METHOD OF PREDICTING THE WATER LEVEL IN THE DNIESTER RIVER, DEPENDING ON WEATHER CONDITIONS

O.M. Mandryk, M.I. Horbiychuk

IFNTUOG Karpatska 15, Ivano-Frankivsk, 76019
Ukraine
gorb@nung.edu.ua

Abstract. This paper deals with a new method of constructing mathematical models of complex processes aliquant vibration frequencies, based on the ideas of genetic algorithms. Compared with the method of data handling group, this method makes it possible to synthesize mathematical models of any complexity without selecting the number of rows selection. Moreover, the use of genetic algorithms will significantly reduce the number of calculations when iterating the models. The new point method can be used to predict both the physical phenomena and complicated processes. The authors created the model of changes in the water level in the Dniester river on the basis that the above method gives satisfactory results of forecasting.

Keywords: *synthesis of mathematical models, genetic algorithm, chromosomes, external criterion aliquant frequency harmonic series forecasting accuracy*

Introduction

The gas transportation network of Ukraine includes about 35 thousand kilometers of gas lines, 71 compressor stations with total capacity of 5.4 million kilowatts. An important part of Ukraine's gas transit network is the Directorate of gas lines "Prykarpattransgaz" which incorporates 18 compressor stations with a total capacity of 380×106 nm^3/d. Also, laid across the Dniester river are such gas lines as: "Soyuz" gas line (Koropets village, Ternopil Province), "Progress" gas line and "Uryngoy - Pomary - Uzhgorod" gas line (Myhalche village, Ivano--Frankivsk Province).

Therefore, to ensure reliable and trouble-free operation of gas lines that pass through the Dniester river, the forecasting of its floods is important. Often floods cause the erosion of the banks, which could damage the "Torzhok - Dolyna" gas line (Ivano-Frankivsk Province), which is laid near the bed of the Dniester river.

According to the results of the observations for many years of the Dniester river regime at the hydro-observation station of Galych town, the most characteristic feature of the water regime of the Dniester river is frequent floods throughout the year of rain and snow origin. It should be noted that this flood level may occur

in all seasons. Overall the Dniester river as well as all the Carpathian rivers are characterized by flood regime, flood elevation (near the town of Galych) being in the range of 0.5-5 m.

Therefore, an important issue is to construct mathematical models of changes in the water level of the Dniester river, which will result in obtaining the prediction accuracy of its state and introduce the preventive measures to avoid the catastrophic consequences of floods.

1. Analysis of present-day research and publications

When studying a great number of processes occurring in the environment, we have to deal with the data obtained as a dynamic series. Moreover, the dynamic series that are related to environmental processes often have significant seasonal or periodic components, these components tend to vary in time and in most cases are described by seasonal statistical models [1]. In doing so, it is assumed that the structure of the model in some way is chosen and the identification problem was only a procedure for determining the parameters of this model, which were based in most cases on the least squares procedure. It is obvious that the accuracy of adjusting the models to the empirical data depends on the structure of the model itself. For example, when choosing the model in the form of a regression equation as the number of members of the equation increases, the error, which is defined for all experimental points, shows a monotonous decrease. As soon as the number of members regression is equal to the number of experimental points, the error becomes equal to zero. It should be noted that across a given number of points it is always possible to draw an infinitely large number of curves [2], that is the least squares procedure produces an infinitely large number of models for a given numerical series. The above statement is true if for choosing the models one criterion is used. The Gödel incompleteness theorem implies that using the training data sequence, which has been used to determine the coefficients for the least squares principle, cannot be used for finding an optimal and single model. Only the external additions will make it possible to find a single model of the optimal complexity [2].

The problem of creating mathematical models of complex technical and ecological processes is thoroughly examined in a number of publications by A.G. Ivakhnenko and his students [1, 2].

Of the known methods of predicting [1, 2] the most noteworthy one is the inductive method of self-organizing complex models [3], which allows to explicitly obtain an empirical model of optimal complexity. The disadvantage of this method is a considerable expenditure of computer time for the practical implementation even for relatively simple models.

The aim of this study is to develop a method of building empirical models using the ideas of genetic algorithms to predict ecological processes, which processes are largely free from the above drawbacks.

2. The principal section

During the year 2007 the observations of the water level of the Dniester river near the village of Nyzhniv, Ivano-Frankivsk Province, were carried out from April 1 to August 31. During this same period of time, the data on temperature, rainfall, average wind speed and average barometric pressure were being collected. As a result of the observations done and the experimental data set containing $N = 184$ points was obtained.

