IRSTI 76.13.23; 15.81.00

UDC 004.94-57.090

CONSTRUCTION OF A MATHEMATICAL MODEL OF THE INFLUENCE OF VARIOUS ENVIRONMENTAL FACTORS ON THE PLANT BIOMASS ON THE CONTAMINATED SOIL BY TOXIC ELEMENTS

T. Mazakov[1], Sh. Dzhomartova[2], P. Kisala[3], Ch. Nurzhanov[4], A. Mazakova[5], B. Zhakyp[6]

[1]Institute of Information and Computational Technologies,Almaty, Kazakhstan

[3]Lublin Technical University, Poland

[1,2,4,5,6]al-Farabi Kazakh National University, Almaty, Kazakhstan

[1]tmazakov@mail.ru, [2]jomartova@mail.ru, , [3]p.kisala@pollub.pl ,

[4]DarkEremite@yandex.kz, [5]aigerym97@mail.ru, [6]zhakyp.botagoz@mail.ru

[1]ORCID ID: https://orcid.org/0000-0001-9345-5167

[2]ORCID ID: https://orcid.org/0000-0002-5882-5588

[4]ORCID ID: https://orcid.org/0000-0002-9760-3101

Abstract. The article has been devoted the development of a program for constructing a mathematical model based on the processing of experimental data for modeling plant biomass on soil contaminated with toxic elements.

In general, the problem under consideration is relevant in areas where forecasting based on analysis of data for previous points in time is required in order to take into account the relationship between the values of a certain set of factors and the behavior of an object or in the process which represented by a time series

The proposed software product for solving the problem of analyzing multidimensional time series is universal and can be applied in various fields: in ecology, medicine, chemistry, pharmacology, economics, and others. Models constructed using this method are superior to regression models in their predictive properties, due to the fact that, thanks to the application of heuristic principles of self-organization, automatic selection of informative input variables and selection of the structure of the optimal regression model. The construction of mathematical models from experimental data is carried out automatically. In the process of modeling, the main task is the identity of the form dependence and the choice of factors that have a significant impact on the dependent variable. Moreover, the structure of the model, in contrast to the regression analysis, is not fixed in advance, but is selected from a variety of options according to the absolute error criterion.

Using the self-organization program, we estimated the dependence of plant biomass (*Miscanthus*), which has a high biological absorption of heavy metals from contaminated soil from the precipitation, air temperature, potential evaporation of moisture in the soil and photosynthetic active radiation during the year, taking into account the approximation confidence value. Was obtained the mathematical models of biomass ‘plant depending from the environmental factors. The analysis showed that the predictors that have the greatest effect on plant biomass growing on soil contaminated with heavy metals are evaporation of soil moisture, photosynthetic active radiation and precipitation. The data obtained are relevant for predicting the processes of cleaning contaminated soil using plants.

Keywords: self-organization, selection, regression equation, time series, identification, heuristics, pollutants, environment, toxic elements.

Introduction

Acceleration of scientific and technological progress is largely determined by the level of application of computer technology in scientific research, in design and construction work, as well as for managing various production processes. Conducting applied research related to product quality control, process optimization, certification of complex technical products, solving sociological and economic problems, biomedical and agricultural research, research in demography, etc. often leads to the need for data analysis.

The task of analyzing time series is a very urgent problem in various fields of science and technology. Mathematical models that describe the statistical relationships between quantities, the change in time due to factors hidden from the observer, are an instrument for studying complex

systems of processes that occur in the world around us. In a large number of cases, the initial information used to build these models and evaluate their adequacy is a time series (a sequence of results of measurements of the current values of one or more parameters carried out at ordered points in time). To explain the reasons for this or that behavior of a complex system or process that gave rise to a given time series, to identify and explain the laws of their dynamics, it is necessary to solve the analysis problem. To predict the dynamics of the development of a complex system or process, as a rule, based on the results of the analysis, the task of synthesizing a time series model is solved.

In general, the problem under consideration is relevant in areas where a forecasting problem is required based on analysis of data from previous time instants in order to take into account the relationship between the values of a certain set of factors and the behavior of the object or process which represented by a time series.

The current situation necessitates the study and formalization of the processes of building, analyzing and optimizing decision support systems based on time series analysis from the point of view of the technical, algorithmic and structural aspects and confirms the relevance and scientific and practical significance of this development.

