

## Construction of Empirical Models of Complex Oscillation Processes with Non-Multiple Frequencies Based on the Principles of Genetic Algorithms

Mykhailo Horbiychuk, Olha Bila<sup>\*</sup>, Nataliia Lazoriv

*Ivano-Frankivsk National Technical University of Oil and Gas, 15, Karpatska Str., Ivano-Frankivsk, 76019, Ukraine*

Received: October 30, 2018. Revised: December 10, 2018. Accepted: February 28, 2019.

© 2019 The Authors. Published by Lviv Polytechnic National University.

### Abstract

A method for constructing the empirical models of complex processes has been developed on the basis of genetic algorithms which, compared to the inductive method of self-organization of models, significantly reduces computer time for their implementation. An approach has been used that allows a complex model to be considered as a composition of three components, i.e. a linear trend, an oscillatory component with non-multiple frequencies and a regression equation which simplifies the process of building complex models. To implement the proposed method, algorithms and software have been developed based on a specific example of the dependence of the water level in the river. The Dniester River weather conditions show that a model built on the basis of the proposed method describes the behavior of complex processes with sufficient accuracy. The resulting empirical model can be used to predict the water level depending on weather conditions.

**Keywords:** complex process; empirical model; unpredictable frequencies; genetic algorithm.

### 1. Introduction

The data obtained from the study of a number of environmental processes are usually presented in the form of time series. In addition, time series that are associated with environmental processes often have significant seasonal or periodic components. These components, as a rule, vary in time and in most cases are described by seasonal statistical models [1]. It is assumed that the structure of the model is selected in a certain way and the identification task is only to determine the parameters of such a model, which in most cases is solved using the method of least squares. It is obvious that the accuracy of the model's correspondence with the empirical data depends on the structure of the model itself. For example, when choosing a model in the form of a regression equation with increasing its terms, the error determined monotonically at all experimental points decreases. As soon as the number of regression members becomes equal to the number of experimental points, the error vanishes. In the case when the number of experimental points is greater than the number of terms in the regression model, it is always possible to obtain an infinitely large number of empirical models [2], that is, the method of least squares generates an infinitely large number of models for a given numerical series. This statement is valid if only one internal criterion is used to select models.

Based on the Gödel incompleteness theorem [3] the use of points of a sequence that have already been used to find coefficients by the method of least squares is impossible in principle to find the optimal and unique model. It is external additions that make it possible to find a single model optimal by this criterion of complexity [4].

<sup>\*</sup> Corresponding author. Email address: ksm@nung.edu.ua

**This paper should be cited as:** M. Horbiychuk, O. Bila, N. Lazoriv. Construction of empirical models of complex oscillation processes with non-multiple frequencies based on the principles of genetic algorithms. *Energy Engineering and Control Systems*, 2019, Vol. 5, No. 1, pp. 29–38. <https://doi.org/10.23939/jeecs2019.01.029>

The choice of the structure of the model is based on the type of the process curve. Then the question of choosing a basic function for a mathematical model arises. For example, it is recommended to use harmonic models for oscillatory processes with non-multiple frequencies with zero mean value, and if the oscillatory process tends to grow or fall, then a model that represents the sum of the polynomial trend and the harmonic remainder should be formed. In the general case, the construction of empirical models is based on the condition that a function is chosen from a given class, which in a certain sense best describes the experimental data.

To solve the problem, academician A.G. Ivakhnenko proposed a number of methods united by a common name – an inductive method of self-organization of models which generates three structures of algorithms [5, 6]: a multi-row algorithm for selecting the method of group accounting of arguments (GAA), a combinatorial algorithm of GAA with a complete search of all possible models from a given polynomial and multi-row algorithm, where the coefficients of a certain complete polynomial are equalized to zero on each selection row. The multi-row GAA algorithm contains some intermediate quantities generated by its multi-row nature as arguments, and the GAA combinatorial algorithm requires significant computational efforts for its implementation [7], which grow exponentially with increasing number of arguments of the empirical model.

To reduce computational efforts and expand the field of application of empirical models for the implementation of which combinatorial algorithms are used, a method for constructing empirical models using genetic algorithms was proposed in studies [8, 9]. Later a similar algorithm for constructing empirical models was described in study [10].