Analysis of the changes of the water level in the Dniester river shows that over a certain time there is a trend $h(t)$ that is linear and there is harmonic component $G(t)$ caused by seasonal changes in weather conditions [1], i.e.

$$\tilde{H}_t = H_t + G(t) + h(t) \tag{1}$$

where: $\tilde{H}_t$ - the current water level [cm]; $G(t)$ - harmonic component of the water level [cm]; $h(t)$ - linear trend [cm].

From relation (1), determined by the change in water level in the Dniester river, linear trend was isolated

$$h(t) = \theta_0 + \theta_1 t, \tag{2}$$

where θ_0, θ_1 - the parameters of linear trend.

The coefficients of model (2) are found using the least squares procedure. As a result of the calculations done, the following values of the coefficients of the linear trend: $\theta_0 = 308.5712$, $\theta_1 = -0.4484$ were obtained.

From numerical series $\tilde{H}_t$ we allocate the fixed component of oscillating process $G(t)$, which will be presented as the harmonic series [2] of aliquant frequencies

$$G(t) = A_0 + \sum_{j=1}^{m} \left(A_j \sin(t\omega_j) + B_j \cos(t\omega_j) \right) \tag{3}$$

where t - timing cycles, $t = 1,2,3,...,N$; A_0, A_j, B_j - the parameters of the harmonic series (3); $\omega_j = \omega_{j-1} + \Delta\omega_j$ - aliquant frequency $j = 1,2,3,\ldots$.

To be able to estimate the basis of the observations process $G(t) = \tilde{H}_t - h(t)$, it is necessary to evaluate the series parameters (3) with the required condition [2] $N \geq 3m + 1$. The sum of several harmonic series (3) where the coefficients A_0, A_j, B_j determined by the least squares procedure, and the number of harmonics and their frequencies selected so as to obtain at least some external criterion of selection is called [2] harmonic trend of optimal complexity.

The problem of optimal synthesis of the harmonic series (3) is solved in [3], which yielded the following procedure.

We shall select the function

$$g(i+p)=\sum_{j=1}^{m}\left(A_j\sin\left((i+p)\omega_j\right)+B_j\cos\left((i+p)\omega_j\right)\right) \tag{4}$$

$$g(i-p)=\sum_{j=1}^{m}\left(A_j\sin\left((i-p)\omega_j\right)+B_j\cos\left((i-p)\omega_j\right)\right). \tag{5}$$

To determine the frequency ω_j it is necessary to solve the equation

$$P_m z^m + P_{m-1}z^{m-1} + \ldots + P_1 z + P_0 = 0, \tag{6}$$

where $z=\cos\omega$.

Equation (6) has m roots that explicitly define ω_j $j=\overline{1,m}$.

So, to find the parameters A_0, A_j, B_j and ω_j of the harmonic trend we first need to determine the weights of the terms of minimizing the discrepancy

$$B=\sum_{i=m+1}^{N-m} b_i^2 \tag{7}$$

where

$$b_i = g(i+m)-\sum_{p=0}^{m-1}\alpha_p\left(g(i+p)+g(i-p)\right)+g(i-m), \quad i=\overline{m+1,N-m} \tag{8}$$

describes the accuracy with which the oscillating process is expressed through a given amount of harmonic components. In formula (8) the values g corresponding to discrete arguments should be replaced by $g_t = G(t)$.

So, let's solve the problem

$$\min_{\bar{\alpha}} : J(\bar{\alpha})=\sum_{i=m+1}^{N-m}\left(z_{i,m}-\sum_{p=0}^{m-1}\alpha_p g_{i,p}\right)^2, \tag{9}$$

where $\bar{\alpha}=\left(\alpha_0,\alpha_1,...,\alpha_{m-1}\right)^T$ - the vector of weight coefficients;

$$z_{i,m}=\bar{g}(i+m)+\bar{g}(i-m); \quad g_{i,p}=\bar{g}(i+p)+\bar{g}(i-p);$$

T - the symbol of matrices transposition.