Among the tasks successfully simulated on computers, a special place is occupied by biological sciences, in particular, molecular biology aimed at genomic research [1], an agricultural complex for planning, forecasting, analysis and modeling of agricultural processes [2, 3] and ecology aimed at preventing the harmful effects of economic and other activities on natural ecological systems [4-6]. Improving the environmental information system with increasing levels of soil/water pollution by various toxic elements as a result of human activities is one of the priority areas in the field of information technology. The concentration of toxic substances in the environment is an acute environmental problem, especially in agricultural and industrial regions [7-11].

The practical value of the obtained results lies in the creation of a set of tools applicable for the construction of operational decision support systems based on the trend analysis of time series.

Software implementation

The use of the concepts and ideas of cybernetics in matters of physics, chemistry, biology, sociology, psychology and other sciences gave excellent shoots, allowed to deeply advance into the essence of the processes taking place in inanimate and living nature. There is no doubt that the coming 21st century and the progress of natural science and science throughout will proceed along the lines of studying the laws of control processes in complexly organized systems. A self-organizing system is a cognitive model of 21st century science.

One of the methods of self-organization is the method of group accounting of arguments (MGUA) [12-13]. The MGUA is based on the principle of multi-row selection of self-organization models, and the MGUA algorithms reproduce the scheme of mass selection. In the MSUA algorithms, the members of the generalized Kolmogorov – Gabor polynomial are synthesized and selected in a special way [14-15]. This synthesis and selection is carried out with increasing complication, and it is impossible to predict in advance what final form the generalized polynomial will have. First, simple pairwise combinations of the initial features are usually considered, from which the equations of the decisive functions are composed, usually not higher than the second order. Each equation is analyzed as an independent decisive function, and according to the training development in one way or another, the values of the parameters of the composed equations are found. Then, in a certain sense, the best ones are selected from the obtained set of decision functions. The selected partial decision functions are considered below as intermediate variables that serve as initial arguments for a similar synthesis of new decision functions, etc. The process of such hierarchical synthesis continues until an extreme criterion of the quality of the decision function is reached, which in practice is manifested in deterioration of this quality when trying to further increase the order of members of the polynomial relative to the original features.

Self-organization algorithms are used to solve problems of pattern recognition, predicting random processes, identifying multi-extreme static and dynamic characteristics and optimal control of complex objects.

Suppose that there is a set of input data in the form of a matrix X of n observations in the space of variable variables of dimension $m > 1$, which is characteristic of the standard problem of multiple regression. Let a training sequence of examples be formed in which each row of the matrix X is associated with a known value of the response Y, measured in a quantitative scale. It is necessary, using self-organization methods, to obtain a model expressing the law of change in the response $\boldsymbol{Y}$ depending on the specific values of the independent variables $\boldsymbol{X}$ [16].

The essence of the proposed algorithm of the multi-row heuristic method of self-organization is that it reproduces the selection scheme. Here is the full description of the object

$$y = f(x_1, x_2, ..., x_n) \tag{1}$$

replaced by several rows of private descriptions:

First row;

$$\begin{aligned} &z_1 = a_{11} * x_1 + b_{11} * x_2 + c_{11} * x_1 * x_2,\ z_2 = a_{12} * x_2 + b_{12} * x_3 + c_{12} * x_2 * x_3, \\ &..., \ z_k = a_{1k} * x_{n-1} + b_{1k} * x_n + c_{1k} * x_{n-1} * x_n \end{aligned} \tag{2}$$

where $k = n*(n+1)/2$.

Second row:

$$\begin{aligned} &\varphi_1 = a_{21} * z_1 + b_{21} * z_2 + c_{21} * z_1 * z_2,\ \varphi_2 = a_{22} * z_2 + b_{22} * z_3 + c_{22} * z_2 * z_3, \\ &..., \ \varphi_k = a_{2k} * z_{n-1} + b_{2k} * z_n + c_{2k} * z_{n-1} * z_n \end{aligned}$$

etc.

For a linear model, the coefficients C_{ij} are taken equal to zero.

Each particular description is a function of only two variables. Therefore, the coefficients of such a regression equation can be easily determined even by a small number of observations using the least squares method.

Using nonlinear private descriptions (2), models of almost any complexity can be obtained, since the degree of the polynomial doubles on each selection row.

