This paper proposes a method for improving empirical models of complex technological objects with insufficient information about the input and output values of an object's parameters. It has been established that most methods for constructing empirical models require knowledge of the statistical characteristics of the input and output values of an object. When modeling complex non-reproducible stochastic processes that evolve over time, information about the parameters and structure of an object is usually not available. A method has been proposed where input and output values are treated as fuzzy quantities with a triangular membership function. Since at some points in the region, the triangular membership function is undifferentiated, it is inconvenient to use it in its typical form to solve the problem of optimal control. Therefore, it is proposed to approximate it with the Gaussian membership function. It is shown that such an approximation is reduced to finding one parameter, which is determined by the least squares method. Its value practically does not depend on the magnitude of the uncertainty interval while the value characterizing the accuracy of approximation is a monotonously increasing function that has a linear character. This makes it possible to define the main operations on fuzzy numbers and derive an empirical model for the case of a polynomial "base" model. The resulting model is linear in its parameters, so a genetic algorithm can be used to find them. It is shown that genetic algorithms can be used in the construction of empirical polynomial models when input parameters are interpreted as fuzzy numbers.

Thus, it can be argued that when constructing an empirical model of an object that is affected by external disturbances that cannot be measured, it is advisable to approximate all input quantities with a triangular Gaussian membership function

Keywords: empirical model, membership function, function approximation, fuzzy numbers, genetic algorithm

UDC 681.5.015:622.24
DOI: 10.15587/1729-4061.2023.276586

IMPROVING EMPIRICAL MODELS OF COMPLEX TECHNOLOGICAL OBJECTS UNDER CONDITIONS OF UNCERTAINTY

Mykhail Gorbiychuk
Doctor of Technical Sciences, Professor*
Dmytro Kropyvnytskyi
Postgraduate Student*
Vitalia Kropyvnytska
Corresponding author
PhD, Associate Professor
Department of Computer Systems and Networks**
E-mail: vitalia.krop@gmail.com
*Department of Automation and Computer-Integrated Technology**
**Ivano-Frankivsk National Technical University of Oil and Gas
Karpatska str., 15, Ivano-Frankivsk, Ukraine, 76019

Received date 26.01.2023
Accepted date 30.03.2023
Published date 28.04.2023

How to Cite: *Horbiychuk, M., Kropyvnytskyi, D., Kropyvnytska, V. (2023). Improving empirical models of complex technological objects under conditions of uncertainty. Eastern-European Journal of Enterprise Technologies, 2 (2 (122)), 53–63. doi: https://doi.org/10.15587/1729-4061.2023.276586*

1. Introduction

When solving problems of optimal control over complex technological objects, when their analytical models are unknown, it becomes necessary to build empirical models. Most methods for constructing such models (least squares method, reference vector method, nuclear evaluation method) require knowledge of the statistical characteristics of the input and output values of the object. Under real conditions, such information is usually not available.

Managing technological objects that operate under conditions of uncertainty and incompleteness of information puts forward high demands on the content of information and methods of its processing.

Assume that the object is affected by external disturbances that are not measurable. The statistical characteristics of the disturbances are unknown. In such a situation, it is natural to assume that the input values are fuzzy numbers with triangular membership functions. The choice of this type of membership function is due to the fact that it is convenient to calculate and can be used for any number of terms. However, the triangular membership function cannot be differentiated at some points in the field of definition. This disadvantage can be avoided when approximating it with the Gaussian function.

Therefore, a relevant task is to improve the method for constructing empirical models of complex technological processes, which is based on the assumption that input and output quantities are fuzzy numbers with a triangular membership function that is approximated by the Gaussian membership function, which makes it possible to build an empirical model taking into account the fuzziness of both input and output quantities.

2. Literature review and problem statement

The well-known least squares method (LSM), which is used to build empirical models, assumes that the arguments of the model are measured accurately, and an additive perturbation is superimposed on the original value. In the case when the disturbances in each observation of the original value are uncorrelated and have the same variance, then LSM gives the best unbiased estimate for the linear regression model [1]. This situation is rare in practice. As a rule, input and output variables are under the influence of disturbances, taking into account which in the construction of empirical models, is possible if one uses the Bayesian approach [2].

Over the past few decades, considerable attention has been paid to non-parametric models. The main reason for their growing popularity is that they do not take a specific form of regression function that characterizes the relationship between input and output values [3]. They require only weak identification assumptions and thus minimize the risk of incorrect model specification. However, the main disadvantage of this method is that its use does not take into account the features of non-parametric evaluation, such as kernel choice and asymptotically descending support for local evaluation. This disadvantage can be avoided using the method of reference vectors [4], which occupies an important place in the modern theory of the construction of empirical models. Its linear version is close to the method of ridge regression [5]. The idea of this method is to build a hyperplane that divides the sample elements in an optimal way. Moreover, the hyperplane should as much as possible separate positive and negative examples from the set that is being trained. Underlying the construction of the algorithm for training reference vectors is the concept of the core of the scalar product of the reference vector and the vector taken from the input space, which is a significant drawback. The core must be calculated for all possible pairs of points, which may be impossible when training, thereby leading to long calculations when predicting new points. The effectiveness of the method of reference vectors strongly depends on the selection of parameters.

To build regression models of objects that function under conditions of uncertainty, the apparatus of fuzzy sets can be used. Thus, in [6], an empirical model is synthesized in the form of a first-order polynomial, in which the coefficients of the model are represented as fuzzy numbers with a triangular membership function. The procedure for determining the coefficients of such a model is reduced to the problem of linear programming.

A similar approach to the construction of regression models is proposed in [7]. The empirical model is a first-order polynomial with fuzzy coefficients. The calculation of the parameters of the model was carried out according to the criterion of minimums of the sums of the areas of the membership functions of fuzzy coefficients. Its use ensures the selection of fuzzy numbers (model coefficients) that are the least blurred and take modal values as close to zero as possible.

In works [8, 9], for the synthesis of regression models of the first order, a procedure is proposed that is similar to the one developed in work [7]. Additionally, the method for artificial orthogonalization of plans for a passive experiment, which is based on fuzzy data clustering, was used.

The implementation of the described procedures that form the methodology makes it possible to obtain adequate mathematical models and find the optimal control, in terms of the final state, over complex technological processes under conditions of uncertainty.

Another approach [10] to the construction of first-order regression models is based on the assumption that the input values of the object are fuzzy numbers, and the coefficients of the model are the Gaussian membership functions. After conducting machine computational experiments, a refined membership function is obtained. These two functions are compared and, on the basis of such a comparison, the procedure for the least squares for calculating the coefficients of the empirical model is formed.

The proposed methods for constructing empirical models of objects that function under conditions of uncertainty have the disadvantage that they cover only a small class of models, and it is necessary to use rather complex computational procedures to calculate their coefficients.

In general, when constructing empirical models, it is assumed that the structure of the model is known, and the task of identification is to determine the parameters of such a model. However, in practice, situations often arise when the structure of the model is a priori unknown. Then it is necessary to choose a suitable empirical model by conducting a series of machine experiments based on a certain criterion. At the same time, there is no certainty that the chosen model is optimal in its class of models.

For the first time, the idea of choosing a model structure from a class of polynomial models with a certain criterion was proposed in [11]. In the process of implementing this idea, a method was developed that was termed the method for group consideration of arguments [11], where in the process of identification the model structure is selected from a given class of models. The method is based on sorting out the models that are gradually becoming more complex. Such a complication occurs due to an increase in the power of polynomial or the addition of a new one term of the series, leading to an increase in the volume of calculations. To reduce the volume of calculations, it is proposed to consider a method that uses the theory of genetic algorithms in its implementation [12].

The main advantage of using genetic algorithms is the absence of the need for additional information. Moreover, it was shown in [12] that this algorithm has internal parallelism. This makes it possible to develop software that will reduce the cost of machine time for calculations.

Therefore, it can be argued that in order to solve the problem of optimizing multiparametric functions, it is advisable to apply the method of choosing the optimal empirical model based on genetic algorithms.