Problem (9) can be written in matrix-vector form

$$\min_{\bar{\alpha}} : J\left(\bar{\alpha}\right) = \left(\bar{z}_m - F_m\bar{\alpha}\right)^T \left(\bar{z}_m - F_m\bar{\alpha}\right), \tag{10}$$

where:

$$\bar{z}_m = \begin{bmatrix} \bar{g}\left(2m+1\right) + \bar{g}\left(1\right) \\ \bar{g}\left(2m+2\right) + \bar{g}\left(2\right) \\ \cdots \\ \bar{g}\left(N\right) + \bar{g}\left(N-2m\right) \end{bmatrix};$$

$$F_m = \begin{bmatrix} 2\bar{g}\left(m+1\right) & \bar{g}\left(m+2\right)+\bar{g}\left(m\right) & \bar{g}\left(m+3\right)+\bar{g}\left(m-1\right) & \cdots & \bar{g}\left(2m\right)+\bar{g}\left(2\right) \\ 2\bar{g}\left(m+2\right) & \bar{g}\left(m+3\right)+\bar{g}\left(m+1\right) & \bar{g}\left(m+4\right)+\bar{g}\left(m\right) & \cdots & \bar{g}\left(2m+1\right)+\bar{g}\left(3\right) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 2\bar{g}\left(N-m\right) & \bar{g}\left(N-m+1\right)+\bar{g}\left(N-m-1\right) & \bar{g}\left(N-m+2\right)+\bar{g}\left(N-m-2\right) & \cdots & \bar{g}\left(N-1\right)+\bar{g}\left(N-2m+1\right) \end{bmatrix}$$

Minimizing the expression (10) leads to the normal Gauss equation, which in the matrix form will look as follows:

$$F_m^T F_m\bar{\alpha} = F_m^T\bar{z}_m. \tag{11}$$

From the last equation the following can be found

$$\bar{\alpha} = \left(F_m^T F_m\right)^{-1} F_m^T\bar{z}_m. \tag{12}$$

The use of the formula (12) is possible only when the vector $\bar{\alpha}$ dimension is small and the matrix $F_m^T F_m$ is well conditioned. If the condition is not satisfied, then the finding should solve the equation (11) using one of the numerical methods such as the Gauss method of backflow [4].

We have proposed a different approach for building mathematical models of oscillatory processes, based on the idea of genetic algorithms, whose essence is as follows.

The entire set of observation points of the initial value of the process or phenomenon is divided into three parts in the following proportion [5] $N_R = 0,7N$, $N_Q = 0,2N$ and $N_S = 0,1N$. For the data set $N_R + N_Q$ weight coefficients are defined α_p as a solution of linear algebraic equation (11) by the Gauss exclusion method with a choice of the main element [4]. The solution of the equation (6) with respect to variable z makes it possible to find the frequency ω_j, $j = \overline{1,m}$. Using the known frequencies ω_j on the set of points $N_R + N_Q$ it is necessary to find the parameters of the model (3) A_0, A_j and B_j.

The above problem will be solved using the genetic algorithms [6]. Let's create a chromosome of m length, in which on i spot there will be zero or 1, depending on whether the frequency ω_j is removed from the selected number of full m or left. A set of chromosomes forms a population. Of the entire population the most adapted the persons that are the most adapted, i.e. those which have the greatest (smallest) value of the fitness function. In the problem of the synthesis of models of oscillatory processes the fitness function is the combination of selection criterion [7]

$$\rho = \sqrt{n_d^2 + B^2}, \tag{13}$$

where n_d^2 - a criterion shift, which is calculated by the following formula:

$$n_d^2 = \frac{\sum_{i=1}^{N} \left(g_i(R) - g_i(S) \right)^2}{\sum_{i=1}^{N} g_i^{-2}}, \tag{14}$$

where: B - the discrepancy function is defined as (7); $g_i(R)$, $g_i(S)$ - quantities whose values are calculated according to the set of points N formula (3) and coefficients of the model (3) are found according to the sets $N_R + N_Q$ and N_S.

Thus, the problem of the synthesis of oscillating process models is to choose the initial population of such chromosomes, which provides the minimum selection criterion (14).

Genetic algorithm consists of the following steps [6]:

K1. Formation of the initial population (initialization).

At the first step of the algorithm is randomly formed a population of I individuals, each of which is a chromosome of m length.

K2. Estimation of chromosome adjustment in the population.

For each chromosome a selection criterion is calculated (14). Namely, according to the model (3) a matrix is formed

$$F = \begin{bmatrix} 1 & \sin\omega_1 & \cos\omega_1 & \sin\omega_2 & \cos\omega_2 & \dots & \sin\omega_m & \cos\omega_m \\ 1 & \sin(2\omega_1) & \cos(2\omega_1) & \sin(2\omega_2) & \cos(2\omega_2) & \dots & \sin(2\omega_m) & \cos(2\omega_m) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \sin(N\omega_1) & \cos(N\omega_1) & \sin(N\omega_2) & \cos(N\omega_2) & \dots & \sin(N\omega_m) & \cos(N\omega_m) \end{bmatrix}$$

Since each frequency ω_j corresponds to a couple of factors A_j, B_j, in the formed chromosome we shall double all 1s and zeros. If, for example, the chromosome generated in the first step was: $Ch = [1001011]$, then after doubling it looks

like: $Ch_d = [11000011001111]$. In the model (3) there is always a factor A_0, then to the chromosomes Ch_d in the first position a single gene is added and we obtain $Ch_{d0} = [111000011001111]$.