The choice of polynomials is due to the property that, according by Weierstrass theorem [17-18], any function continuous on a finite interval can be represented with arbitrarily high accuracy as a polynomial of a certain degree.

The degree of a complete description of an object increases with each row of selection, and as a result, it is possible to determine the numerical values of the coefficients of an arbitrarily complex full description from a small number of field (experimental data).

During the selection process, variables are selected in accordance with the criteria for minimizing the absolute error functional:

$$F = \sum_{t=1}^{M} (y_t - \hat{y_t})^2, \tag{3}$$

where y_t is the value of the indicator (1) at the time t; $\hat{y_t}$ is the predicted value of the indicator at time t.

At each step of the selection, the coefficients of the regression equation are determined using the least squares method for the three arguments.

Let the function $Y = Y(U,V)$ be given by the table, that is, the numbers U_i, V_i, $Y_i (i = 1,...,n)$ are known from experience. We will look for the relationship between these data in the form:

$$Y(U,V) = a*U + b*V + c*U*V, \tag{4}$$

where a,b,c are unknown parameters.

We choose the values of these parameters so that the smallest sum of the squared deviations of the experimental data Y_i and theoretical

$$Y_i^{\wedge} = a*U_i + b*V_i + c*U_i*V_i,$$

i.e. the amount:

$$\sigma = \sum_{t=1}^{M} (Y_i^{\wedge} - a*U_i - b*V_i - c*U_i*V_i)^2 -> \min \tag{5}$$

The value σ is a function of the three variables a, b, c. A necessary and sufficient condition for the existence of a minimum of this function is the equality to zero of the partial derivatives of the function σ with respect to all variables, i.e.

$$\frac{\partial\sigma}{\partial \mathrm{a}} = 0, \frac{\partial\sigma}{\partial \mathrm{b}} = 0, \frac{\partial\sigma}{\partial \mathrm{c}} = 0 \tag{6}$$

Since

$$\begin{aligned}
\frac{\partial\sigma}{\partial \mathrm{a}} &= -2\sum_{i=1}^{n} (Y_i - aU_i - bV_i - cU_iV_i)U_i \\
\frac{\partial\sigma}{\partial \mathrm{b}} &= -2\sum_{i=1}^{n} (Y_i - aU_i - bV_i - cU_iV_i)V_i \\
\frac{\partial\sigma}{\partial \mathrm{c}} &= -2\sum_{i=1}^{n} (Y_i - aU_i - bV_i - cU_iV_i)U_iV_i
\end{aligned} \tag{7}$$

then the system for finding a, b, c will have the form:

$$\begin{aligned}
a\sum_{i=1}^{n} U_i^2 + b\sum_{i=1}^{n} U_iV_i + c\sum_{i=1}^{n} U_i^2V_i &= \sum_{i=1}^{n} Y_iU_i \\
a\sum_{i=1}^{n} U_iV_i + b\sum_{i=1}^{n} V_i^2 + c\sum_{i=1}^{n} V_i^2U_i &= \sum_{i=1}^{n} Y_iV_i \\
a\sum_{i=1}^{n} U_i^2V_i + b\sum_{i=1}^{n} U_iV_i^2 + c\sum_{i=1}^{n} U_i^2V_i^2 &= \sum_{i=1}^{n} Y_iU_iV_i
\end{aligned} \tag{8}$$

This system is solved by the Cramer method [19-20].

In accordance with the self-organization algorithm, after each series of selection, K regression equations of the form are selected:

first row $z = f(x_i, x_j)$,

second row $\varphi = f(z_i, z_j)$,

third row $v = f(\varphi_i, \varphi_j)$,

fourth row $\omega = f(v_i, v_j)$, etc.

After completing the first selection row, the first N equations with the smallest error are

selected. The resulting regression equations are denoted by $z_k = f(x_i, x_j), \mathrm{k} = \overline{1, \mathrm{N}}$.

The second and subsequent series of breeding are constructed similarly to the first.

Stopping the generation of models on subsequent rows occurs when, with an increase in the layer number, i.e., with a complication of the models, the external criterion of the best model does not decrease (3). The complication of the model ceases when further improvement of the selection criterion does not exceed a certain number ε (algorithm parameter).