3. The aim and objectives of the study

The aim of this work is to improve the inductive method for self-organization of empirical models based on genetic algorithms, taking into account the fuzziness of both input and output quantities. This will make it possible to optimize multiparametric functions that describe complex technological objects, the input and output parameters of which can be estimated only with a certain approximation.

To accomplish the aim, the following tasks have been set:

– to devise a method for approximating the triangular membership function with a Gaussian function to construct an empirical polynomial model;

– to devise a method for identifying the parameters of empirical models of complex technological objects based on a genetic algorithm, taking into account the fuzziness of the input parameters.

4. The study materials and methods

The object of our research is methods and algorithms for constructing empirical models of complex technological objects under conditions of uncertainty. Empirical models that belong to the class of polynomial ones and have one output value and an arbitrary number of input quantities are considered. The output quantity and input values are interpreted as fuzzy quantities with a triangular membership function.

The data used to build empirical models contain inaccuracies that significantly affect the result of the identification problem [1]. Taking into account the disturbances that are superimposed on the input and output values of the model requires knowledge, as a rule, of the statistical characteristics of such disturbances that are problematic to obtain under the conditions of operation at industrial facilities.

An alternative approach to taking into account the inaccuracies that accompany the observation of the input and output quantities of an object is the interpretation of such quantities in terms of fuzzy set theory, namely the possibility of using the Gaussian membership function in the construction of empirical models.

To find the parameters of such a model, it is advisable to use a genetic algorithm. Its essence is the fact that an ordered sequence of ones and zeros is built. Unity will correspond to a nonzero value of the model parameter, and zero will be the case when the corresponding parameter of the model becomes zero. At the initial stage, a pool of chromosomal relatives is randomly formed. Each of these chromosomes corresponds to some partial model from the selected class of models. The most "adapted" chromosome is selected from the pool of offspring by crossing and mutation using the fitness function (criterion of regularity or displacement). The selected chromosome determines the structure of the empirical model.

Improvement of the inductive method for self-organization of empirical models based on genetic algorithms, taking into account the fuzziness of both input and output quantities.

5. Results of research into the construction of empirical models of complex objects under conditions of uncertainty

5. 1. Method for approximating the triangular membership function with the Gaussian function to construct an empirical polynomial model

When solving a number of problems in the field of drilling, oil and gas production, gas transportation by main pipelines, etc., we have to deal with the fact that technological parameters are measured with certain errors. After all, the objects are subject to numerous disturbances. All this entails the need to consider technological parameters as fuzzy quantities, which are conveniently characterized by a triangular membership function [13].

To take into account the factor of fuzziness, in the mathematical description of objects, it is necessary to perform certain arithmetic operations on fuzzy values. The process of performing arithmetic operations (addition, subtraction, multiplication, and division) becomes possible if fuzzy numbers are defined as numbers of the $(L–R)$-type.

Let x be a fuzzy value of the $(L–R)$-type. Then its membership function can be represented as a composition of L and R-functions [13]:

$$\mu_{L-R}(z)=\begin{cases} L\left(\dfrac{a_x-x}{\alpha_L}\right), & x\le a_x, \\ R\left(\dfrac{x-a_x}{\alpha_R}\right), & x>a_x, \end{cases}$$

where $\alpha_L>0$, $\alpha_R>0$ are the left and right fuzziness coefficients; Z_0 is the modal value of a fuzzy number.

So, a fuzzy number of the $(L–R)$-type is uniquely determined by the trio of its parameters $\langle a_x, \alpha_L, \alpha_R\rangle$.

Note that the triangular membership function, which is symmetric with respect to a_x, is a function of the $(L–R)$-type. Such a function is inconvenient for practical use because it is undifferentiated at some points from the definition area.

Therefore, the triangular membership function:

$$\mu(x)=\begin{cases} \dfrac{2}{\Delta}(x-a_x)+1, \; x\in[a_x-\Delta/2; a_x], \\ -\dfrac{2}{\Delta}(x-a_x)+1, \; x\in[a_x; a_x+\Delta/2] \end{cases} \quad (1)$$

is proposed to be approximated with the Gaussian function:

$$\mu_G(x)=\exp\left(-\frac{(x-a_x)^2}{2\alpha^2}\right), \quad (2)$$

where Δ is the uncertainty interval of the fuzzy value x; $\mu(a_x)=\mu_G(a_x)=1$; α is the concentration coefficient of fuzzy value x.

Since functions (1) and (2) at each of the definition intervals $x\in[a_x–\Delta/2; a_x]$ and $x\in[a_x; a_x+\Delta/2]$ are monotonous, then when they are approximated, they will have no more than two common points. The first of these is determined by the value of a_x, and the second will occur when $x=x_a$. With this x value, there will be an interelation:

$$\mu(x_a)=\mu_G(x_a)=\theta. \quad (3)$$

Obviously, the a_x value does not affect the form of membership functions (1) and (2) but only determines their position on the abscissa axis. Therefore, the a_x value does not affect the accuracy of the approximation of function (1) with function (2). Let $a_x=0$. Then formulas (1) and (2) will take the following form:

$$\mu(x)=\begin{cases} \dfrac{2}{\Delta}x+1, \; x\in[-\Delta/2; 0], \\ -\dfrac{2}{\Delta}x+1, \; x\in[0; \Delta/2] \end{cases} \quad (4)$$

and

$$\mu_G(x)=\exp\left(-\frac{x^2}{2\alpha^2}\right). \quad (5)$$

Given that functions (4) and (5) are symmetrical with respect to the origin of coordinates, approximation was carried out on the range of values $x\in[0; \Delta/2]$.

On the range of values $x\in[0; \Delta/2]$, equation (4) produced:

$$\mu(x_a)=-\frac{2}{\Delta}x_a+1.$$

Taking into account condition (3), we get the following expression:

$$\theta=-\frac{2}{\Delta}x_a+1.$$

The last equation yielded:

$$x_a = \frac{(1-\theta)\Delta}{2}. \quad (6)$$

Taking into account the a_x value, defined by formula (6), the membership function (5) is as follows:

$$\mu_G(x_a) = \exp\left(-\frac{(1-\theta)^2 \Delta^2}{8\alpha^2}\right).$$

Since $\mu_G(x_a)=\theta$, then:

$$\exp\left(-\frac{(1-\theta)^2 \Delta^2}{8\alpha^2}\right) = \theta.$$

Henceforth:

$$\alpha^2 = -\frac{(1-\theta)^2 \Delta^2}{8\ln\theta}, \quad (7)$$

where $0<\theta<1$.

Analysis of formula (7) shows that the coefficient of concentration α of membership function (2) depends on the basis Δ of the triangular membership function and the value of the ordinate, which is determined by the intersection point of membership functions (1) and (2) at $x \in [a_x - \Delta/2; a_x]$.

Since Δ is a priori known quantity, the accuracy of the approximation of function (1) by function (2) will depend on the value of the ordinate θ. It is possible to estimate the value of Δ by constructing a confidence interval using the Chebyshev inequality [14].

The accuracy of approximation is defined as the sum of the squares of deviation of the ordinates of function (5) from the corresponding ordinates of function (4):

$$E = \sum_{i=1}^{n} \left(\mu(x_i) - \mu_G(x_i)\right)^2, \quad (8)$$

where $x \in [0; \Delta/(2iT)]$; T – discreteness step; n is the number of ordinates of the function $\mu(x)$ on the segment $x \in [0; \Delta/2]$.

Parameter θ is selected from the minimum condition of expression (8). To do this, the value $\mu_G(x_i)$, which is determined by formula (5), is substituted into ratio (8). This takes into account the value of the value of α^2 according to formula (7). As a result, the following was obtained:

$$E(\theta) = \sum_{i=1}^{n} \left(\mu_i - \exp\left(\frac{4x_i^2 \ln\theta}{(1-\theta)^2 \Delta^2}\right)\right)^2, \quad (9)$$

where $\mu_i = \mu(x_i)$.

The function $E(\theta)$ is nonlinear and the value of θ, which minimizes (9), can only be found by the numerical method. Since known numerical methods find only the local minimum [15] at a certain change interval θ, we construct a plot of the function $E(\theta)$ (Fig. 1).