According to chromosomes Ch_{d0} obtained from matrices F we form a new matrix F_{new} by deleting those columns of the matrix F which are associated with the zeros of the chromosome Ch_{d0}. From the resulting matrix two matrices F_R and F_S will be created of $\left(N_R + N_Q\right) \times m_1$ and $N_S \times m_1$ size.

On the sets of points $N_R + N_Q$ and N_S are calculated the non-zero coefficients A_0 and B_j of model (3), by solving the normal Gauss equation

$$F_R^T F_R \overline{A}_R = F_R^T \overline{g}_R, \tag{15}$$

$$F_S^T F_S \overline{A}_S = F_S^T \overline{g}_S, \tag{16}$$

where: $\overline{A}_R = \left(A_0^{(R)}, A_1^{(R)}, B_1^{(R)}, A_2^{(R)}, B_2^{(R)}, \ldots, A_{m_1}^{(R)}, B_{m_1}^{(R)}\right)$, $\overline{A}_S = \left(A_0^{(S)}, A_1^{(S)}, B_1^{(S)}, A_2^{(S)}, B_2^{(S)}, \ldots, A_{m_1}^{(S)}, B_{m_1}^{(S)}\right)$ - vectors of model parameters, which is associated with another chromosome of the initial population and calculated using the formulas (13) and (14); $\overline{g}_R = \left(\overline{g}^{(1)}, \overline{g}^{(2)}, \ldots, \overline{g}^{\left(N_R+N_Q\right)}\right)^T$, $\overline{g}_S = \left(\overline{g}^{(1)}, \overline{g}^{(2)}, \ldots, \overline{g}^{(N_S)}\right)^T$ - vectors of experimental data on the points sets $N_R + N_Q$ and N_S.

According to a certain set of coefficients $\overline{A}_R$ and $\overline{A}_S$ model (3) on the set of N points is calculated

$$g\left(R\right) = F_{new} \overline{A}_R, \tag{17}$$

$$g\left(S\right) = F_{new} \overline{A}_S. \tag{18}$$

Using the formula (13) the selection criterion is calculated, where B is in accordance with (7) and (8). The value of selection criterion is calculated for each chromosome and the resulting values are the set ρ_i, $i = \overline{1, M}$, where M - the number of chromosomes in the population.

K3. The check procedure of the algorithm condition cease.
We define

$$\rho_{\min} = \min_{i \in M} \rho_i. \tag{19}$$

If the minimum value (19) of the selection criterion (13) does not exceed a certain positive value ε, there is the algorithm cease. An algorithm cease can also occur when its execution does not lead to improved adaptation function or when the algorithm has performed a given number of iterations.

After one of the three conditions is executed, selected of the population is the chromosome Ch^*, for which condition (19) is executed. After the operation of doubling and joining a single gene to chromosome Ch_d^* we obtain - Ch_{d0}^*. This chromosome structure defines the optimal model complexity and creates a matrix F^* so that with the original matrix F are removed the columns associated with zero genes chromosome Ch_{d0}^*. The recalculation of model parameters (3) is done on the set of all points of the original data set.

K4. Chromosomes selection.

At the second step, the calculated values of the adaptation functions selected are the chromosomes that are involved in creating the descendants for the next population. The greatest chance to create a new population of chromosomes with the best value of adjustment function, i.e. providing the minimum criteria for selection (13).

The most common methods of selection [6] are the method of roulette and the tournament selection method. In this algorithm is used the tournament method. In tournament selection all the chromosomes are divided into subgroups and with further selection from each new subgroup the chromosomes with the best adjustment are formed. The subgroups may be of arbitrary size, but most of the population is divided into subgroups of 2-3 persons in each one.

K5. The formation of a new population of descendants.

The descendants' populations are formed by two main operators - mating and mutation. It should be noted that in the genetic algorithm mating is performed almost always and mutations are rare. Chance of mating is large enough ($0,5 \le P_c \le 1$), while the probability of mutation is chosen as sufficiently small ($0 \le P_m \le 0,1$).