The developed program is universal and can be applied in various fields to solve the problem of analyzing multidimensional time series - in space physics, seismology, medicine, finance, and others.

The program is implemented in the Delphi programming language [21-22].

Application of a software product

Next, we consider the application of the developed program to build a mathematical model of the influence of various environmental factors on the biomass of a plant, which has a high ability to absorb toxic elements from contaminated soil.

Let us consider the use of a multi-row self-organization algorithm for analyzing the relationship between the dynamics of the biomass plant (Miscanthus) and the following parameters: 1 – evaporation of soil moisture (PE), 2 – photosynthetic radiation activity (PAR), 3 – precipitation (Rainfall), 4 – temperature air (Temperature). The following information is located in the FisxP.txt file:

4 100 2
Biomass
PE
PAR
Rainfall
Temperature.

In the first line, the first number indicates the number of parameters, the second number indicates the amount of data, the third number indicates the type of model (linear – 1, nonlinear -– 2).

The FisxD.txt file contains line-by-line experimental data.

In the line, the first number indicates the value of the effective parameter, then the values of the auxiliary parameters are listed.

The result of the program is displayed in the Rezult.txt file, the contents of which are presented below.

```
            SELF-ORGANIZATION program parameters
Parametric source data
Number of arguments – 4
Amount of points –100
Initial data

     Biomass
  1 - PE
  2 - PAR
  3 - RAINFALL
  4 - Temperature
  1)     1.050 =      2.408     8.280     0.250    15.500
  2)     2.532 =      3.515     7.314     7.110    17.000
  3)     3.535 =      2.152     3.542     5.080    14.000
  . . . . . .
100)  1264.565 =      2.014     6.624     0.000    22.500
Nonlinear model
=======================================
SELF-ORGANIZATION RESULT
           Max      Min      Avg
Y     =  1264.565     1.050   544.753
```

X(1) =	4.807	0.000	2.291
X(2) =	14.030	1.460	7.782
X(3) =	38.100	0.000	5.684
X(4) =	111.000	6.500	21.242

1 row results

	Error	Coef-t-1	Coef-t-2	Coef-t-3
3(1- 4) =	16.442745	0.687433	0.639918	-1.078038
2(1- 3) =	16.641309	0.686506	0.656516	-1.019773
1(1- 2) =	16.738916	0.790674	0.763545	-1.305119
4(2- 3) =	17.036969	0.715034	0.855330	-1.359676
5(2- 4) =	17.412309	0.649833	0.670954	-1.035383
6(3- 4) =	17.424851	0.742313	0.770697	-1.275642

z1(1) = 0,687*X(1) + 0,640*X(4) - 1,078*X(1)*X(4)
z1(2) = 0,687*X(1) + 0,657*X(3) - 1,020*X(1)*X(3)
z1(3) = 0,791*X(1) + 0,764*X(2) - 1,305*X(1)*X(2)
z1(4) = 0,715*X(2) + 0,855*X(3) - 1,360*X(2)*X(3)

2 row results

	Error	Coef-t-1	Coef-t-2	Coef-t-3
3(1- 3) =	14.452810	1.618656	1.360954	-9.061547
4(1- 4) =	14.871161	1.020455	1.879339	-8.893387
7(3- 4) =	15.046972	1.463690	0.953528	-3.064651
6(2- 4) =	15.327753	1.094837	1.005669	-2.758885
5(2- 3) =	16.314729	1.077392	1.105448	-2.506673
1(1- 2) =	16.442745	1.000000	0.000000	0.000000
2(1- 2) =	16.571151	1.459143	1.204295	-4.596856

z2(1) = 1,619*Z1(1) + 1,361*Z1(3) - 9,062*Z1(1)*Z1(3)
z2(2) = 1,020*Z1(1) + 1,879*Z1(4) - 8,893*Z1(1)*Z1(4)
z2(3) = 1,464*Z1(3) + 0,954*Z1(4) - 3,065*Z1(3)*Z1(4)
z2(4) = 1,095*Z1(2) + 1,006*Z1(4) - 2,759*Z1(2)*Z1(4)