The plot, which is constructed for Δ=0.5, demonstrate that function (9) reaches its smallest value on the segment $\theta \in [0.4; 0.7]$. To find the minimum of function (9), we use the method of golden section [1∈6].

The following program parameters were selected:

– the starting point for finding the interval of the local minimum is 0.4;

– an error of finding the minimum function (9) is 10^{-6};

– the uncertainty interval of a fuzzy value is 0.5.

As a result, we got the following solution to the problem:

θ^*=0.5152; $E(\alpha^*)$=0.0703.

Fig. 2 illustrates the process of approximation of function (4) with function (5).

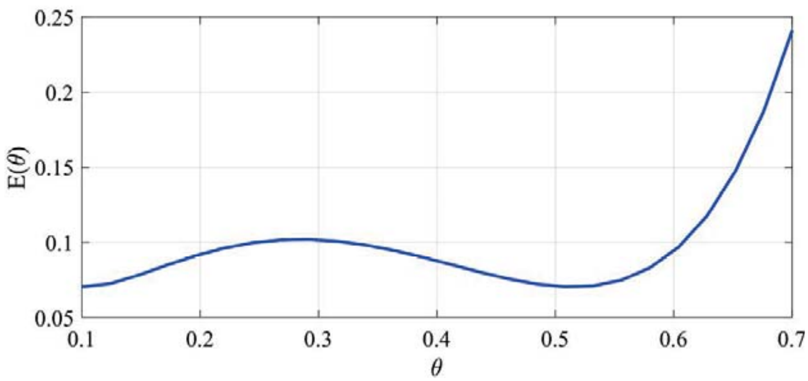

Fig. 1. Dependence of the approximation error on the value of θ

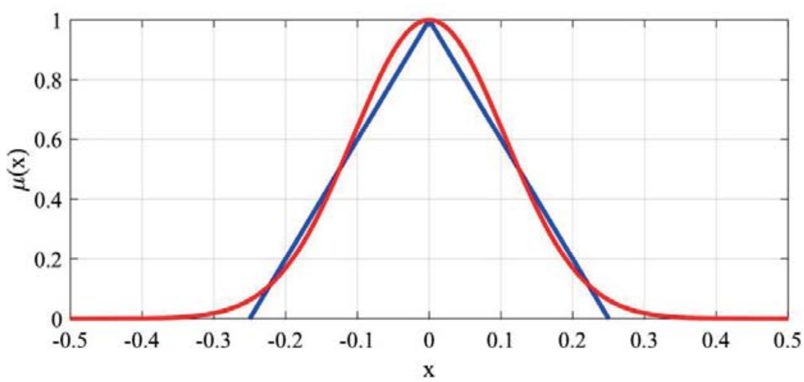

Fig. 2. Approximation of a triangular function with the Gaussian membership function

Analysis of formula (9) shows that the value of $E(\theta)$ depends not only on θ but also on the magnitude of the uncertainty interval of the fuzzy value Δ, which is an a priori quantity; its value is set by the researcher.

To identify the influence of value Δ on the approximation process, the problem of minimizing function (9) was solved for different values of Δ (Table 1).

Table 1

Results of approximating a triangular function with the Gaussian membership function at different values of Δ

No. of entry	Δ	θ^*	$E(\theta^*)$
1	1.0	0.5138	0.1380
2	0.8	0.5141	0.1112
3	0.6	0.5147	0.0839
4	0.4	0.5159	0.0567
5	0.2	0.5197	0.0297

Fig. 3 shows how the values θ* and $E(\theta^*)$ change depending on the change in the value of Δ.

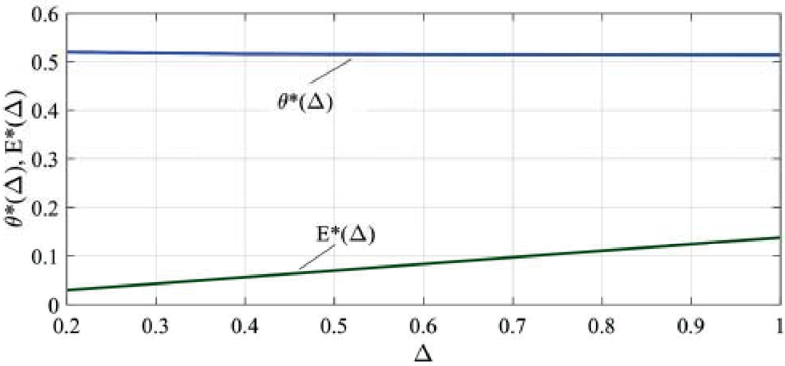

Fig. 3. The dependence of θ*(Δ) and E^*(Δ) on changes in the value of Δ

From the analysis of Fig. 3, it follows that the value of θ* is practically independent of the value of the uncertainty interval Δ, and the value of $E^*(\Delta)$ is a monotonously increasing function that has a linear character.

The least squares method determined the coefficients of dependence:

$$E^*(\Delta)=a_0+a_1\Delta. \tag{10}$$

As a result, we obtained; a_0=0.0026 and a_1=0.1355.

The accuracy of the approximation of triangular membership function (4) with Gaussian function (5) was estimated by the coefficient of determination, which is calculated by the following formula [17]:

$$R^2=\frac{\left(\sum_{i=1}^{N+1}\left(\mu_a^{(i)}-\bar{\mu}_a\right)\left(\mu_G^{(i)}-\bar{\mu}_a\right)\right)^2}{\sum_{i=1}^{N+1}\left(\mu_a^{(i)}-\bar{\mu}_a\right)^2\sum_{i=1}^{N+1}\left(\mu_G^{(i)}-\bar{\mu}_a\right)^2},$$

where $\mu_a^{(i)}$, $\mu_G^{(i)}$ are the value of the triangular and Gaussian membership functions, calculated from formulas (4) and (5) at values $x_i=\frac{\Delta}{2N}(i-1)$, $i=\overline{1,N+1}$, $\bar{\mu}_a=\frac{1}{N+1}\sum_{i=1}^{N+1}\mu_a^{(i)}$, N=100. As a result of the calculations, we received: R=0.9913.

Since the value of the coefficient of determination R is close to unity, the accuracy of approximating membership function (4) with Gaussian function (5) is quite high.

It should be noted that in [18], without justification, based on intuitive reasoning, the value θ=0.5 was chosen. As follows from Table 1, the value θ=0.5 differs little from the values θ^*, which are obtained as a result of solving minimization problem (9).

As an example of the application of the devised method for approximating a triangular membership function with a Gaussian function, the process of constructing an empirical model of an object that has m inputs and one output can be considered.

We assume that the object is affected by external disturbances that are not measurable. The statistical characteristics of the disturbances are unknown. In such a situation, it is natural to assume that the input values are fuzzy numbers with triangular membership functions, which we approximate with Gaussian functions.

So, the following model of the object is considered:

$$y(\bar{x})=\sum_{i=0}^{N-1}c_i\prod_{j=1}^{m}x_j^{\varphi_{ij}}, \tag{11}$$

where $\sum_{j=1}^{m}\varphi_{ij}\le r$, r is the power of polynomial (11); m is the number of variables.

The number of terms N in polynomial (11) is calculated by the following formula [19]:

$$N=\frac{(r+m)!}{r!m!}.$$

The structure of empirical model (11) is determined by the matrix of the powers of polynomial (11), which takes the following form:

$$\Phi=\begin{bmatrix}\varphi_{01} & \varphi_{02} & \cdots & \varphi_{0m}\\ \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1m}\\ \cdots & \cdots & \cdots & \cdots\\ \varphi_{N-1,1} & \varphi_{N-1,2} & \cdots & \varphi_{N-1,m}\end{bmatrix}.$$

Since the input values x_j are the fuzzy numbers that have a Gaussian membership function, the output value y, which is linear with respect to its parameters, is also a fuzzy value with a Gaussian membership function [13]:

$$\mu(y)=\exp\left(-\frac{\left(y-a_y\right)^2}{2\alpha_y^2}\right), \tag{12}$$

where a_y, α_y are the modal value and concentration coefficient of fuzzy quantity y.