The purpose of this work is to further improve the inductive method of constructing empirical models of complex processes containing a harmonic component with non-multiple frequencies based on the principles of genetic algorithms, which will make it possible to synthesize the empirical models that are optimal in complexity for predicting, in particular, ecological processes.

## 2. An empirical model of a complex oscillatory process with non-multiple frequencies

In the study of complex processes the analytical description of which is not known, the researcher has observable values of certain physical quantities characterizing such a process. These quantities form numerical sets; one of them forms a numerical series, called the values of the output quantity, and the rest are external factors that cause the flow of a certain physical process. The observed values of the output quantity in the form of a numerical series will be denoted by  $Y_c(t)$ , and external factors will be represented in the form of a vector  $\bar{x}(t)$  with components  $x_1(t)$ ,  $x_2(t)$ , ...,  $x_n(t)$ , where  $n$  – the number of external factors.

Then a number of complex processes (for example, ecological ones) can be represented by an empirical model containing a trend  $y_s(t)$ , a component  $G(t)$  caused by a harmonic change of some conditions, and a residual component  $\tilde{Y}(\bar{x}(t))$  that is a function of external factors, that is,

$$Y(t) = \tilde{Y}(\bar{x}(t)) + G(t) + y_s(t). \quad (1)$$

The component  $G(t)$  is modelled by the following expression [6]:

$$G(t) = A_0 + \sum_{j=1}^m (A_j \sin(\omega_j t) + B_j \cos(\omega_j t)), \quad (2)$$

where  $t$  – the timing cycle;  $A_0$ ,  $A_j$ ,  $B_j$  – parameters of the harmonic series (2);  $\omega_j$  – frequencies of harmonics of the series (2);  $j = \overline{1, m}$ .

In contrast to the harmonic Fourier series, for which  $\omega_j = j\omega$ , a harmonic series with unconstrained frequencies ( $\omega_j \neq j\omega$ ) (2) can more accurately approximate the results of observations of a physical process where the periodicity is not observed [11].

The trend  $y_s(t)$  that characterizes the non-stationary component of the process can be described by an empirical model in the form of a regression equation of order  $n_s$

$$y_s(t) = \sum_{i=0}^{n_s} q_i t^i ,$$

where  $\theta_i$  are the parameters of the model.

In order for the parameters of the series (2) to be estimated from the observations of the process, it is necessary for the condition [6],  $N \geq 3m+1$  to be met. The identification of model parameters (2) consists of several stages [2].

The first stage is the calculation of the balance coefficients  $\alpha_p$  based on the condition

$$\sum_{p=0}^{m-1} a_p (g(i+p) + g(i-p)) = g(i+m) + g(i-m), \quad i = \overline{m+1, N-m}, \quad (3)$$

where  $g$  are the samples of the realization of the process at discrete instants of time symmetrically placed with respect to an arbitrary point.

The balance coefficients  $\alpha_p$  are found [2] from the condition for minimizing the residual function

$$B = \sum_{i=m+1}^{N-m} b_i^2, \quad (4)$$

where  $b_i = g(i+m) - \sum_{p=0}^{m-1} a_p (g(i+p) + g(i-p)) + g(i-m)$ ,  $i = \overline{m+1, N-m}$ ,

which taking into account the value of  $b_i$  will take the following form:

$$\min_{\bar{a}} : J(\bar{a}) = \sum_{i=m+1}^{N-m} \left( z_{i,m} - \sum_{p=0}^{m-1} a_p g_{i,p} \right)^2, \quad (5)$$

where  $b_i = (\alpha_0, \alpha_1, \dots, \alpha_{m-1})^T$  is the vector of weight coefficients;  $z_{i,m} = \tilde{g}(i+m) + \tilde{g}(i-m)$   
 $g_{i,p} = \tilde{g}(i+p) + \tilde{g}(i-p)$ ;  $T$  is a matrix transposition symbol.