The operator of the mutation probability P_m changes the value of the gene in the chromosome to the opposite, i.e. from 1 to 0 or from 0 to 1. Chance of mutations P_m can be done by randomly chosen numbers from the interval [0; 0,1] for each gene and the selection for this operation of those genes for which the number drawn is less than or equal to P_m. Mutation can be made both within a pool of relatives and the descendants of the pool.

The operator of mating consists of two stages. In the first phase formed are the subgroups of z person from which the best chromosome is selected on the criterion of selection $R\left(Ch_{d0}^*\right) = \min_i R\left(Ch_{d0,i}\right)$. As a result we obtain a new population of chromosomes, which uses the operator of the second stage.

At the second stage the mating is done. For this purpose from the pool of relatives $M(k)$ randomly with probability P_c the pairs are in this way formed. From

the population of individuals is randomly selected a pair of chromosomes. Generated is a random number P_z from the interval [0,1] and if the value is not more than P_c, the mating of a pair of chromosomes occurs. Otherwise, the pair of chromosomes remains unchanged. Then, for each pair of relatives the position of a gene (locus) is drawn in a chromosome that determines the point of mating. If the chromosome of each family includes m genes, the mating point L_c is a positive integer which is less than m. Therefore, the fixation of the mating point is reduced to random integer from the interval [1, $L_c - 1$].

The effects of crossbreeding operator leads to the fact that a couple of relatives forms a new pair of descendants in the following way: the first descendant in the pair whose chromosome, which in the positions from 1 to L_c consists of the genes of the first cousin, and on the position from $L_c + 1$ to m of the genes of a second cousin; the second descendant in a pair whose chromosome, which is in the positions from 1 to L_c and consists of the genes of a second family member, and on the position from $L_c + 1$ to m the genes of the first cousin. After the mating operator is done, we go to K2.

Using the method developed in MatLab environment the program of discharging the harmonic trend $G(t)$ of aliquant frequencies was written. The maximum number of frequencies $m = 30$, the number of observation points $N = 160$ was chosen. Probability of mating was $P_c = 0,9$ and that of mutation was $P_m = 0,1$. Thus, the maximum number of coefficients of the model (3), which were defined, was $2m + 1 = 61$; among them - eight are equal to zero. The results of the program are given in Figure 1, where the "o" shows the experimental data, and the "+" - the calculation value using formula (3).

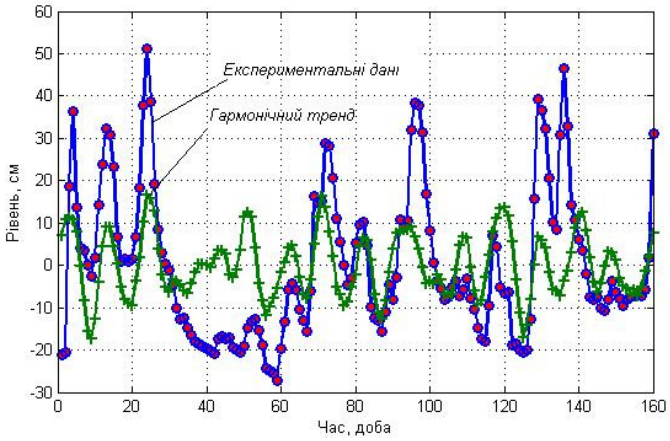

Fig. 1. The harmonic trend of the oscillating process (the Dniester river)

After the separation of the harmonic trend from the experimental data we obtained the residue (Fig. 2). The magnitude of the residue is determined from

the equation (1), provided that the pre-defined linear trend and harmonic components are as follows - $H_t = \tilde{H}_t - \left(G(t) + h(t)\right)$. The value H_t is a function of the parameters that determine the weather conditions in the area of observation, i.e.

$$H_t = \phi\left(T_t, f_{t-k}, \nu_t, p_t\right), \tag{20}$$

where: T_t - the average daily air temperature [°C]; f_t - amount of precipitation [mm/day]; t - current discrete time period [day]; k - time shift; ν_t - average wind speed [m/s]; p_t - average barometric pressure [mercury column mm].

Let's consider this process as a system whose operation is characterized by a set of input values $x_1 = T_t$, $x_2 = f_t$, $x_3 = f_{t-1}$, $x_4 = f_{t-2}$, $x_5 = f_{t-3}$, $x_6 = \nu_t$, $x_7 = p_t$ and the initial value $y = H_t$.