3 row results

	Error	Coef-t-1	Coef-t-2	Coef-t-3
7(3- 4) =	11.940320	0.909004	0.998533	-1.489588
6(2- 4) =	12.143334	0.783708	0.993105	-1.172293
4(1- 4) =	12.239301	0.735509	1.109891	-1.440375
5(2- 3) =	12.419342	1.026507	1.106805	-2.174982
2(1- 2) =	12.480062	1.197777	1.237831	-3.162864
3(1- 3) =	12.616615	0.925663	1.116772	-2.057170
1(1- 2) =	14.452810	1.000000	0.000000	0.000000

z3(1) = 0,909*Z2(3) + 0,999*Z2(4) - 1,490*Z2(3)*Z2(4)
z3(2) = 0,784*Z2(2) + 0,993*Z2(4) - 1,172*Z2(2)*Z2(4)
z3(3) = 0,736*Z2(1) + 1,110*Z2(4) - 1,440*Z2(1)*Z2(4)
z3(4) = 1,027*Z2(2) + 1,107*Z2(3) - 2,175*Z2(2)*Z2(3)

4 row results

	Error	Coef-t-1	Coef-t-2	Coef-t-3
1(1- 2) =	11.940320	1.000000	0.000000	0.000000
4(1- 4) =	12.802728	1.009403	1.719933	-4.853063
3(1- 3) =	12.880619	1.218947	1.385676	-4.595693
6(2- 4) =	13.302587	1.158373	1.204906	-3.611096
7(3- 4) =	14.603084	0.890151	0.937029	-1.608685

5(2- 3) =	15.269376	0.872656	1.045008	-2.020681
2(1- 2) =	15.358824	1.041783	1.226586	-3.371989

z4(1) = 0,909*Z3(3) + 0,999*Z3(4) - 1,490*Z3(3)*Z3(4)
z4(2) = 1,000*Z3(1) + 0,000*Z3(2) + 0,000*Z3(1)*Z3(2)
z4(3) = 0,784*Z3(2) + 0,993*Z3(4) - 1,172*Z3(2)*Z3(4)
z4(4) = 0,736*Z3(1) + 1,110*Z3(4) - 1,440*Z3(1)*Z3(4)

5 row results

	Error	Coef-t-1	Coef-t-2	Coef-t-3
1(1- 2) =	11.940320	1.000000	0.000000	0.000000
4(1- 4) =	14.851743	1.078460	1.163604	-3.101076
2(1- 2) =	14.876404	1.179823	1.011738	-2.886278
3(1- 3) =	14.884729	1.115208	1.117868	-3.023926
6(2- 4) =	15.421893	1.077140	1.047707	-2.720931
5(2- 3) =	15.452147	1.139382	0.944632	-2.586491
7(3- 4) =	16.021664	1.080676	0.910812	-2.341144

z5(1) = 0,909*Z4(3) + 0,999*Z4(4) - 1,490*Z4(3)*Z4(4)
z5(2) = 1,000*Z4(1) + 0,000*Z4(2) + 0,000*Z4(1)*Z4(2)
z5(3) = 1,000*Z4(1) + 0,000*Z4(2) + 0,000*Z4(1)*Z4(2)
z5(4) = 0,784*Z4(2) + 0,993*Z4(4) - 1,172*Z4(2)*Z4(4)

	Row	Error	Coef-t-1	Coef-t-2	Coef-t-3
7(3- 4) =	3	11.940320	0.909004	0.998533	-1.489588
1(1- 2) =	4	11.940320	1.000000	0.000000	0.000000
1(1- 2) =	5	11.940320	1.000000	0.000000	0.000000
6(2- 4) =	3	12.143334	0.783708	0.993105	-1.172293

Y(1) = 0,909*Z2(3) + 0,999*Z2(4) - 1,490*Z2(3)*Z2(4)
Y(2) = 1,000*Z3(1) + 0,000*Z3(2) + 0,000*Z3(1)*Z3(2)
Y(3) = 1,000*Z4(1) + 0,000*Z4(2) + 0,000*Z4(1)*Z4(2)
Y(4) = 0,784*Z2(2) + 0,993*Z2(4) - 1,172*Z2(2)*Z2(4)
End of breeding

At each iteration step, 4 optimal candidate models were selected. The best model for predicting biomass growth was obtained on the 3rd row of selection and is based on 3 initial parameters from 4. The parameter X4 – Temperature (Temperature) is excluded.