To find the parameters a_y and α_y for membership function (11), the following actions on fuzzy numbers are necessary: adding fuzzy numbers, multiplying positive fuzzy numbers, multiplying a fuzzy number by a defined value, and raising to the power of a fuzzy number.

Based on the rules for performing arithmetic operations on fuzzy numbers [13], we adapt them for the case of Gaussian membership functions (2). Then any fuzzy number will be characterized by two parameters – a modal value and a fuzziness coefficient.

Let A_{LR} and B_{LR} be fuzzy numbers with a Gaussian membership function, which are characterized by two parameters A_{LR}=áa_1, α_añ and B_{LR}=áa_2, α_bñ. Therefore, $C_{LR}=A_{LR}\pm$ $\pm B_{LR}$=áa_c, α_cñ, where $a_c=a_1\pm a_2$; $\alpha_c=\alpha_a+\alpha_b$. $C_{LR}=A_{LR}B_{LR}$= =áa_c, α_cñ, where $a_c=a_1a_2$; $a_c=a_1\alpha_b+a_2\alpha_a$ under condition a_1>0 i a_2>0. If a_1 and a_2 take different signs, then $C_{LR}=A_{LR}B_{LR}$=áa_c, α_cñ, where $a_c=a_1a_2$; $a_c=a_2\alpha_a-a_1\alpha_b$. For negative values a_1 and a_2, the multiplication operation will be as follows: $C_{LR}=A_{LR}B_{LR}$=áa_c, α_cñ, where $a_c=a_1a_2$; $a_c=-a_2\alpha_a-a_1\alpha_b$. The operation of multiplying a fuzzy number a_1 by a definite number q follows from the statement that a definite number can be considered as a fuzzy number B_{LR} with parameters $a_2=q$, $\alpha_b=b_b=0$. Then, from the multiplication ratios of two fuzzy numbers for which $a_2=q$, $\alpha_b=b_b=0$, it follows that $C_{LR}=qA_{LR}$=áa_c, α_c,ñ, where $a_c=qa_1$; $a_c=q\alpha_a$. In the last formulas, a_c is treated as some physical quantity, therefore a_c>0.

Now we can find $\sum_{i=1}^{k} q_i A_i$, where A_i, q_i are fuzzy and definite numbers. The modal value of the fuzzy number A_i is a_i, and the fuzziness coefficients α_i, $i=\overline{1,k}$. We find $q_1A_1+q_2A_2$ first. We introduce the following notation $V_1=q_1A_1$ and $V_2=q_2A_2$. Then we get the fuzzy number $V=V_1+V_2$ with the parameters $V_{LR}=\langle \nu, \alpha_\nu \rangle$. According to the rule of adding fuzzy numbers V_1 and V_2, we have $\nu=\nu_1+\nu_2$. Taking into account that $\nu_1=q_1a_1$, $\nu_2=q_2a_2$, $\alpha_{v_1}=q_1\alpha_1$ and $\alpha_{v_2}=q_2\alpha_2$, we get $(q_1A_1+q_2A_2)_{LR}=\langle a_s, \alpha_s \rangle$, where $\nu_s=q_1a_1+q_2a_2$, $\alpha_s=q_1\alpha_1+q_2\alpha_2$. Obviously, the result can be extended to an arbitrary number of terms. Then:

$$\left\langle \sum_{i=1}^{k} q_i A_i \right\rangle_{LR} = \langle a_s, \alpha_s \rangle,$$

where

$$a_s = \sum_{i=1}^{k} q_i a_i, \quad \alpha_s = \sum_{i=1}^{k} q_i \alpha_i. \tag{13}$$

Now we find the power of fuzzy number $A_{LR}^n = \langle a_d, \alpha_d \rangle$. Let $n=2$. Then $A_{LR}^2 = A_{LR} \cdot A_{LR}$. Using the rule of multiplication of two fuzzy numbers for which $a_1=a_2$ and $\alpha_a=\alpha$, we get:

$$A_{LR}^2 = \langle a_{d,2}, \alpha_{d,2} \rangle,$$

where

$$a_{d,2} = a_1^2, \quad \alpha_{d,2} = 2a_1\alpha_a.$$

For $n=3$, we have $A_{LR}^3 = A_{LR}^2 \cdot A_{LR}$. Power A_{LR}^3 is the product of two fuzzy numbers A_{LR}^2 and A_{LR} with parameters $a_{d,2}=a_1^2$, a_1, $\alpha_{d,2}=2a_1\alpha_a$, and α. Based on the rule of multiplication of two fuzzy numbers, we have:

$$A_{LR}^2 \cdot A_{LR} = \langle a_{d,3}, \alpha_{d,3} \rangle,$$

where

$$a_{d,3} = a_1^3, \quad \alpha_{d,3} = 3a_1^2\alpha_a.$$

Extending the result to an arbitrary integer value n≥0, we come to the conclusion that:

$$A_{LR}^n = \langle a_d, \alpha_d \rangle,$$

where

$$a_d = a_1^n, \quad \alpha_d = n a_1^{n-1}\alpha_a. \tag{14}$$

Assume that the following product of fuzzy numbers is assigned: $A = \prod_{i=1}^{n} A_i$. It is necessary to find the parameters of a fuzzy number $(A)_{LR}=\langle a_p, \alpha_p \rangle$.

For the product of two fuzzy numbers with parameters a_1, a_2, $\alpha_{a,1}$ and $\alpha_{a,2}$, we get the following ratio: $(A_1A_2)_{LR} = \langle a_{p,2}, \alpha_{p,2} \rangle$, where $a_{p,2}=a_1a_2$, $\alpha_{p,2}=a_1\alpha_{a,2}+a_2\alpha_{a,1}$.

Now we must derive the product of three fuzzy numbers, which can be written in the following form: $A=(A_1A_2)A_3$. The product of two numbers in parentheses is a fuzzy number with parameters $a_{p,2}=a_1a_2$ and $\alpha_{p,2}=a_1\alpha_{a,2}+a_2\alpha_{a,1}$, and the fuzzy number A_3 has parameters a_3 and $\alpha_{a,3}$. According to the rule of multiplication of two fuzzy numbers (A_1A_2) and A_3, we have $((A_1A_2)A_3)_{LR} = \langle a_{p,3}, \alpha_{p,3} \rangle$, where $a_{p,3}=a_{p,2}a_3$, $\alpha_{p,3}=a_{p,2}\alpha_3+a_3\alpha_{p,2}$. Taking into account the $a_{p,2}$ and $\alpha_{p,2}$ values, we come to the following result: $\alpha_{p,3}=a_1a_2a_3$, $\alpha_{p,3}=a_2a_3\alpha_{a,1}+a_1a_3\alpha_{a,2}+a_1a_2\alpha_{a,3}$.

Extending the result to an arbitrary number of factors, we get:

$$\left(\prod_{i=1}^{n} A_i\right)_{LR} = \langle a_p, \alpha_{a,p} \rangle,$$

where

$$a_p = \prod_{i=1}^{n} a_i, \quad \alpha_{a,p} = \sum_{i=1}^{n} \alpha_{a,i} \prod_{k=1,k\neq i}^{n} a_k. \tag{15}$$

In the partial case, when $A_1=A_2=\ldots=A_n=A$, we arrive at formula (14).

The resulting formulas (13) to (15) make it possible to find the parameters of fuzzy value y, which is determined by formula (11), where c_i, $i=\overline{1,N}$ are definite numbers. We introduce the following notation: $\lambda_i = \prod_{j=1}^{m} x_j^{\varphi_{ij}}$. Then, if we take into account formula (13), then $\left(\sum_{i=0}^{N} c_i\lambda_i\right)_{LR} = \langle a_y, \alpha_\lambda \rangle$, where $a_y = \sum_{i=0}^{N} c_i a_{\lambda,i}$, $\alpha_\lambda = \sum_{i=0}^{N} c_i \alpha_{\lambda,i}$. We introduce another notation $\pi_{ij} = x_j^{\varphi_{ij}}$. Then $\lambda_i = \prod_{j=1}^{m} \pi_{ij}$. Taking into account the accepted designation and formula (15), we get $a_{\lambda,i} = \prod_{j=1}^{m} a_{\pi,ji}$, $\alpha_{\lambda,i} = \sum_{j=1}^{m} \alpha_{\pi,ji} \prod_{k=1,k\neq j}^{m} a_{\pi,ki}$. Define the $a_{\pi,ji}$ and $\alpha_{\pi,ji}$ parameters for the fuzzy value $x_j^{\varphi_{ij}}$. Taking into account formula (14), we come to the conclusion that $a_{\pi,ji} = a_{x,i}^{\varphi_{ij}}$ and $\alpha_{\pi,ji} = \varphi_{ij} a_{x,i}^{\varphi_{ij}-1}\alpha_{x,i}$, where $a_{x,j}$, $\alpha_{x,j}$ are the modal value and concentration coefficient of the fuzzy value x_j.