We write problem (5) in matrix-vector form

$$\min_{\bar{a}} : J(\bar{a}) = (\bar{z}_m - F_m \bar{a})^T (\bar{z}_m - F_m \bar{a}), \quad (6)$$

where  $\bar{z}_m = \begin{bmatrix} \tilde{g}(2m+1) + \tilde{g}(1) \\ \tilde{g}(2m+2) + \tilde{g}(2) \\ \dots \\ \tilde{g}(N) + \tilde{g}(N-2m) \end{bmatrix}$ ;

$$F_m = \begin{bmatrix} 2\tilde{g}(m+1) & \tilde{g}(m+2) + \tilde{g}(m) & \tilde{g}(m+3) + \tilde{g}(m-1) & \mathbf{K} & \tilde{g}(2m) + \tilde{g}(2) \\ 2\tilde{g}(m+2) & \tilde{g}(m+3) + \tilde{g}(m+1) & \tilde{g}(m+4) + \tilde{g}(m) & \mathbf{K} & \tilde{g}(2m+1) + \tilde{g}(3) \\ \mathbf{K} & \mathbf{K} & \mathbf{K} & \mathbf{K} & \mathbf{K} \\ 2\tilde{g}(N-m) & \tilde{g}(N-m+1) + \tilde{g}(N-m-1) & \tilde{g}(N-m+2) + \tilde{g}(N-m-2) & \mathbf{K} & \tilde{g}(N-1) + \tilde{g}(N-2m+1) \end{bmatrix}.$$

In the last expression, the values  $g$  of the corresponding discrete arguments are replaced by the data  $g$  obtained after subtracting the component  $Y_e(t)$  from the observed values  $y_s(t)$ .

Expression minimization (6) leads to a normal Gaussian equation, which in matrix form will have the following form:

$$F_m^T F_m \bar{\mathbf{a}} = F_m^T \bar{\mathbf{z}}_m. \quad (7)$$

Solving the equation (7) by one of the numerical methods, for example, by the Gauss method with the reverse motion [12], we determine the balance coefficients  $b_i = (\alpha_0, \alpha_1, \dots, \alpha_{m-1})^T$ .

In the second stage, equation

$$\mathbf{a}_0 + \sum_{p=1}^{m-1} \mathbf{a}_p \cos(p\omega) = \cos(m\omega)$$

for an arbitrary frequency  $\omega$ , which, with the help of the recurrence relation [6]

$$\cos(p\omega) = 2\cos((p-1)\omega)\cos\omega - \cos((p-2)\omega), \quad p = \overline{2, m}$$

is reduced to an algebraic equation in accordance to  $\cos\omega$

$$P_m z^m + P_{m-1} z^{m-1} + \dots + P_1 z + P_0 = 0, \quad (8)$$

where  $z = \cos\omega$ .

Equation (8) has  $m$  roots that unambiguously determine the frequencies  $\omega_j, j = \overline{1, m}$ .

The essence of the third stage is that, knowing the weight coefficients  $\alpha_p$ , we can write equation (8) whose solution relatively allows one to unambiguously determine the frequencies of the harmonics  $\omega_j, j = \overline{1, m}$  series (2).

To calculate the coefficients  $P_i, i = \overline{0, m}$  of equation (8), we use the following recurrent procedure [13] which consists of the following steps:

Step 1. Initialization. Form a zero matrix  $T$  by dimension  $m \times (m+1)$ .

Step 2. Put  $T_{12} = I, T_{21} = -I$ .

Step 3. Compute the elements of the matrix  $T$ :  $T_{ij} = 2T_{i-1, j-1} - T_{i-2, j}, i = \overline{3, m}, j = \overline{2, i+1}$ .

Step 4. Put  $T_{i1} = T_{i-2, 1}, i = \overline{3, m}$ .

Step 5. Calculate  $T_{mj} = T_{mj} - T_{ij} \mathbf{a}_p^{(i+1)}, i = \overline{1, m-1}, j = \overline{1, m+1}, T_{m1} := T_{m1} - \mathbf{a}_p^{(1)}$ .

Step 6. Form a vector of coefficients of equation (8)  $P_j = T_{mj}, j = \overline{1, m+1}$ .

After the coefficients of equation (8) have been determined, the integrated roots function of the Matlab system can be used to solve it.