The values $Y_t = H_t$ are defined in discrete points of time $t = \overline{1,N}$. Input values x_j, $j = \overline{1,k}$ in each observation are taking a certain value, so that their combination forms the matrix

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_k^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_k^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ x_1^{(N)} & x_2^{(N)} & \cdots & x_k^{(N)} \end{bmatrix}$$

Relation (20) will be sought in the form of a polynomial

$$y_t = \sum_{i=0}^{M-1} a_i \prod_{j=1}^{k} \left(x_j^{(t)}\right)^{s_{ji}}, \; t = \overline{1,N}. \tag{21}$$

where: M - the polynomial number of members; a_i - polynomial coefficients; s_{ji} - degree arguments must satisfy the restriction $\sum_{j=1}^{n} s_{ij} \le m$.

The number M of the polynomial members (21) is determined by the following formula [7]

$$M = \frac{(m+k)!}{m!k!} \tag{22}$$

The system of equations (21) can be conveniently represented in matrix-vector form

$$\bar{y} = F\bar{a}, \tag{23}$$

where: $\bar{y} = \left(y_1, y_2, \ldots, y_N\right)^T$ - the calculated output value of the model (26) at each point of observation; F - the matrix of $N \times M$ size whose elements are the products of arguments for parameters a_i that

$$F = \begin{bmatrix} \prod_{j=1}^{k}\left(x_j^{(1)}\right)^{s_{j0}} & \prod_{j=1}^{k}\left(x_j^{(1)}\right)^{s_{j1}} & \cdots & \prod_{j=1}^{k}\left(x_j^{(1)}\right)^{s_{jM-1}} \\ \prod_{j=1}^{k}\left(x_j^{(2)}\right)^{s_{j0}} & \prod_{j=1}^{k}\left(x_j^{(2)}\right)^{s_{j1}} & \cdots & \prod_{j=1}^{k}\left(x_j^{(2)}\right)^{s_{jM-1}} \\ \cdots & \cdots & \cdots & \cdots \\ \prod_{j=1}^{k}\left(x_j^{(N)}\right)^{s_{j0}} & \prod_{j=1}^{k}\left(x_j^{(N)}\right)^{s_{j1}} & \cdots & \prod_{j=1}^{k}\left(x_j^{(N)}\right)^{s_{jM-1}} \end{bmatrix};$$

$\bar{a} = \left(a_0, a_1, \ldots, a_{M-1}\right)^T$ - vector of model parameters (21).

In practice, as a rule, the structure of the model (21) is unknown, which leads to the need for an arbitrary choice of both the number of features and functions in the same kind of model (21). Therefore, for the choice of the model structure (21) an inductive method of self-organizing models [2], which defines the idea of Gödel's theorem, was proposed. With regard to the problem of determining the structure of the model (21) Gödel's approach means the use of an external criterion which allows unambiguous selection of a single model from a given class of models. The definition of the external criterion was done using the new data that were not used in the synthesis of model (21). Thus, all the data obtained in the experiment are divided into two groups - the training one - N_R and the testing one - N_Q.

In most cases, for the choice of model structure the criteria of regularity are used

$$\Delta^2\left(Q\right) = \frac{\sum_{i=1}^{N_Q}\left(Y_i - y_i\right)^2}{\sum_{i=1}^{N_Q} Y_i^2} \tag{24}$$

and the minimum of bias

$$\Delta^2\left(R,Q\right) = \frac{\sum_{i=1}^{N}\left(y_i\left(R\right) - y_i\left(Q\right)\right)^2}{\sum_{i=1}^{N} Y_i^2}. \tag{25}$$

If the criterion of regularity (24) is selected, then we choose the following distribution of experimental data [5] $N_R = 0,7N$ and $N_Q = 0,3N$, and when selecting the minimum displacement criterion we use (25) - $N_R = 0,5N$ and $N_Q = 0,5N$.

The implementation of the inductive method of self-organizing models was done in two stages: first stage - generation of applicants models (in a certain order of complexity increase), the second stage - the selection of the best model according to the selection criterion (24) or (25).

After a thorough analysis of the disadvantages of the known methods of applicants - models generation a combinatorial method was selected because, unlike the other methods, it is able to present a model where the arguments are the input values of the system. Note that when creating models with the use of a combinatorial method we would have to try $2^M - 1$ models. According to the formula (22) for $n = 7$ and $m = 4$ - $M = 330$. Then $2^M - 1 = 2{,}1873 \cdot 10^{99}$ the options that are impossible to realize with today's personal computers. To remove the problem of high dimensionality a genetic approach is applied.

