Медицинская
и биологическая кибернетика

M.Yu. ANTONOMOV[1], Dr (Biology), Professor,
Chief Researcher of the Laboratory of Epidemiological
Research and Medical Informatics
e-mail: antomonov@gmail.com
A.A. SHEVCHENKO[2], Dr (Medicine), Professor,
Head of Research Center of State University
"Dnipropetrovsk Medical Academy of Ministry
of Health of Ukraine"
e-mail: toxysan@ukr.net
A.A. KULAGIN[3], Ph.D (Medicine),
Professor of the Department of Hygiene and Ecology
e-mail: kulagin111188@ukr.net
[1] O.M. Marzeev Institute for Public Health of the National
Medical Academy of Sciences of Ukraine,
Popudrenko Str., 50, Kyiv-94, 02660, Ukraine
[2] Research Center of State University "Dnipropetrovsk Medical
Academy of Ministry of Health of Ukraine",
Vernadsky Str., 9, Dnipro, 49027, Ukraine
[3] Dnipropetrovsk Medical Academy of the Ministry of Health of Ukraine
Soborna Square, 4, Dnipro, 49027, Ukraine

ALGORITHMS OF MULTIFACTORIAL REGRESSION MODELING IN ECOLOGICAL AND HYGIENIC STUDIES

Introduction. One of the most urgent problems of environmental health is soil contamination by oil and petroleum products (PP) and its impact on crop productivity and population health. The main task here is to determine the acceptable or safe concentrations of PP in the soil. However, at present time there is no unified approach of PP regulating in different countries. One possible solution of the problem is mathematical modeling of the results of experimental studies. With the help of mathematical models, it becomes possible not only to describe the investigated dependencies, but also to calculate safe levels of PP in the soil.

The purpose of the paper is to develop a methodology for constructing and using multi-factor nonlinear regression models from data obtained in a real ecological and biological experiment.

Results. The article presents the results of an experimental study of the influence of one type of petroleum products — diesel fuel, when it enters the soil to germinate one of the most important crops — wheat. It is shown that the mathematical model describing the suppres

sion of wheat growth should be a multifactorial function ("time — concentration — effect"), which has a nonlinear character. An algorithm for constructing multifactorial nonlinear regression models is proposed. On the basis of experimental data, an adequate multifactor nonlinear mathematical model was developed. This made it possible to calculate the threshold concentration of diesel fuel, which does not affect the growth of wheat.

Conclusion. On the basis of the proposed algorithm from experimental data an adequate multifactor nonlinear regression model was constructed. With the help of this model, the concentration of diesel fuel that does not cause a negative effect on the growth of wheat was calculated.

Keywords: petroleum products, environmental contamination, threshold concentration, mathematical modeling, nonlinear multifactor regression models.

INTRODUCTION

One of the top problems of environmental hygiene is the soil pollution by petrochemical products and its negative impact on the productivity of agricultural crops.

World's industrial growth and the increasing of transnational traffic leads to an average 2% of world's annual growth of petrol production. About half of the world oil losses leaks into the soil. Sources of such pollution are usually — petrol extracting and petroleum refining facilities, pipelines, oil terminals, petrol stations, all types of transport, that either use or/and transport petrol and petroleum products (PP). The loss of petrol and PP at oil production, its further processing, transportation and use, reaches 45 mln m.t. per year, or almost 2–3% of the world production [1, 2]. Only in carrying out dumping/filling operations with rail tank cars they lost up to 0.36–0.44 mg/m3, automobile — up to -0.42–0.6 mg/m3 OP [3]. According to official data (2006) at Sumy region, where 50% of the national petroleum production is placed, the area of soil, polluted by oil, consists of 14.2 hectares, 5.8 ha of which — are agricultural lands [4].

Petrol and PP affect all components of ecosystems: soil, microflora, flora and fauna. However petrol and PP are subject to extremely slow and difficult biodegradable oxidation in soil environment [5]. Once in the soil, such petrol components as oil resins and asphaltenes (heavy fractions) clog all pores, stopping the movement of water and oxygen that has negative impact onto the viability of soil biota. Methane and aromatic hydrocarbons (light oil fractions: petrol, diesel, etc.) have a direct toxic effect, leading to a number of negative impacts, such as: oppression of self-purification or vegetation degradation, disruption the ecological balance in the soil ecosystem, negatively affect organisms living in soil, alter the soil structure, reduce the productivity of agricultural lands, etc. [6]. It is important to note that gasoline and diesel fuel, unlike crude oil, have global scale of distribution and use, while diesel fuel comparable to gasoline remains in soil for much longer [7].