### 3. Synthesis of a harmonic series model with non-multiple frequencies based on the principles of genetic algorithms

Now the problem is [15] in the optimal synthesis of the harmonic series (2). There are two approaches [6] to solving the problem. The first of these involves deleting the harmonics in various combinations from the full series  $\omega_j, j = \overline{1, m}$ . The second method is based on the ideas of multi-row algorithms of group accounting of arguments (MGAA). In accordance with this method, the number of harmonics included in the model constantly increases as long as this leads to a decrease in the selection criterion. The simplest algorithm is the sequential selection of the best model in each row. But the algorithm is more effective when several harmonics are allocated in each row [6].

The disadvantage of the first approach to solving this problem is the need to search through a large number of options the number of which is  $S_v = 2^m - 1$ . For example, for  $m=30$  you need to take 1073741823 variants, which requires considerable computer time. The second approach is characterized by the fact that as a result of the

implementation of the multi-row MGAA algorithm, it is impossible to obtain a mathematical model explicitly and this is a significant drawback of such a method. Therefore, to remove the problem of large dimension and obtain the model in explicit form, it is suggested to use a different approach to construct mathematical models of harmonic processes with non-multiple frequencies based on the ideas of genetic algorithms.

The essence of this approach is as follows [15]. The whole realization of the output value of the process or phenomenon is divided into two parts  $N_R$  and  $N_S$  in a certain proportion depending on the selection criterion [14]. For a set of data  $N_R$ , the weighting coefficients are defined as the solution of the linear algebraic equation (7) by Gauss elimination with the choice of the principal element [13]. The solution of equation (8) with respect to the variable  $z$  makes it possible to find the frequencies  $\omega_j$ ,  $j = \overline{1, m}$ . From known frequencies  $\omega_j$  on a set of points  $N_R$ , it is necessary to determine the parameters  $A_0$ ,  $A_j$  and  $B_j$  of the model (2). To do this, we form an ordered structure of length  $m$  in which in the  $i$ -th place will be zero or one, depending on whether the frequency  $\omega_j$  is removed from the chosen full series or left. In the theory of genetic algorithms, such an ordered sequence is called a chromosome or an individual, and the atomic element of a chromosome (one or zero) is a gene. A set of chromosomes forms a population. An important concept in the theory of genetic algorithms is the function of adaptation. Minimization of this function allows the selection from the whole population of those individuals which are the most adapted.

In the problem of synthesizing models of oscillatory processes, the adaptation function is the combined selection criterion [4]

$$r = \sqrt{n_d^2 + B^2}, \tag{9}$$

where  $n_d^2$  is the bias criterion, which is calculated by the formula [4]:

$$n_d^2 = \frac{\sum_{i=1}^N (g_i(R) - g_i(S))^2}{\sum_{i=1}^N \tilde{g}_i^2}; \tag{10}$$

$$B = \sum_{i=m+1}^{N-m} b_i^2; \tag{11}$$

$g_i(R)$ ,  $g_i(S)$  are the quantities whose values are computed on the set of points  $N_R$  and  $N_S$ , and the coefficients  $A_0$ ,  $A_j$  and  $B_j$  of the model (2) are found respectively on the sets  $N_R$  and  $N_S$ .

For the displacement criterion (9), the sets of values  $G(t)$  of the length  $N$  are divided into two parts in a proportion:  $N_R = N_S = 0,5N$ .

Thus, the problem of synthesizing the model of the oscillatory process will be formed as follows: from the initial population of chromosomes, by choosing an evolutionary selection, choose one, the chromosome which will ensure the best value of the adaptation function (the minimum value of the selection criterion (9)).

Formally, the algorithm of the GA method can be described as follows:

$$GA = \langle I, CP, SA, SE, NP \rangle. \tag{12}$$

The operator randomly generates an initial population of  $I$  individuals. The length of each chromosome from the population is determined by the number of members of the series (2). In the formed chromosome we double the ones and zeros. Let, for example, a chromosome  $Ch = [1001011]$  be generated. After the doubling operation, it will take the following form:  $Ch_d = [11 00 00 11 00 1111]$ . Since the coefficient  $A_0$  is always present in model (2), we add a single gene to the first position in the chromosome  $Ch_d$ . As a result, we get  $Ch_{d0} = [111 00 00 1100 1111]$ .