As you can see the optimal result obtained in the third row of selection:

$$Y(1) = 0,909*Z2(3) + 0,999*Z2(4) - 1,490*Z2(3)*Z2(4). \quad (9)$$

The analysis showed that the equation describing the dependence of biomass growth on evaporation of soil moisture, photosynthetic active radiation, precipitation, and temperature has a non-linear form. At the same time, air temperature does not significantly affect the process under study.

We rewrite equation (9) in the variables Y, X1, X2, X3, X4:

$$y = 1.804x_1 + 2.355x_2 + 2.319x_3 - 7.927x_1x_2 - 8.237x_1x_3 + 10.77x_2x_3 - 1.298x_1^2 - 3.452x_2^2 - 3.469x_3^2 + 38.77x_1x_2x_3 + 12.017x_1x_2^2 + 13.283x_1x_3^2 + 17.624x_2x_3^2 + 6.423x_1^2x_2 + 7.043x_1^2x_3 + 17.51x_2^2x_3 - 1.056x_1^2x_2^2 - 12.601x_1^2x_3^2 - 31.067x_2^2x_3^2 - 51.256x_1x_2^2x_3 - 66.02x_1x_2x_3^2 - 34.498x_1^2x_2x_3 + 37.15x_1^2x_2x_3^2 + 111.269x_1x_2^2x_3^2 + 56.245x_1^2x_2^2x_3 - 99.157x_1^2x_2^2x_3^2 + 1.794x_2^3 + 1.886x_3^3 - 7.714x_1x_3^3 - 10.624x_2x_3^3 - 10.052x_2^3x_3 - 6.443x_1x_2^3 + 39.674x_1x_2x_3^3 - 11.814x_1^2x_2x_3^3 + 7.428x_1^2x_3^3 + 19.486x_2^2x_3^3 - 67.499x_1x_2^2x_3^3 + 18.784x_2^3x_3^2 - 63.404x_1x_2^3x_3^2 + 23.597x_1x_2^3x_3 + 38.159x_1x_2^3x_3^3 - 11.707x_2^3x_3^3 + 5.772x_1^2x_2^3 + 53.491x_1^2x_2^3x_3^2 - 30.527x_1^2x_2^3x_3 + 57.817x_1^2x_2^2x_3^3 - 31.02x_1^2x_2^3x_3^3 \quad (10)$$

As can be seen from formula (10), the relationship between the dynamics of the biomass plant Miscanthus (Biomass) and the following parameters: 1 is evaporation of soil moisture, 2 is photosynthetic active radiation, 3 - precipitation is described by an 8th degree polynomial.

Analyzing it should be noted that the predictors that have the greatest influence on the biomass on the contaminated soil by the plant depend on the evaporation of soil moisture, photosynthetic active radiation and precipitation.

Conclusion

This article developed a construction program for constructing regression equations based on experimental data processing, which is used to derive equations of a mathematical model of the influence of various environmental factors on the biomass plant on the contaminated soil with toxic elements.

In the method of self-organization, the construction of mathematical models from experimental data is carried out automatically. In the process of modeling, the task is to identify the form of dependence and the choice of factors that have a significant impact on the dependent variable. Moreover, the structure of the model is not fixed in advance but is selected from a variety of options according to specified criteria.

The fundamental difference between the method of self-organization from regression analysis is that its goal is to achieve the minimum of the selection criterion that is appropriately chosen, and the goal of the regression analysis is to achieve the minimum mean-square error at all extreme points for a given type of regression equation. From here various results follow. The accuracy of the self-organization method implemented in the article, in contrast to the regression analysis, is higher due to the optimization of the complexity of the model. Self-organization algorithms are used to solve the problems of pattern recognition, predicting random processes, identifying multi-extreme static and dynamic characteristics and optimal control of complex objects.

The method of self-organization has great prospects in contrast to regression analysis and artificial neural networks and therefore is widely used in the world. The practical value of the obtained results lies in the creation of a set of tools applicable for the construction of operational decision support systems based on the trend analysis of time series. The task of analyzing and forecasting time series is in demand in many developing areas, such as: data mining, analysis of the relationship between economic data, forecasting of environmental processes, etc.

Using a self-organization program, we estimated the dependence of plant biomass growing on

soil contaminated with heavy metals on temperature, acidity of soil moisture evaporation, photosynthetic active radiation, and precipitation, taking into account the approximation confidence value. Mathematical models of plant biomass are obtained depending on environmental factors. The analysis showed that the predictors that have the greatest effect on the biomass of plants growing on a medium contaminated with heavy metals are evaporation of soil moisture, photosynthetic active radiation and precipitation.

