machine Learning & Pattern Recognition series*. Chapman & Hall/CRCs.
19. Mangalova, E.S., Chubarova, O.S., Zhalnin, D. (2017). Decision trees control of static system under incomplete information, *Applied Methods of Statistical Analysis. Nonparametric Methods in Cybernetics and System Analysis (AMSA)*. Krasnoyarsk.
20. Hamid, R.M. (2017). A Hybrid Approach to Parameter Identification of Linear Delay Differential Equations Involving Multiply Delays. *International Journal of Control*, vol. 91, 1-44.
21. Guo, L., Tomizuka, M. (1999). Parameter Identification with Derivative Shift Operator Parametrization. *Automatica*, vol. 35, 1073-1080.
22. Yingkui, G., Jing, L. (2012). Multi-State System Reliability: A New and Systematic Review. *Procedia Engineering*, vol. 29, 531-536
23. Batalovic, V., Danilovic, D., Zivkovic, M., (2011). Model of oil and gas pipeline rinsing using the fluid flow. *Journal of Applied Engineering Science*, vol.9, 237-242.
24. Loginov, K.V., Myznikov, A.M., Faizullin, R.T. (2006). Calculation, Optimization and Control Modes of Operation of Large Hydraulic Networks. *Mathematical Modelling*, vol. 18, 92-106.
25. Medvedev, A.V., Raskina, A.V. (2017). On the nonparametric identification and dual adaptive control of dynamic processes. *Journal of Siberian Federal University: mathematics and physics*, vol. 10, 96-107.
26. Zhurabok, A.N., Shumsky, A.E., Pavlov, S.V. (2017). Diagnosis of linear dynamic systems by the nonparametric method. *Automation and Remote Control*, vol. 78, 1173-1188.

*Paper submitted: 05.03.2018.*

*Paper accepted: 29.05.2018.*

*This is an open access article distributed under the CC BY-NC-ND 4.0 terms and conditions.*