The necessity of the doubling operation is explained by the fact that each frequency corresponds to a pair of coefficients  $A_j$  and  $B_j$ .

The operator CP assesses the fitness of the chromosome in the population; the operator SA checks the termination conditions of the algorithm of the GA method; the operator SE selects chromosomes; the operator NP forms a new population of descendants from the selected chromosomes by the operator SA. After executing the operator SE, we transit to the operator SA.

The operator tuple (12) forms an algorithm similar to the one described in [8, 11].

#### 4. An empirical model of the remainder $\tilde{Y}(\bar{x}(t))$

In general, the empirical model (1) includes three components – a trend  $y_s(t)$  that can be described by a polynomial of a degree  $n_s$ , a harmonic component  $G(t)$  and a remainder  $\tilde{Y}(\bar{x}(t))$ , which is a function of external factors  $\bar{x}(t)$ .

After the results of observations of the value  $Y_e(t)$  the components of the process  $y_s(t)$  and  $G(t)$  are determined, we subtract the values of  $y_s(t)$  and  $G(t)$  from the number series  $Y_e(t)$ . As a result, we obtain a new number series

$$\tilde{Y}_e(\bar{x}(t)) = Y_e(t) - G(t) - y_s(t), \quad (13)$$

which will be approximated by a polynomial of degree  $\nu$

$$\tilde{Y}(\bar{x}(t)) = \sum_{i=0}^{M-1} a_i \prod_{j=1}^n x(t)_j^{s_{ij}}, \quad (14)$$

where  $M$  is the number of terms of the polynomial;  $n$  – the number of factors that are arguments to the function  $\tilde{Y}(\bar{x}(t))$  (size of the vector  $\bar{x}(t)$ );  $a_i$  – are the coefficients of the polynomial;  $S_{ji}$  – the degree of arguments that must

satisfy the constraint  $\sum_{j=1}^n s_{ij} \leq n$ .

The number of terms of a polynomial  $M$  is determined by the formula:

$$M = \frac{(\nu + n)!}{\nu! n!}. \quad (15)$$

As a result of monitoring the process, the researcher has a set of experimental values  $\tilde{Y}_e(\bar{x}(t))$  and  $\bar{x}(t)$ . Then, using the method of least squares, one can determine the parameters of the empirical model (14).

In practice, as a rule, the structure of the model (14) is unknown, which leads to the necessity of an arbitrary choice of both the number of functions and the form of the functions themselves in the model (14). To select the structure of the model (14) for a given degree of polynomial  $\nu$ , an inductive method for self-organizing models was suggested [6].

The implementation of the inductive method of self-organization of models is carried out in stages: the first stage is the generation of candidate models (in a certain order of increasing complexity); the second stage is the selection of the best model according to one of the selection criteria (regularity or minimum offset).

Of the three ways of generating candidate models [6], the combinatorial algorithm of MGAA [8] is the most attractive one [7] since it makes it possible to obtain a model (14), in which its arguments manifest themselves explicitly.

The disadvantage of the combinatorial algorithm is the need to search through a large number of candidate models the number of which is  $2^M \square I$ .

To solve the problem of large dimension, which is inherent in the combinatorial algorithm, we apply the genetic approach. As an empirical model, we will consider the polynomial (14) of degree  $v$ .

We form an ordered structure  $M$  in length in which the unit or zero will stand in the  $i$ -th place, depending on whether the parameter  $a_i$  of  $i = \overline{1, M}$  model (15) is different from zero, or zero. As before, such an ordered structure consisting of zeros and ones will be called a chromosome.

Thus, the problem of synthesizing the empirical model will be formed as follows: from the initial population of chromosomes, by means of evolutionary selection, choose a chromosome that provides the best value of the adaptation function (the minimum value of the selection criterion (20) or (21)). The algorithm for solving the problem is developed and described in studies [8, 11].

### 5. An example of synthesis of an empirical model of a complex process based on the principles of genetic algorithms

We apply the developed method of synthesis of mathematical models of complex processes on the principles of genetic algorithms for constructing a mathematical model of the water level in the river Dniester, depending on weather conditions.