The natural restoration of soil ecosystems contaminated by petrol and PP is a long and complicated process. Under the influence of oil pollution, soil is restored much worse than water and air environment, as it is capable to accumulate and consolidate toxic substances. It was demonstrated that the natural penetration of mineral oils into the soil as a result of accidental leakage is quite durable and can last up to 45 years or more [8].

The task of high priority in assessing the level of soil contamination in PP is to determine their permissible or safe concentration.

One of the possible solutions to the problem of establishing the sanitary rules of PP is the mathematical modeling of the results of experimental studies. Mathematical models can help not only to describe the dependencies, but also to study and calculate safe concentrations of PP in the soil.

PROBLEM STATEMENT

To date, the maximum allowable concentrations (MACs) of petroleum products in soil on the territory of the former USSR are legally adopted only in Tatarstan and Belarus. In Tatarstan, allowed MAC is 1.5 g/kg — according to the translocation level of harmfulness. In Belarus MAC of mineral oil in soil depends on the category of land: agricultural lands-50 mg/kg, localities lands, horticultures, summerhouse cooperatives -100 mg/kg, industrial lands, for transportation, communications, energy, defense and other purposes-500 mg/kg, lands of environmental, health, recreational, historical and cultural purposes, along with the lands of forest and water fund, and land reserve is 50 mg/kg [9].

Three levels of soil pollution assessment are applied in the Netherlands, i.e. background pollution with PP concentration 50 mg/kg or less, increased pollution-1000 mg/kg, high pollution-5000 mg/kg. In Germany, the permissible PP concentration in soil at water reserve zones and in nature reserves is 300 mg/kg, soils of ancient river valleys-3000 mg/kg, watersheds-5000 mg/kg. Some other West European countries take 1000 mg/kg [10] as the upper safe levels of PP in soil. Soil pollution by PP in Russia assessed by the levels of: acceptable — up to 1000 mg/kg, low — 1000–2000 mg/kg, average — 2000–3000 mg/kg, high-3000-5000 mg/kg, very high- more than 5000 mg/kg [11]. In Ukraine MACs for petrol and petroleum products in the soil are not legally adopted, while all relevant studies commonly refer to the AAC- approximately allowable concentration (0.2 mg/kg) [12].

Thus, different countries have not any common approach to environmental and hygienic regulation of PP, because of the diversity and specifics of the object of study, and traditional methodological difficulties in determining threshold levels of chemicals in the soil in general [13].

The study of the effect of different PP is sometimes complicated, as it is almost impossible to create and to consider during the experiment all variables present in real environment, particular in soil. That results in shown above diversity of values regulations sometimes substantially differ from each other. This creates significant problems in risk assessing of soil pollution by PP and, that is the most important, of relative levels of such pollution to health indicators of population living in contaminated territories.

Moreover, it is very difficult to consider during the experiment all time processes of biological and physical-chemical transformations of PP that also reduces the accuracy of obtained thresholds, despite considerable cost and duration of such studies.

Different techniques and methods of mathematical processing are currently widely used to solve a variety of medical, ecological and hygiene tasks. Computer data processing with the use of appropriate statistical packages is not only effective, but also is mandatory for any scientific research. However, as a rule in mathematical processing one uses traditional for this scientific field techniques and methods.

For clinical studies, odds ratios and formal diagnostic methods can be calculated or individual mechanisms of pathological processes such as treatment of apoplexy [14], limb diseases [15], tumor growth [16, 17], including that in the brain [18], changes in pulmonary pressure in hyperthyroidism [19], etc are described.

For full-scale data, various elements of epidemiological processes [20] are usually studied for their dependence on risk factors [21], dynamic aspects are spread epidemics [22], the spread of non-infectious diseases [23]. General methodological questions concerning the use of applied mathematical methods are mainly presented in authoritative reference publications [24–27].