The data obtained are relevant for predicting the processes of cleaning technogenic contaminated soil using plants. Models can be used to make a quantitative forecast when deciding how to clean contaminated soils with plants.

Acknowledgements

The work was carried out at the expense of grant funding for scientific research for 2018-2020 under the project AR05132044 “Development of a hardware-medical complex for assessing the psychophysiological parameters of a person”.

Funding

The authors received no direct funding for this research.

References

[1] Kane M., Brewer J. An information technology emphasis in biomedical informatics education. J Biomed Inform. 2007. 40. 67–72.

[2] Patel S., Sayyed I. U Impact of information technology in agriculture sector. JFAV. 2014. 4(2). 17–22.

[3] Zhu H., Research on Agricultural Ecological Factors Information Technology based on Internet +, Advances in Computer Science Research. 3rd International Workshop on Materials Engineering and Computer Sciences (IWMECS 2018), 2018. 78. 19–22.

[4] Rodhain F., Fallery B. ICT and Ecology: In favour of research based on the Responsibility principle. AIS Electronic Library (AISeL). Mediterranean Conference of Information Systems, 2009. Athens, Greece. 173–186.

[5] Khomyakov D. M., Iskandaryan R. A. Information technology and mathematical modeling in environmental management when implementing the concept of sustainable development, In the book: Ecological and socio-economic aspects of Russia's development in the context of global changes in the natural environment and climate. M.: Geos. 1997. 102–119.

[6] Eryomin A.L. Information Ecology – a Viewpoint. Int. J. Environ. Sci., Sections A & B. 1998. 3(4). 241–253.

[7]Panin M. S. The influence of technogenic factors and human agrochemical activity on the content of heavy metal migration in the soil-plant system. State and rational use of soils of the Republic of Kazakhstan,1998. 76–79.

[8] Kabata-Pendias A., Pendias H. Trace elements in soils and plants, M.: Mir, 1989. 439 p.

[9] Damalas C. A. Understanding benefits and risks of pesticide use, Science Research Essay, 2009. 4 (10). 945–949.

[10] Issimov N., Mazakov T., Mamyrbayev O., Ziyatbekova G. Application of fuzzy and interval analysis to the study of the prediction and control model of the epidemiologic situation, Journal of Theoretical and Applied Information Technology, 2018. 96(14). 4358–4368.

[11] Mamyrbayev O., Alimhan K., Zhumazhanov B., Turdalykyzy T., Gusmanova F. End-to-End Speech Recognition in Agglutinative Languages, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020.

[12] Ivakhnenko A. G., Zaichenko Yu. P., Dimitrov V. D. Decision making on the basis of

self-organization. M.: Soviet Radio, 1976. 280 p.

[13] Ivakhnenko V. I., Labkovsky V. A. The problem of uncertainty in decision making tasks, Kiev: Naukova Dumka, 1990, 132 p.

[14] Seber J. Linear Regression Analysis, M.: Mir, 1980, 456 p.

[15] Belov V. V., Chistyakova V. I. Modeling and forecasting of business processes using self-organization algorithms of formal descriptions, Business Informatics, 2008. 4 (06). 37–48.

[16] Ayvazyan S. A. and others. Classification and reduction of dimension, M.: Finance and Statistics, 1989, 607 p.

[17] Ilyin V. A., Sadovnichy V. A., Sendov Bl. Kh. Mathematical analysis. Beginner course. M.: Moscow State University, 1985. 662 p.

[18] Fichtenholtz G. M. Fundamentals of mathematical analysis. M.: Nauka, 1968. 1. 440 p.

[19] Kalitkin N. N. Numerical methods, M.: Nauka, 1978. 512 p.

[20] Bakhvalov N. S. Numerical methods. M.: Science, 1975. 1. 632 p.

[21] Tyukachev N. A., Rybak K. S., Mikhailova E. E. Programming in Delphi for beginners, SPb: BHV-Petersburg, 2007. 672 p.

[22] Bucknell J. M. Fundamental algorithms and data structures in Delphi, St. Petersburg: DiaSoftUP LLC, 2003. 560 p.