The necessary information was collected (Nyzhniv village, Ivano-Frankivsk region) by monitoring the water level and the parameters that characterize the meteorological conditions – air temperature, precipitation, average wind speed and average daily barometric pressure. As a result of the observations, a numerical series of experimental data containing  $N=160$  points was obtained.

Analysis  $Y_t$  of changes in water level in the Dniester river shows that there is a time trend  $y_s(t)$ , which is linear in nature, and there is also a harmonic component  $G(t)$ .

From the time series  $Y_e(t)$  a linear trend was identified

$$y_s(t) = q_0 + q_1 t, \tag{16}$$

where  $\theta_0, \theta_1$  are the parameters of the linear trend.

The coefficients of the model (16) are determined by the method of least squares. As a result of the calculations, the following values of the coefficients of the linear trend were obtained:  $\theta_0=308.5712, \theta_1=-0.4484$ .

After separating the linear trend  $y_s(t)$  from the numerical series  $Y_e(t)$  (Fig. 1), a new numerical series is obtained containing the components  $G(t)$  and  $\tilde{Y}(\bar{x}(t))$ .

In the Matlab environment, a program for separating the harmonic component  $G(t)$  was created. The maximum number  $m=30$  of frequencies was chosen; the number of observation points  $N=160$  for building a model was chosen; probability of crossing  $P_c=0.9$ ; the probability of a mutation was  $P_m=0.1$ . The maximum number of coefficients of the model (2) that were determined was  $2m+1=61$ ; 8 of which are zero.

The result of the program reproduces Fig. 1, where the “o” sign is marked with experimental data after the linear trend  $y_s(t)$  is extracted from them, and “+” is the result of calculation of  $G(t)$  according to the model (2).

After separating the linear trend and the harmonic component from the experimental data, a residual component is obtained

$$\tilde{Y}_e(\bar{x}(t)) = Y_t - (G(t) + y_s(t)).$$

The value  $\tilde{Y}_e(\bar{x}(t))$  is a function of the parameters that determine the weather conditions in the observation area. At the same time, the water level in the river depends not only on the current amount of precipitation, but also on the precipitation that fell out the day before [2, 16]

$$\tilde{Y}_e(\bar{x}(t)) = j(T(t), f(t), f(t-1), f(t-2), \mathbf{K}, f(t-k), \mathbf{n}(t), p(t)), \quad (17)$$

where  $T(t)$  is an average daily air temperature, °C;  $f(t)$  is precipitation, mm / day;  $v(t)$  is average daily wind speed, m/s;  $p(t)$  is average daily barometric pressure, mm. gt;  $k$  is time shifting.

Based on the numerical experiments, it was found that  $k=1,2,3$ . Thus, the dependence (17) will be a function of seven variables

$$\tilde{Y}_e(\bar{x}(t)) = j(T(t), f(t), f(t-1), f(t-2), f(t-3), \mathbf{n}(t), p(t)). \quad (18)$$

We will consider the section of the river that is being monitored as a certain system characterized by a set of input quantities  $\bar{x}(t) = (x_1(t), x_2(t), \mathbf{K}, x_n(t))^T$  and an output quantity  $\tilde{Y}(\bar{x}(t))$ . For the case under consideration –  $x_1(t) = T(t)$ ,  $x_2(t) = f(t)$ ,  $x_3(t) = f(t-1)$ ,  $x_4 = f(t-2)$ ,  $x_5(t) = f(t-3)$ ,  $x_6(t) = \mathbf{n}(t)$ ,  $x_7(t) = p(t)$ .

The numerical series of the residual component  $\tilde{Y}_e(\bar{x}(t))$  will be approximated by the polynomial (15) of degree  $\nu=4$ . The degree of the polynomial was chosen from the condition of providing a certain “freedom” of choosing the degree  $\nu$  of the polynomial (14) by an algorithm for constructing mathematical models of complex processes on the principles of genetic algorithms.

Based on the developed algorithm [7, 8], using a genetic algorithm, a program was written in the Matlab environment for constructing a mathematical model of the residual component. The developed program allowed synthesizing a model having 144 non-zero (from the total number  $M=300$ ) and 156 zero parameters, a polynomial (14).

It is necessary to determine the sum of the three components of the dependence (1) for each discrete time  $t$  to learn the empirical model of the water level change of the Dniester, depending on the weather conditions.