Using the combinatorial method of synthesis of complete polynomial model (21) we obtain an empirical model where some parameters take the zero value. The other remaining parameters will be different from zero. Let's form an ordered structure of M length, in which on i spot there will be 1 or zero depending on whether the parameter a_i $i = \overline{1, M}$ model (21) is different from zero, or is equal to zero.

Thus, the problem of synthesis of the empirical model is to select the initial population of such chromosomes, which provides the best value of the adjustment function (minimum selection criterion value (24) or (25)). The algorithm for the solution of the problem is similar to the previously developed to highlight the harmonic trend.

Based on the algorithm the program in MatLab environment was written to construct a mathematical model of the residue which was obtained after removing the linear trend and harmonic components. $m = 4$ was selected. Using the developed program the model was synthesized that accommodates 154 non-zero and 330-154 = 176 zero parameters a_i $i = \overline{0, M-1}$ of polynomial (21). The results of the program are shown in Figure 2, where the «○» marks the experimental data, and the "+" - y values that are calculated as the output of the synthesized model.

Adequacy of the model was checked by the coefficient of correlation K_{Yy} between values Y_i and its output $y^{(i)}$

$$K_{Yy} = \frac{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)\left(y^{(i)} - \overline{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2 \sum_{i=1}^{N}\left(y^{(i)} - \overline{y}\right)^2}}. \qquad (26)$$

where $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ $\overline{y} = \frac{1}{N}\sum_{i=1}^{N} y^{(i)}$ - evaluation of mathematical expectations for the values of Y_i and $y^{(i)}$. We obtained: K_{Yy} = 0.9662, indicating a high degree of correlation between the values of Y_i and $y^{(i)}$.

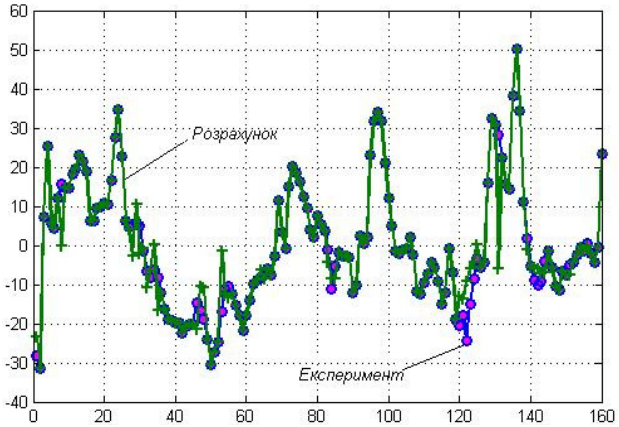

Fig. 2. Dependence of the water level on the parameters of weather conditions (after removing the linear and harmonic trends)

The dependencies found $h(t)$, $G(t)$ and y_t provide an opportunity to find

$$\tilde{H}_t = G(t) + h(t) + y_t, \tag{27}$$

where $h(t)$, $G(t)$ and y_t - calculated in accordance with equations (2), (3) and (21).

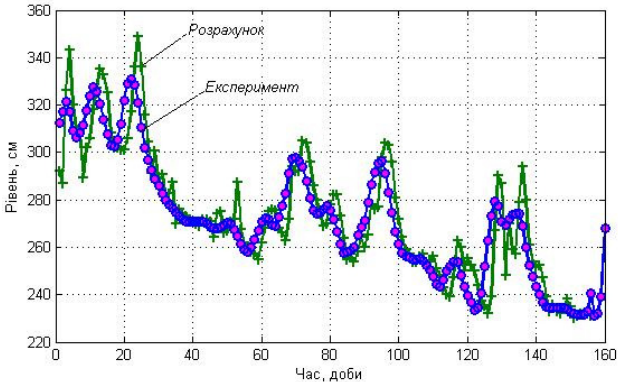

Fig. 3. Dependence of the water level in the Dniester river on the weather condition

The dependence plot (27) is shown in Figure 3, where the "+" shows the calculated values according to formula (27), and with "o" the experimental values of the water

level in the Dniester river are marked. As it is seen from the plot, there exists a quite satisfactory coincidence between the calculated and experimental data.

The importance of the prediction performed is largely determined by its quality, which depends on the degree of coincidence of the future process values with those made in advance by means of the mathematical model.