Typically, the proposed mathematical method, describing the relationship, is a different version of the regression analysis. To model the actions of one predictor, we use linear or nonlinear one-factor models. In the case of nonlinear models, we mainly use logistic functions, which are linearized using the simplest methods.

If the linearization of nonlinear models is difficult or initially assumed that it will not be used at all, the most complex problem of regression analysis arises — that is, the construction of multifactor models with nonlinear regression without the linearization procedure.

The feasibility of solving such a problem is primarily due to the need of a mathematical description of complex dynamic processes, including for ecological and hygienic research.

Given the a priori sufficient complexity of the influence of PP on soil quality and its biological productivity, it can be assumed that to describe these processes it is advisable to use the apparatus of nonlinear regression analysis. Since the final result obviously depends on the concentration of the PP and its duration, the regression models, by definition, must be multifactorial.

However, the methodology for constructing models that describe real natural processes is far from being finalized.

The purpose of the paper is to develop a methodology for constructing and using multifactor nonlinear regression models from data obtained in a real ecological and biological experiment.

MULTIFACTOR NONLINEAR REGRESSION MODELS DESIGN ALGORITHM

In general terms, any mathematical model has the form:

$$\hat{y} = f\big(\{a_j\}, \{x_k\}\big),$$

where: $\hat{y}$ — function (dependent variable); $\{a_j\}$ $(j = 0, ..., m)$ — is the set of m parameters of the model; $\{x_k\}$ — vector of arguments (factors, predictors, independent variables); f — dependence operator.

The task of building a model is to select the type of function f and to calculate its parameters $\{a_j\}$ according to the array of source data $\{y_k\}$, $k = 1, K$, where K — is the number of the registration points of source data [28–30].

Consider a common algorithm of building such models, let designate impact factors in large Latin letters $A, B, C, ...$ (as in variance analysis), and the coefficients (parameters) of models — in small Latin letters: $a, b, c...$

Table 1. **Source data for two variables**

The values of the factor A	The values of factor B					
	B1	*B2*	...	*Bj*	...	*Bm*
A1	*y11*	*y12*	...	*y1j*	...	*y1m*
A2	*y21*	*y22*	...	*y2j*	...	*y2m*
...	...	...	...	...	...	...
Ai	*yi1*	*yi2*	...	*yij*	...	*yim*
...	...	...	...	...	...	...
An	*yn1*	*yn2*	...	*ynj*	...	*ynm*

To simplify, let's consider firstly the procedure of model building for only two factors A and B:

$$\hat{y} = f(\{a\},\{b\},A,B).$$

1. At the first stage all source data are collected into a single table. For two variables — A (with gradations of Ai $i = 1, n$) and B (with gradations Bj $j = 1, m$) the table usually has this form (Table 1)

2. At the second step, let's choose one of the factors (e.g. A) and build the original function dependency plots from this variable f(A) for each value of another factor (B):

$$\hat{y} = f_B(A).$$

The number of graphs is equal to the number of parameters, i.e. m. We define the type (class) of the regression model by the appearance of the graphs. The accuracy and adequacy of further modeling depends on the choosing the type of the function. Therefore, at this stage it is advisable to make full use of software tools for plotting diagrams. The most suitable way is to choose the leveling function from the exponential class with constrained models (the so-called "biological" curves).

Ideally, this function should match equally well the alignment of the plots for all values of m of the second factor. It is not excluded that the chosen type of function will eventually lead to the creation of an adequate model. If this is the case, all subsequent modeling steps must be removed and the construction process resumes using a different type of model.

3. The chosen mathematical model (the first-level model) is written analytically and its parameters are calculated in accordance with the initial data by numerical methods. Modern software provides many such methods. Therefore, if the model is chosen prudently, and the requirements for sufficient points to calculate the selected number of settings are maintained to get the desired result, this does not create any problems.

As a rule, the implementation of numerical methods involves setting initial approximations for all unknown parameters of the model. The better (closer to the final result) they are set initially, the greater the likelihood that the numerical refinement procedure will be correctly implemented. Therefore, the settings of the initial parameter values should be performed with the utmost care, using all possible expert information and graphical analysis results.

Calculation of parameters of the model is made for all m values of B factor.