Fig. 2 shows the results of calculations using the empirical relationship (1), where “+” denotes the computed values for model (1), and the “o” sign indicates the experimental values of the water level in the river Dniester. It can be seen from the graph that there are quite satisfactory coincidences between calculated and experimental data.

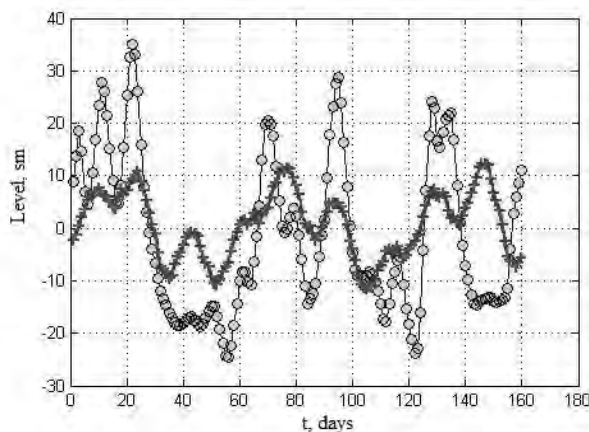


Fig. 1. Harmonic component of the oscillatory process of water level change (the Dniester River).

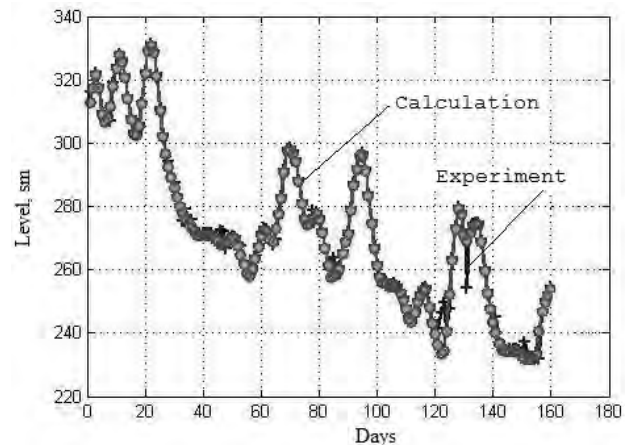


Fig. 2. Dependence of the water level in the river Dniester from weather conditions.

The adequacy of the model was verified using the correlation coefficient between the actual values  $Y_e(t)$  and the values  $Y(t)$  at the output of the model (1). The calculated value of the correlation coefficient was 0.9662, which indicates a high degree of similarities between the calculated and experimental values of the water level of the Dniester.



The obtained empirical model (1) can be used to predict the water level depending on the weather conditions. To implement such a forecast, a forecast of the parameters characterizing the weather conditions is necessary for a certain depth of the forecast. For this purpose, the method of empirical models described above can be used on the principles of genetic algorithms.

## 6. Conclusion

The inductive method was improved for constructing the empirical models of complex processes based on the principles of genetic algorithms, which made it possible to obtain models of optimal complexity and simplify the procedure for obtaining them. The case is considered when the process model can be represented as the sum of three components, i.e. a trend in the form of a polynomial of a certain degree, a harmonic series with non-multiple frequencies and a residual.

On the basis of the proposed method, an algorithm and corresponding software have been developed. Their working capacity is confirmed by a test example of building an empirical model for the water level change of the Dniester River.