To assess the quality of the prediction for self-organizing prediction models of the input data, besides the training N_R and the testing N_Q samples, some portion was selected for the examination data sample N_S. On the N_S examination set the mean-square deviation was determined, which was calculated using the model y_i (predicted) value and the actual value Y_i:

$$\Delta^2(S) = \frac{\sum_{i-1}^{N_S}(Y_i - y_i)^2}{\sum_{i=1}^{N_S} Y_i^2} \to \min. \tag{27}$$

The above program calculates the value of the criterion based on the examination sample $N_S = 24$. The value $\Delta^2(S) = 0.0058$ proves a fairly good prediction accuracy of the Dniester river level.

Figure 4 shows the experimental values obtained on the examination set and the values according to the constructed model (28). The figure shows that the values obtained according to the prognostic model (1) are quite similar to the actual water level.

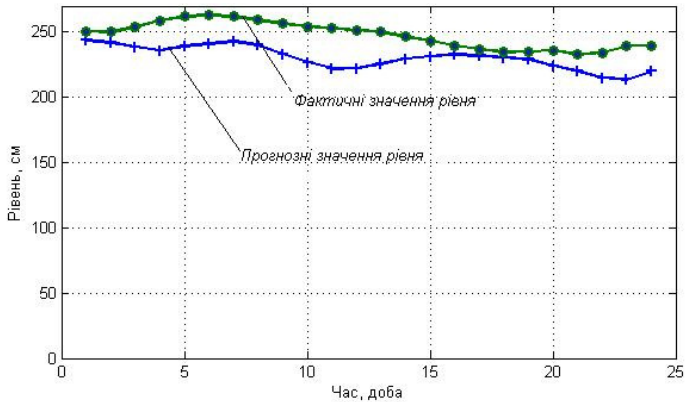

Fig. 4. The prognostic results of the water level on the examination set

Conclusions

Thus, the application of the ideas of genetic algorithms to construct a mathematical model of the change in water level in the Dniester river has made it possible

to obtain an adequate model and significantly reduce the amount of computation. The verification of the prognostic accuracy according to the empirical model on the examination set has shown that the above model is able to predict with high probability the changes in water level in the Dniester river. The later indicates that the resulting model of changes in water level in the Dniester river, depending on weather conditions, can be used to predict floods that, as the events of 2008 showed, is a very urgent task for the Carpathian region.

References

[1] Бокс Дж, Анализ временных рядов. Прогноз и управление (Дж. Бокс, Г. Дженкинс), Пер. с англ. А.Л. Левшина, Мир, М.: 1974, 406 с.

[2] Ивахненко А.Г., Индуктивный метод самоорганизации моделей сложных систем: монография (А.Г. Ивахненко), Наукова думка, К.: 1981, 296 с.

[3] Горбійчук М.І., Метод побудови математичних моделей складних процесів на засадах генетичних алгоритмів (М.І. Горбійчук, М.А. Шуфнарович), Штучний інтелект 2010, 4, 50-57.

[4] Вержбицкий В.М., Основы численных методов: учебник для вузов (В.М. Вержбицкий), Высшая школа, М.: 2002, 840 с.

[5] Ивахненко А.Г., Справочник по типовым программам моделирования (А.Г. Ивахненко, Ю.В. Коппа, В.С. Степашко и др.), под ред. А.Г. Ивахненко, Техніка, К.: 1980, 180 с.

[6] Рутковская Д., Нейронные сети, генетические алгоритмы и нечеткие системы (Д. Рутковская, М. Пилиньский, Л. Рутковский), пер. с польск. И.Д. Рудинского, Горячая линия--Телеком, М.: 2004, 452 с.

[7] Івах Горбійчук М.І., Індуктивний метод побудови математичних моделей газоперекачувальних агрегатів природного газу (М.І. Горбійчук, М.І. Когутяк, Я.І. Заячук), Нафтова і газова промисловість 2008, 5, 32-35.

[8] Горбійчук М.І., Індуктивний метод побудови математичних моделей газоперекачувальних агрегатів природного газу (М.І. Горбійчук, М.І. Когутяк, Я.І. Заячук), Нафтова і газова промисловість 2008, 5, 32-35.

[9] Ермаков С.М., Математическая теория оптимального эксперимента (С.М. Ермаков, А.А. Жиглявский), Учеб. Пособие, Наука, М.: 1987, 320 с.

[10] Ивахненко А.Г., Самоорганизация прогнозирующих моделей (А.Г. Ивахненко, Й.А. Мюллер), Техніка, К.: 1985; Берлин: ФЕБ Ферлаг Техник, 1984. - 224 с., ил. - Библиогр.: 220 с.