Knowing $a_{\pi,ji}$ and $\alpha_{\pi,ji}$, we find $a_{\lambda,i} = \prod_{j=1}^{m} a_{x,\mathrm{j}}^{\varphi_{ij}}$,

$$\alpha_{\lambda,i} = \sum_{j=1}^{m} \varphi_{ij} a_{x,j}^{\varphi_{ij}-1} \alpha_{x,j} \prod_{k=1,k\neq j}^{m} a_{x,k}^{\varphi_{ik}}.$$

Taking into account the results obtained, we find the modal value a_y and the fuzziness coefficient α_y for the fuzzy value y. So:

$$\left(\sum_{i=0}^{N} c_i \prod_{j=1}^{m} x_j^{\varphi_{ij}}\right)_{LR} = \langle a_y, \alpha_y \rangle, \tag{16}$$

where

$$a_y = \sum_{i=0}^{N} c_i \prod_{j=1}^{m} a_{x,\mathrm{j}}^{\varphi_{ij}}, \quad \alpha_y = \sum_{i=0}^{N}\sum_{j=1}^{m} c_i \varphi_{ij} a_{x,\mathrm{j}}^{\varphi_{ij}-1} \alpha_{x,\mathrm{j}} \prod_{k=1,k\neq j}^{m} a_{x,\mathrm{k}}^{\varphi_{ik}}.$$

Let γ be the slice for membership function (12). Then:

$$\exp\left(-\frac{(y-a_y)^2}{2\alpha_y^2}\right) = \gamma,$$

where $0<\gamma<1$.

From the last equation, we find:

$$y = a_y + \alpha_y \sqrt{\ln\frac{1}{\gamma^2}}.$$

If we take into account a_y and α_y, which are determined by ratios (16), we obtain an empirical model of the object, provided that the arguments of dependence (11) are interpreted as fuzzy quantities. So:

$$y=\sum_{i=0}^{N-1}c_i\prod_{j=1}^{m}a_{x,j}^{\varphi_{ij}}+a_\gamma\sum_{i=0}^{N-1}\sum_{j=1}^{m}c_i\varphi_{ij}a_{x,j}^{\varphi_{ij}-1}\alpha_{x,j}\prod_{k=1,k\neq j}^{m}a_{x,k}^{\varphi_{ik}}, \qquad (17)$$

where $a_\gamma=\sqrt{\ln\frac{1}{\gamma^2}}.$

Since all the variables x_j of empirical model (11) are interpreted as fuzzy quantities with a triangular membership function (1), which are approximated with exponential function (2), $\alpha_{x,j},\ j=\overline{1,m}$ is then calculated by (7):

$$\alpha_{x,j}=\eta\Delta_j,$$
$$j=\overline{1,m}, \qquad (18)$$

where $\eta=(1-\theta)\left(8\ln\frac{1}{\theta}\right)^{-1/2}.$

Taking into account formula (18), empirical model (17) will take the following form:

$$y=\sum_{i=0}^{N-1}c_i\prod_{j=1}^{m}a_{x,j}^{\varphi_{ij}}+A_\gamma\sum_{i=0}^{N-1}\sum_{j=1}^{m}c_i\varphi_{ij}a_{x,j}^{\varphi_{ij}-1}\Delta_j\prod_{k=1,k\neq j}^{m}a_{x,k}^{\varphi_{ik}}, \qquad (19)$$

where $A_\gamma=a_\gamma\eta.$

If we take into account the value of a_γ and η, then

$$A_\gamma=\frac{1}{2}(1-\theta)\sqrt{\frac{\ln\gamma}{\ln\theta}}.$$

Analysis of formulas (11) and (19) shows that the fuzziness of the arguments $x_j,\ j=\overline{1,m}$ leads to the appearance of an additional term in model (19), which is a kind of payment for the impossibility of accurately determining the values of $x_j,\ j=\overline{1,m}.$ In the case when the uncertainty intervals $\Delta_j,\ j=\overline{1,m}$ tend to zero $a_{x,j}$®$x_j,\ j=\overline{1,m},$, then models (11) and (19) become identical.

Model (19) will be represented in a slightly different form

$$y=\sum_{i=0}^{N-1}c_i\left(\prod_{j=1}^{m}a_{x,j}^{\varphi_{ij}}+A_\gamma\sum_{j=1}^{m}\varphi_{ij}a_{x,j}^{\varphi_{ij}-1}\Delta_j\prod_{k=1,k\neq j}^{m}a_{x,k}^{\varphi_{ik}}\right). \qquad (20)$$

So, the empirical model (20) is obtained, which is linear with respect to its parameters c_i. In the case when $\Delta_j=0$, $j=\overline{1,m},$ then $a_{x,j}=x_j,\ j=\overline{1,m}$, and we come to the empirical model (11).

We introduce the following designation:

$$X_i=\prod_{j=1}^{m}a_{x,j}^{\varphi_{ij}}+A_\gamma\sum_{j=1}^{m}\varphi_{ij}a_{x,j}^{\varphi_{ij}-1}\Delta_j\prod_{k=1,k\neq j}^{m}a_{x,k}^{\varphi_{ik}},\quad i=\overline{0,N-1}. \qquad (21)$$

Variables $X_i,\ i=\overline{0,N-1}$ can be interpreted as arguments of a fuzzy empirical model (20) and their number is determined both by the number of m input (independent) quantities of model (11) and the power r of polynomial (11).

Taking into account the accepted notation (21), empirical model (21) would be as follows:

$$y=\sum_{i=0}^{N}c_iX_i. \qquad (22)$$

Assume that the result of observing the work of an object is a set of values of both input and output values. We form from the obtained values the ordered structures X_r and $\bar{Y}$. The first ordered structure is X_r – the matrix of observations of the input values of the object, and the second structure $\bar{Y}$ is the vector of observations of the output value of the object.

As an approximation of the experimental data that are represented by the vector $\bar{Y}$, to model (22), we take the sum of the squares of deviations:

$$J(\bar{c})=\sum_{k=1}^{M}\left(Y_k-\bar{c}^T\bar{X}_k\right)^2, \qquad (23)$$

where M is the dimensionality of the vector $\bar{Y}$ (the number of observations); $\bar{X}_k$ – is the argument vector of fuzzy model (22), the components of which are calculated for the values of the input variables of model (11), in each experimental study.

The matrix of observations X_r has the size $M\times m$, in which the first row is the value of variables $x_j,\ j=\overline{1,m}$ in the first observation, the second row – the value of x_j in the second observation, etc. Thus,

$$X_r=\begin{bmatrix}x_{11} & x_{12} & \cdots & x_{1m}\\ x_{12} & x_{22} & \cdots & x_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ x_{M1} & x_{M2} & \cdots & x_{Mm}\end{bmatrix},$$

where $x_{ij},\ i=\overline{1,M},\ j=\overline{1,\mathrm{m}}$ is the value of the j-th input value in the i-th observation.