## References

- [1] Box J. Analysis of time series. Forecast and management / J. Box, G. Jenkins; transl. from the English .. – M.: Mir, 1974. – 406 p. (in Russian)
- [2] Ivakhnenko A. G. Long – term forecasting and management of complex systems / A. G.Ivakhnenko/ – K.: Tekhnika, 1975. – 312 p. (in Russian)
- [3] V. A. Uspenskii. Gödel's incompleteness theorem / V. A. Uspenskii. – Moscow: Nauka, 1982. – 110 p. (in Russian)
- [4] Ivakhnenko A. G. Modeling stability against disturbances / A. G. Ivakhnenko, V. S. Stepashko. – Kyiv: Nauk. Dumka, 1985. – 216 p. (in Russian)
- [5] Ivakhnenko A. G. The prediction of random processes / A. G. Ivakhnenko, V. G. Lapa. – Kyiv: Naukova Dumka, 1969. – 420 p. (in Russian)
- [6] Ivakhnenko A. G. Inductive method of self-organization of models of complex systems / A. G. Ivakhnenko. – Kyiv: Naukova Dumka, 1981. – 286 p. (in Russian)
- [7] Lemke F. Self-organising Modelling for Decision Support // International Conference in Inductive Modelling ICIM' 2013. Berlin, Germany. 172–178 pp.
- [8] Gorbichuk M. I. Analysis of the parallel algorithm for synthesizing empirical models on the principles of genetic algorithms / M. I. Gorbichuk, V. M. Medvedchuk, A. N. Lazoriv // Problems of Control and Informatics. – 2016. – No. 1. – p. 112–130. (in Ukrainian)
- [9] Gorbichuk M. I. Method of synthesis of empirical models based on the genetic algorithms / M. I. Gorbichuk, M. I. Kogutyak, O. B. Vasilenko, I. V. Shchupak // Exploration and development of oil and gas fields. – 2009. – No. 4 (33). – p. 72–79. (in Ukrainian)
- [10] Gorbichuk M. I. Synthesis of functions of classification on the basis of genetic algorithms / M. I. Gorbichuk, S. T. Samulyak, I. B. Shchupak // Artificial Intelligence. – 2010. – No. 2. – p. 24–31. (in Ukrainian)
- [11] Moroz O. G. The GMD algorithm with the genetic search for the optimal model / O. G. Moroz // Control systems and machines. – 2016. – No. 6. – p. 73–88. (in Russian)
- [12] Ivakhnenko A. G. Self-organization of predictive models / A. G. Ivakhnenko, J. A. Muller. – K.: Tehnika, 1984. – 223 p. (in Russian)
- [13] Verzhbitsky V. M. Fundamentals of numerical methods: a textbook for high schools / V. M.Verzhbitsky. – M. : Higher School, 2002. – 840 p. (in Russian)
- [14] Ivakhnenko A. G. Handbook on typical programs for modeling / A. G. Ivakhnenko, Yu. V. Koppa, V. S. Stepashko and others; Ed. A. G. Ivakhnenko. – K. : Technique, 1980. – 184 p. (in Russian)
- [15] Gorbichuk M. I. Method of motivating mathematical models of folding processes based on the genetic algorithms / M. I. Gorbichuk, M. A. Shufnarovich / Information Problems of Computer Systems, Jurisprudence, Energy, Economy, Modeling and Control: Proceedings of the International Conference on Science and Technology, Buchach, 01-04.06, 2010, pp. 328–332. (in Ukrainian)
- [16] Rutkovskaya D. Neural networks, genetic algorithms and fuzzy systems / D. Rutkovskaya, M. Pilinsky, L. Rutkovsky; trans. from Polish. I. D. Rudinsky. – Moscow: Hot line-Telecom, 2004. – 452 p. (in Russian)

## **Побудова емпіричних моделей складних коливальних процесів з некрatними частоти на принципах генетичних алгоритмів**

Михайло Горбійчук, Ольга Біла, Наталія Лазорів

*Івано-Франківський національний технічний університет нафти і газу, вул. Карпатська, 15,  
Івано-Франківськ, 76019, Україна*

### **Анотація**

Розроблений метод побудови емпіричних моделей складних процесів на основі генетичних алгоритмів, що дозволяє, порівняно з індуктивним методом самоорганізації моделей, значно скоротити витрати машинного часу на їх реалізацію. Використаний підхід, що дозволяє складну модель розглядати як композицію трьох складових – лінійного тренда, коливальної складової з некрatними частотами і рівняння регресії, що спрощує процес побудови складних моделей. Для реалізації запропонованого методу розроблено алгоритмічне і програмне забезпечення. На конкретному прикладі залежності рівня води в р. Дністер від погодних умов показано, що модель, побудована на основі запропонованого методу, з достатньою точністю описує поведінку складних процесів. Отримана емпірична модель може бути використана для прогнозування рівня води залежно від погодних умов.

**Ключові слова:** складний процес; емпірична модель; генетичний алгоритм; експеримент; програмне забезпечення.