Table 2. **Presentation of results for the first level modeling**

Model parameters	The values of factor B					
	B_1	B_2	...	B_j	...	B_m
a_1	a_{11}	a_{12}	...	a_{1j}	...	a_{1m}
a_2	a_{21}	a_{22}	...	a_{2j}	...	a_{2m}
...	...	...	...	...	...	...

As a result we get a table of values for all model parameters of the first level for each value of factor B (Table 2).

4. Then we construct a second-level model, that is, mathematically describes the dependence of the changes for each parameter of the first-level model when the coefficient B is changed:

$$a = f(\{b\}, B).$$

To do this, we repeat the previous steps and put the parameter values in the new tables. After that, we create new plots separately for each parameter. For these plots, we also choose approximating functions. Naturally, these functions may be different for different graphs, including non-linear functions. Then, for each parameter a, the coefficients b of the corresponding dependencies are calculated.

5. Then we assemble an entire model. Two-tier model is written down analytically. Instead of coefficients in a symbolic form we fill in their numeric values. These values are considered to be initial approximations of the parameters of the resulting model.

According to them we make corrections with the help of numerical methods and model acquires the final form. At this stage we verify the accuracy of the model, i.e. the accordance of calculated values to source data. For this purpose we select the corresponding criterion and perform necessary calculations.

At this stage we perform graphical representation of the results. For two-factor models with the help of software we build function graphs in 2D and surfaces in 3D. We select the position of the surface in three-dimensional space, select colors and build cross sections of that surface in all directions, etc.

6. If a model satisfies formal and substantive criteria, one may start working with it: to calculate the projection of function changes when modifying factors (direct task simulation), to find the "critical" (threshold) value factors when values function are set (inverse problem), to identify areas of change of some values while other factors are fixed, etc.

For a larger number of factors the procedure of nonlinear simulation is mostly the same. Stages of building models of a third (or more) orders are added. That is, for three factors it is necessary to build plots of changes in parameters b of second order model depending on factor C. Then one should calculate values of parameter c and write down the model in general terms.

PRACTICAL IMPLEMENTATION OF THE ALGORITHM
IN ECOLOGICAL AND HYGIENIC TASKS

In accordance with the task we investigated the impact of PP, namely, diesel fuel (DF), on the development of wheat germs in the possible soil contamination.

It is known that physical-chemical characteristics in particular, their insolubility in water, determine the impossibility of their assimilation by crops root system. Therefore, priority and the most dangerous for vegetation is their phyto-toxic effect that leads to suppression of plant growth and its further death.

In the experiment we used classic methodology of phyto-toxic measurement according MR 2609-82 that involved seeds pre-sprouting in Petri dishes with different quantities of DF and thin layer of soil (50 grams in each cup) [31]. In assessing phyto-toxic effect of DF, we considered the dynamics of germination, crop germination rate and the length of the plants roots.

With the help of eyedropper 25, 50, 75, 100, 200 and 300 mg of DF were evenly put into each dish that makes 0.5, 1, 1.5, 2.0, 4.0 and 6.0 g/kg for 50 g of dry soil. In every dish 30 winter wheat seeds were sown, covered with a 50 g layer of ordinary black-earth soil, brought to 60% of total moisture content. Then on the 3-rd, 4-th, 5-th, 6-th and 7-th day the percentage of germination of plants sprouts was counted.

Thus, the initial function y represented itself the number of seeds germinated (in %), the first obvious factor A — time of observation (day), the second factor B — DF concentration (g/kg). Source data are given in Table 3, dynamics graphs presented in Fig. 1.

The most suitable mathematical form of description of this kind of dynamics (first level model) is the exponential function with the limitation of growth:

$$\hat{y} = f(\{a\}, A) = y(A) = a_1(1-exp(-a_2*(A - a_3)),$$

where $\hat{y}$ — is the theoretical value of the percentage of sprouted seeds, a_1 — is its limit (scale parameter), a_2 — is the speed parameter, A — is the observation time, and a_3 — is the possible time shift before the germination begins (shift parameter).
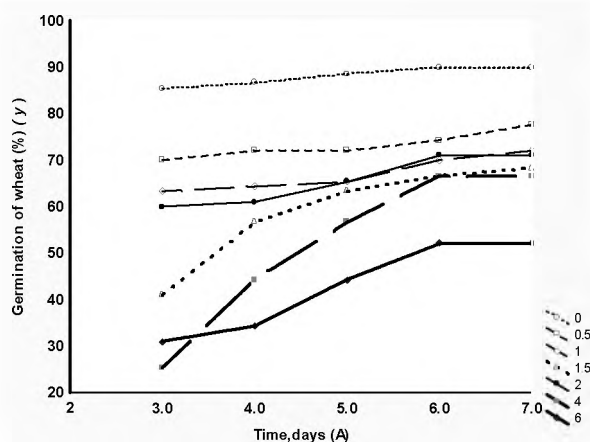

Fig. 1. An experimental graph showing the dynamics
of germination of wheat seeds

Table 3 **Germination of wheat (%) with application of DF**

Time (days)	The concentration of DF (g/kg)						
	0(control)	0.5	1	1,5	2	4	6
3	82.0	82.0	61.0	51.0	44.3	48.6	12.0
4	84.3	84.3	80.0	58.6	57.6	56.6	17.6
5	84.3	84.3	80.0	58.6	57.6	56.6	17.6
6	87.6	90.0	85.3	64.3	62.0	57.6	36.6
7	88.6	90.0	85.3	64.3	62.0	57.6	46.6

Table 4. **The results of calculating the statistical characteristics of the first level mathematical model**

Concentration DF (g/kg)	Statistical characteristics				
	parameter	average	S_a	t_a	p_a
0	$a_1 *10^{-2}$	0.89	0.006	147.45	< 0.001
	a_2	0.97	0.079	12.32	0.001
	a_3	0	-	-	-
0.5	$a_1 *10^{-2}$	0.75	0.015	51.16	< 0.001
	a_2	0.82	0.138	5.96	0.009
	a_3	0	-	-	-
1	$a_1 *10^{-2}$	0.70	0.021	33.06	< 0.001
	a_2	0.68	0.128	5.36	0,013
	a_3	0	-	-	-
1.5	$a_1 *10^{-2}$	0.69	0.003	259.48	< 0.001
	a_2	0.78	0.023	34.13	0.001
	a_3	1.85	0.032	57.65	< 0.001
2	$a_1 *10^{-2}$	0.74	0.080	9.35	0.068
	a_2	0.50	0.561	0.89	0.534
	a_3	0.66	2.931	0.22	0.857
4	$a_1 *10^{-2}$	0.74	0.051	14.40	0.005
	a_2	0.52	0.131	4.03	0.056
	a_3	2.22	0.188	11.82	0.007
6	$a_1 *10^{-2}$	0.54	0.044	12.40	0.051
	a_2	0.76	0.411	1.86	0.314
	a_3	2.73	0.583	4.68	0.134

The values of the model parameters: average value, error S, Student's coefficient t, accuracy p, are calculated by numerical method of Levenberg-Marquardt using the package STATISTICA 10.0 (Table 4).

Then we examined the dependencies of these parameters (a_1, a_2, a_3) according to concentration of B. The original plot of changes of parameter a_1 appears as follows (Fig. 2).

It can be seen that generally the dependence has a decreasing trend. Mathematical model of the second level for parameter a_1 can be represented by decreasing hyperbolic function with a shift:
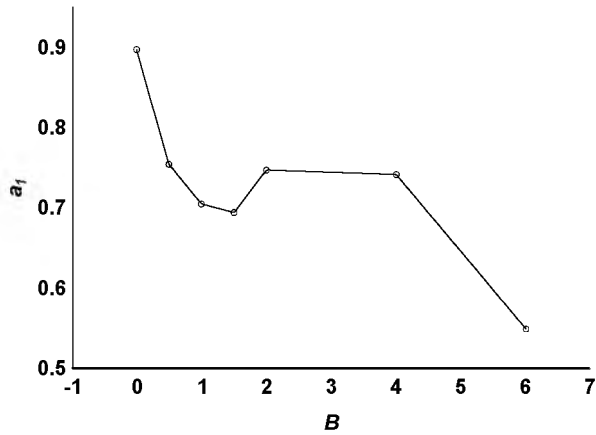
$$a_1(B) = b_1 / (B + b_2)^{\wedge}b_3.$$