It is assumed that $a_{x,j}=\hat{x}_j,\ j=\overline{1,m},$ in formula (21), where $\hat{x}_j$ is the value of the input quantities obtained as a result of observations of the operation of the object. Taking into account the adopted designation, dependence (21) will take the following form:

$$X_{ik}=\prod_{j=1}^{m}\hat{x}_{\mathrm{ij}}^{\varphi_{kj}}+A_\gamma\sum_{j=1}^{m}\varphi_{kj}\hat{x}_{\mathrm{ij}}^{\varphi_{kj}-1}\Delta_j\prod_{r=1,r\neq j}^{m}\hat{x}_{\mathrm{ir}}^{\varphi ir},$$
$$i=\overline{1,\mathrm{M}},$$
$$k=\overline{0,N-1}. \qquad (24)$$

The components of vector $\bar{X}_k$ are calculated according to the following algorithm: for $i=k$: the k-th row of the matrix Φ determines the powers of polynomial (11) at coefficient c_k; the value of $X_{ik},\ i=\overline{1,M}$ is determined at the values of φ_{kj}, when $a_{x},_j$ acquires values $\hat{x}_{ij}$, where $\hat{x}_{ij}$ are the elements of the i-th row of the matrix X_r.

Minimizing the quotient criterion (23) relative to the parameters of the model c_i makes it possible to obtain a matrix normal equation [1]:

$$\left(F^TF\right)\bar{c}=F^T\bar{Y}, \qquad (25)$$

where $F=\left[\bar{X}_0\ \bar{X}_1\ \bar{X}_2\cdots\bar{X}_k\cdots\bar{X}_{N-1}\right]$ is the matrix, whose size is $M\times N$. The columns of matrix F are determined by the elements X_{ik}. For the k-th column of matrix F:

$$\bar{X}_k = \begin{bmatrix} \hat{x}_{11}^{\varphi_{k1}}\hat{x}_{12}^{\varphi_{k2}}\cdots\hat{x}_{1m}^{\varphi_{km}} + A_\gamma\sum_{j=1}^{m}\varphi_{kj}\hat{x}_{1j}^{\varphi_{kj}-1}\Delta_j\prod_{r=1,r\neq j}^{m}\hat{x}_{1j}^{\varphi_{kr}} \\ \hat{x}_{21}^{\varphi_{k1}}\hat{x}_{22}^{\varphi_{k2}}\cdots\hat{x}_{2m}^{\varphi_{km}} + A_\gamma\sum_{j=1}^{m}\varphi_{kj}\hat{x}_{2j}^{\varphi_{kj}-1}\Delta_j\prod_{r=1,r\neq j}^{m}\hat{x}_{2j}^{\varphi_{kr}} \\ \cdots \\ \hat{x}_{i1}^{\varphi_{k1}}\hat{x}_{i2}^{\varphi_{k2}}\cdots\hat{x}_{im}^{\varphi_{km}} + A_\gamma\sum_{j=1}^{m}\varphi_{kj}\hat{x}_{ij}^{\varphi_{kj}-1}\Delta_j\prod_{r=1,r\neq j}^{m}\hat{x}_{ij}^{\varphi_{kr}} \\ \cdots \\ \hat{x}_{M1}^{\varphi_{k1}}\hat{x}_{M2}^{\varphi_{k2}}\cdots\hat{x}_{Mm}^{\varphi_{km}} + A_\gamma\sum_{j=1}^{m}\varphi_{kj}\hat{x}_{Mj}^{\varphi_{kj}-1}\Delta_j\prod_{r=1,r\neq j}^{m}\hat{x}_{Mj}^{\varphi_{kr}} \end{bmatrix},$$

$$k=\overline{0,N-1}.$$

The matrix Φ, the structure of which is determined by the value of r and the number of independent variables m of the model, determines the class of models (23), where a certain number of its parameters can take zero values. The search for the best model from a given class will be carried out using the criterion of displacement or regularity [20]. To do this, the matrix F is divided into two parts F_1 and F_2. The vector of observations $\bar{Y}$ is also divided into two parts $\bar{Y}_1$ and $\bar{Y}_2$. The matrix F_1 will have the size of $wM\times N$, and the size of the matrix F_2 will be $(1-w)M\times N$. Accordingly, $\bar{Y}_1$ and $\bar{Y}_2$ will have wM and $(1-w)M$ components. The value of w is determined by the selected criterion. For the criterion of regularity, $w=0.7$, and for the criterion of displacement $w=0.5$.

5. 2. Method for identifying the parameters of complex empirical models, taking into account the fuzziness of input parameters

Since model (22) is linear relative to its parameters, then, to find them, we apply a method that is based on the theory of genetic algorithms [21].

Let's form an ordered sequence of ones and zeros. Unity will correspond to the nonzero value of the model parameter (22), and zero will be in the case when the corresponding parameter of model (22) acquires a zero value.

Such an ordered sequence in the theory of genetic algorithms is called a chromosome, and its each individual element is termed a genome.

At the initial stage, a pool of chromosome relatives is randomly formed, each of which has a length of N. Each of these chromosomes corresponds to some partial model from the model class (22). By crossing and mutation using a fitness function (the criterion of regularity or displacement), the most "adapted" chromosome is selected from the pool of descendants. The chromosome selected determines the structure of model (22).

The algorithm for selecting the best chromosome was built in [21]; we adapt it to the problem of synthesizing the optimal model, which belongs to the class of models (22). Such an algorithm consists of the following steps:

Step 1. Formation of the initial population. Randomly, the initial population of chromosomes from T individuals is formed. Each individual has N ordered unities and zeros, which set the structure T of models from the model class (22).

Step 2. Evaluation of chromosome fitness in a population. For each chromosome, the fitness function value is calculated as follows. Let the chromosome from the population T have ω zeros. Then ω columns are extracted from the matrices F_1 and F_2. As a result, the matrices F_1 and F_2 will change their size. The size of the matrix $\tilde{F}_1$ will be $wM\times(N-\omega)$, and the matrix $\tilde{F}_2$ will acquire the size of $(1-w)M\times(N-\omega)$. Knowing the matrix $\tilde{F}_1$ and vector $\bar{Y}_1$, we obtain a matrix normal equation that is similar to equation (25):

$$\left(\tilde{F}_1^T\tilde{F}_1\right)\bar{c}_{F,1}=\tilde{F}_1^T\bar{Y}_1. \quad (26)$$

where $\bar{a}_{F,1}=\left(a_{F,1}^{(0)},a_{F,1}^{(1)},\ldots,a_{F,1}^{(N-c-1)}\right)^T$ is the vector of nonzero parameters of the model, which is associated with the next chromosome. Having solved equation (26), we find the parameter vector of the model, which is associated with the next chromosome from the population T. Calculate the value of the fitness function for the next chromosome, which depends on the values of the model output at the points of the test set:

$$\bar{y}\left(F_2\right)=F_2\bar{c}_{F,1}. \quad (27)$$

Similar calculations are carried out for other chromosomes from the population T. As a result, we find the value of the fitness function $H_F(ch_j)$, $j=\overline{1,s}$ (s is the number of ch chromosomes in the population) for all chromosomes from population T.

Step 3. Checking the stop condition of the algorithm. For values calculated in the second step, we find:

$$H_F\left(ch^*\right)=\min_{j\in\{1,2,\ldots s\}}:H_F\left(ch_j\right). \quad (28)$$

If the condition $H_F(ch^*)\leq E$, where $E>0$, is satisfied, then the calculations are terminated. The algorithm stops working, except for meeting condition (28), in two more cases. First, when there is no significant reduction in the fitness function as a result of the calculations; secondly, when the algorithm has carried out a given number of iterations, but condition (28) is not met.

After fulfilling one of the three conditions for stopping the algorithm, the ch^* chromosome is selected, for which condition (28) holds. The ch^* chromosome sets the structure of model (22) of optimal complexity and forms the F^* matrix, removing from it those columns whose indices coincide with the position of the gene on the ch^* chromosome. The recalculation of the parameters of model (22) is carried out by solving a normal equation, which, by analogy with (26), is written in the following form:

$$\left(F^{*T}F^*\right)\bar{c}^*=F^{*T}\bar{Y}.$$

Step 4. Chromosome selection. Among the chromosomes from the T population, a certain number of chromosomes are selected that will participate in the creation of a new population. Such selection is carried out according to the principle of natural selection, when chromosomes with the best value of fitness function have a chance to get into the new population. There are a number of methods for chromosome selection (tournament method, roulette method, elitist method, cut-off method, etc.). To solve the problem of minimizing the function, the tournament method is effective [22]. Tournament selection assumes that the chromosome population is divided into subgroups, followed by the selection of the most adapted chromosomes from each subgroup. The number of individuals in each subgroup may be different but most often subgroups of 2–3 individuals are built.

Step 5. The formation of a new population of descendants. Selected chromosomes in the fourth step are subject to change with the help of two operators: crossing and mutation. The mutation operator is used much less frequently than the crossing operator. The probability of crossing is quite high – $0.5 \le P_c \le 1$, while the probability of mutation is $0 \le P_m \le 0.1$.

For a uniform law of distribution, a random number p_m is generated from the interval [0; 1] and, if the condition $P_m \le p_m$ is met, where P_m is the selected number from the interval [0; 0.1], a mutation of chromosomes occurs. Mutation in the chromosome is carried out over genes for which the $P_m \le p_m$ condition is valid, replacing one with zero and vice versa – zero by one. The mutation operator can be applied both to the pool of relatives and to the pool of descendants.

The crossing operator is applied to the pool of descendants. From the pool of descendants, a pair of chromosomes is randomly selected. Generate a random number p_c from the interval [0; 1]. When the $P_c \le p_c$ condition is satisfied, where P_c is the selected number from the interval [0.5; 1], then the crossing operator is applied to the pair of chromosomes. Otherwise, the pair of chromosomes remains unchanged. For a pair of chromosomes to be crossed, a random integer L_c is selected from the interval [1; N–1]. The number (locus) L_c determines the crossing point for a pair of chromosomes. The action of the crossing operator leads to the fact that in positions from L_c to N a pair of relatives exchange their chromosomes. As a result, a new pair of descendants is formed.

Step 6. After completing step 5, you move on to step 2.

The application of the developed method is discussed below on the example of compression of natural gas with centrifugal superchargers with a gas turbine drive.

When solving the problem of optimal control over the operation of centrifugal superchargers of natural gas [19], it becomes necessary to build an empirical model of consumption G of fuel gas, which is a function of technological parameters and temperature t_c of the environment

$$G = f\left(t_{in}, n_s, \varepsilon, P_{in}, t_c\right), \qquad (29)$$

where t_{in} – temperature of the gas at the inlet to the supercharger, °C; n_s – number of engine shaft revolutions, rpm; $\varepsilon = \frac{P_{out}}{P_{in}}$ – the degree of increase in gas pressure; P_{in}, P_{out} is the pressure at the inlet and outlet of the pressurizer, MPa; t_c is the ambient temperature, °C.

Functional dependence (29) will be approximated with empirical model (22). In accordance with the developed algorithm in the algorithmic language MatLab, the software for problem (23) was developed. The following program parameters were selected:

– the power of polynomial is 4;
– the number of chromosomes in the population – 30;
– the number of chromosomes in the subgroup – 3;
– the maximum number of iterations of the genetic algorithm is 200;
– the probability of crossing $(0.5 \le P_c \le 1)$ – 0.9;
– the mutation probability $(0 \le P_m \le 0.1)$ – 0.1;
– the fitness function is a criterion of regularity.

The parameters of empirical model (22), which characterize the conditions of uncertainty: the level of cut of the membership function $\gamma = 0.75$; the approximation coefficient of triangular function (1) with the Gaussian function (2) is $\theta = 0.52$.

To control the main characteristics of gas pumping units, technological requirements have been compiled [23]. According to them, the following Δ_j values were selected:

– gas temperature at the inlet to the supercharger – 0.004;
– the number of revolutions of the engine shaft – 0.004;
– pressure at the inlet to the supercharger and at the outlet – 0.004;
– degree of increase in gas pressure – 0.01;
– ambient temperature – 0.02.

All $\hat{x}_{ij}$, values, as follows from formula (24), must be different from zero. Therefore, the ambient temperature is given in degrees of the Kelvin scale.

The data used to build empirical model (22) have different units and scales of measurement. Therefore, all physical quantities that are included in formula (24), as well as the mass consumption of fuel gas, are expressed in dimensionless units according to the following formula:

$$x_j = \frac{X_j}{X_{\max,j}}, \quad j = \overline{1, m+1},$$

where X_j are the dimensional physical quantities; $X_{max,j}$ – the maximum value of X_j quantity in the data set; X_{m+1} – output of the empirical model.

Fig. 4 reflects the result of the synthesis of model (29) for the case where the empirical model (22) is chosen.

As a result, we obtained an empirical model in which the number of zero parameters is 64, and the number of parameters other than zero is 62.

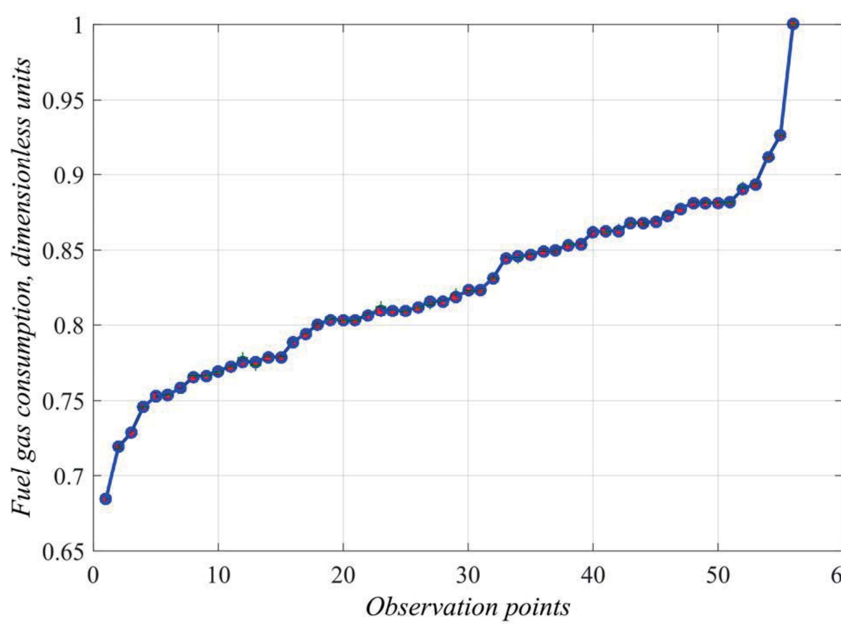

Fig. 4. Fitting experimental data to empirical model (22)

In Fig. 4, icons "o" mark the values that are obtained as a result of solving the problem of minimizing function (29) (experimental data), and the solid line is built on equation (22). In fact, there was a complete coincidence of experimental and calculated data, as evidenced by the approximation error value, calculated as the sum of squares of deviations of the calculated values from the corresponding experimental data.

Verification of the adequacy of the model with experimental data was carried out by calculating the correlation coefficient between the experimental and calculated values of the mass consumption of fuel gas. Its value is 0.997, which may indicate a high degree of consistency between experimental and calculated data.

6. Discussion of results of investigating the construction of an optimal-structure empirical model with fuzzy input values

When solving a number of problems related to the identification and optimal control over technological objects, it is necessary to take into account the uncertainty that arises as a result of the interaction of the object with the external environment. In such conditions, technological parameters are interpreted as fuzzy values. For fuzzy sets, an important characteristic is the membership function. The choice of the type of membership function is crucial in the construction of mathematical models. But the use of the apparatus of fuzzy sets requires a large amount of operations on fuzzy variables. Therefore, for the convenience of performing operations, it is desirable to work with membership functions with a standard form, for example, a triangular membership function. However, the triangular membership function is undifferentiated at some points, which makes it difficult to build a mathematical model. To eliminate this drawback, it is proposed to approximate the triangular membership function (4) with the Gaussian function (5), which is differentiated throughout the field of definition.

It is shown that the concentration coefficient α can be expressed through the uncertainty interval Δ using formula (7), which makes it possible to obtain an effective algorithm for constructing empirical models under uncertainty conditions.

The assessment of the accuracy of the approximation of the triangular membership function (4) with the Gaussian function (5) is made using the coefficient of determination, the value of which is R=0.9913. Since the obtained R value is close to unity, function (5) adequately approximates function (4).

On the basis of our studies, which established the relationship between the parameters Δ and α of membership functions (4) and (5), empirical models (22) were built that take into account the fuzziness of the input values. As a rule, empirical models are chosen in the form of polynomial dependences of a certain power, which means the choice of a certain class of models. When constructing empirical models, restrictions were made on the power of the polynomial. This limitation reduces the amount of calculations.

To select a specific model from this class, a genetic algorithm is used, which makes it possible to choose the optimal model in terms of its structure. On the example of solving the problem of identifying the parameters of the fuel gas consumption model, it is shown that when applying genetic algorithms, the number of significant parameters of the model decreased by almost 2 times, in contrast to the use of the LSM algorithm.

The research results showed that the developed method for synthesis of empirical models makes it possible to choose the optimal models in terms of structure under uncertainty conditions. The effectiveness of the method is confirmed by machine calculations in the empirical modeling of complex technological processes. The developed method, algorithms, and software can be used in solving the problem of optimal control over complex technological objects in the oil and gas sector, such as the processes of well construction, oil and gas production, and their transportation.

Thus, the basis of the improved method is the assumption of the fuzziness of the technological parameters of the object, that is, such parameters are considered as fuzzy quantities with a triangular membership function. It is shown that the effective solution of the problem is achieved by approximating the triangular membership function with the Gaussian function. For polynomial models using fuzzy arithmetic, their structure is determined, which takes into account the parameters of the membership function – the fuzziness interval and its modal value.

The results of theoretical studies are confirmed by a specific example of the construction of an empirical model of fuel gas consumption in the form of polynomial (22), the power of which is 4, and the number of input variables is 5. The developed method for synthesis of empirical model (29) with polynomial (22) made it possible to reduce the number of its coefficients from 128 to 62. The correlation coefficient between the results of the experiment and the values produced by model (22) was calculated. It was established that the value of the correlation coefficient is 0.997, which indicates a high power of consistency between experimental and calculated data.

The main disadvantage of our method for choosing the optimal empirical model in terms of structure is that the researcher sets the parameters of the triangular membership function based on practical experience and his/her intuition. To eliminate this drawback, it is advisable to build an expert database where the necessary information would be stored, and to devise rules for selecting initial data on interval changes in technological parameters that are inherent in the processes of the oil and gas industry.

7. Conclusions

1. It is shown that the implementation of arithmetic operations on fuzzy numbers is greatly simplified if the triangular function is approximated with the Gaussian membership function. Using the least squares method, an approximation parameter is defined, which makes it possible to express the concentration coefficient through the uncertainty interval of the triangular membership function.

2. A method and algorithm for identifying the parameters of complex empirical polynomial models using the theory of genetic algorithms for the case when the input values of the models are interpreted as fuzzy numbers

have been developed. As a result of our calculations carried out using the developed algorithm, it was established that there is a high degree of consistency between the experimental and calculated data, which is confirmed by the calculated value of the correlation coefficient. Its value is 0.997.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study and the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

All data are available in the main text of the manuscript.

References

1. Syniehlazov, V. M., Silvestrov, A. M. (2015). Teoriya identyfikatsiyi. Kyiv: NAU, 471.
2. Bidiuk, P. I., Kalinina, I. O., Hozhyi, O. P. (2021). Baiesivskyi analiz danykh. Kherson: Knyzhkove vydavnytstvo FOP Vyshemyrskyi V. S., 208.
3. Čížek, P., Sadıkoğlu, S. (2020). Robust nonparametric regression: A review. WIREs Computational Statistics, 12 (3), e1492. doi: https://doi.org/10.1002/wics.1492
4. Hable, R., Christmann, A. (2011). On qualitative robustness of support vector machines. Journal of Multivariate Analysis, 102 (6), 993–1007. doi: https://doi.org/10.1016/j.jmva.2011.01.009
5. Draper, N., Smith, H. (1998). Applied Regression Analysis. Wiley. 2016. 736.
6. Chen, F., Chen, Y., Zhou, J., Liu, Y. (2016). Optimizing h value for fuzzy linear regression with asymmetric triangular fuzzy coefficients. Engineering Applications of Artificial Intelligence, 47, 16–24. doi: https://doi.org/10.1016/j.engappai.2015.02.011
7. Raskin, L., Sira, O., Ivanchykhin, Y. (2017). Models and methods of regression analysis under conditions of fuzzy initial data. Eastern-European Journal of Enterprise Technologies, 4 (4 (88), 12–19. doi: https://doi.org/10.15587/1729-4061.2017.107536
8. Muzzioli, S., Ruggieri, A., De Baets, B. (2015). A comparison of fuzzy regression methods for the estimation of the implied volatility smile function. Fuzzy Sets and Systems, 266, 131–143. doi: doi: https://doi.org/10.1016/j.fss.2014.11.015
9. Seraya, O. V., Demin, D. A. (2012). Linear Regression Analysis of a Small Sample of Fuzzy Input Data. Journal of Automation and Information Sciences, 44 (7), 34–48. doi: https://doi.org/10.1615/jautomatinfscien.v44.i7.40
10. Domin, D. (2013). Application of artificial orthogonalization in search for optimal control of technological processes under uncertainty. Eastern-European Journal of Enterprise Technologies, 5 (9 (65)), 45–53. doi: https://doi.org/10.15587/1729-4061.2013.18452
11. Ivakhnenko, A. G. (1981). Induktivniy metod samoorganizatsii modeley slozhnykh sistem. Kyiv: Naukova dumka, 286.
12. Gorbiychuk, M., Medvedchuk, V., Pashkovskyi, B. (2014). The parallelism in the algorithm of the synthesis of models of optimal complexity based on the genetic algorithms. Eastern-European Journal of Enterprise Technologies, 4 (2 (70)), 42–48. doi: https://doi.org/10.15587/1729-4061.2014.26305
13. Raskin, L. G., Seraya, O. V., Volovschikov, V. Yu. (2016). Nechetkaya matematika. Kharki: NTU «KhPI», 203.
14. Ihnatkin, V. U. et al. (2019). Matematychne zabezpechennia tekhnichnykh zasobiv vymiriuvannia i kontroliu. Dnipro: TOV «Favor LTD», 339.
15. Verzhbitskiy, V. M. (2021). Osnovy chislennykh metodov. Moscow-Berlin: Direkt-Media. doi: https://doi.org/10.18720/SPBPU/2/ek21-2
16. Yaroshenko, O. I., Hryhorkiv, M. V. (2018). Chyslovi metody. Chernivtsi: Chernivetskyi nats. un-t, 172.
17. Dykha, M. V., Moroz, V. S. (2016). Ekonometriya. Kyiv: Tsentr navchalnoi literatury, 206.
18. Horbiychuk, M. I., Skripka, O. A., Pashkovskyi, B. V. (2016). Optymalnyi rozpodil kilkosti hazoperekachuvalnykh ahrehativ v umovakh nevyznachenosti pry zadanykh obsiahakh na perekachku pryrodnoho hazu. East European Scientific Journal, 2 (3 (7)), 53–58.
19. Gorbiychuk, M., Pashkovskyi, B., Moyseenko, O., Sabat, N. (2017). Solution of the optimization problem on the control over operation of gas pumping units under fuzzy conditions. Eastern-European Journal of Enterprise Technologies, 5 (2 (89)), 65–71. doi: https://doi.org/10.15587/1729-4061.2017.111349
20. Dranyshnykov, L. V. (2018). Intelektualni metody v upravlinni. Kamianske: DDTU, 416.
21. Gorbiychuk, M. I., Humenyuk, T. V. (2016). Synthesis Method of Empirical Models Optimal by Complexity under Uncertainty Conditions. Journal of Automation and Information Sciences, 48 (9), 64–74. doi: https://doi.org/10.1615/jautomatinfscien.v48.i9.50
22. Oliynyk, A. P., Subbotin, S. O., Oliynyk, O. O. (2012). Intelektualnyi analikh danykh. Zaporizhzhia: ZNTU, 277.
23. SOU 60.3-30019801-011:2004. Kompresorni stantsiyi. Kontrol teplotekhnichnykh ta ekolohichnykh kharakterystyk hazoperekachuvalnykh ahrehativ. Kyiv: DK Ukrtranshaz, 117.