Fig. 2. Plot of changes a_1 (B)

Table 5. **Statistical characteristics of model parameters $a_1(B)$**

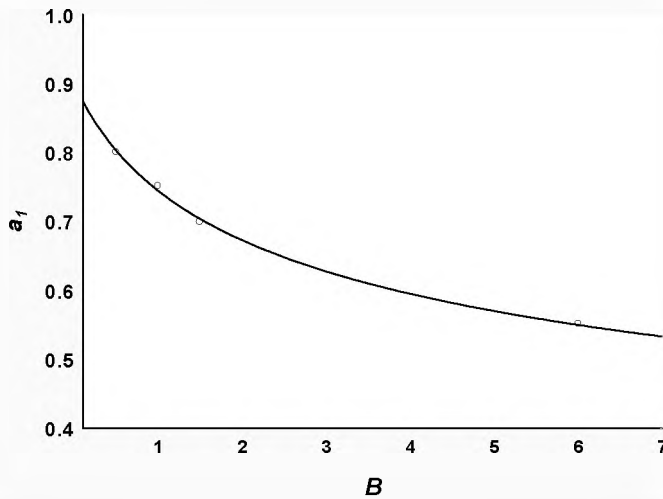| Parameters	Statistical characteristics			
	average	S_b	t_b	p_b
b_1	0.84	0.019	43.55	0.001
b_2	0.75	0.099	7.64	0.017
b_3	0.22	0.014	16.54	0.004

Fig. 3. Mathematical model of dependencies $a_1(B)$

The results of calculations by numerical methods made with the use of STATISTICA 10.0 software are listed in Table 5. All parameters are valid at the high significance level that demonstrates the adequacy of the chosen model. The plot of this function is shown in Fig. 3.
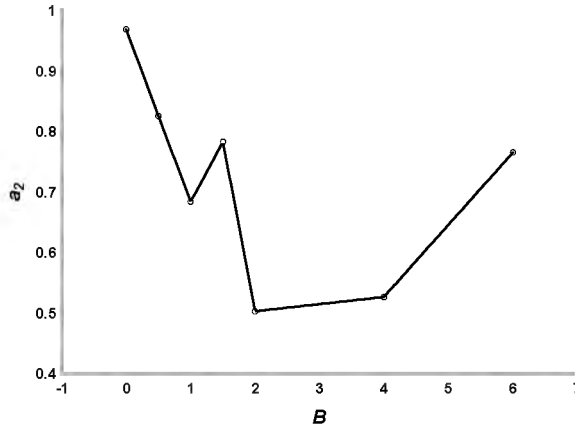
Fig. 4. The original graph of $a_2(B)$ changes

Table 6. **Statistical characteristics of model parameters a $_2$(B)**

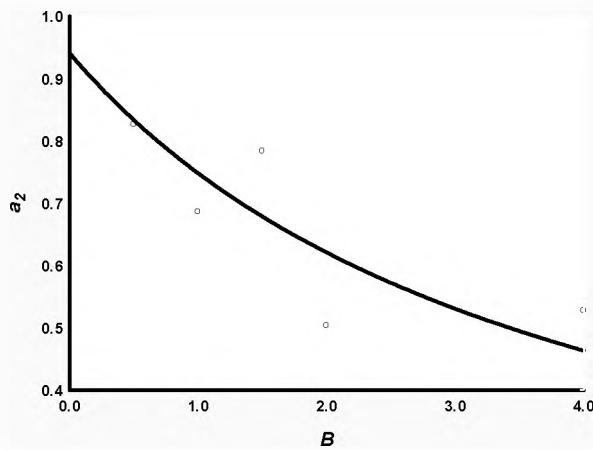| Parameters	Statistical characteristics			
	average	S_b	t_b	p_b
b_4	3.64	1.059	3.44	0.026
b_5	3.87	1.344	2.88	0.045

Fig. 5. Mathematical model of dependencies $a_2(B)$

The original graph of speed parameter (a_2) depending on concentration (B) is shown in Fig. 4. It can be seen that generally, except for the last point, the dependency can be described by decreasing hyperbolic function with a shift:

$$a_2(B) = b_4 / (B + b_5).$$

The results of the calculation of the parameters are given in Table 6. The values of the parameters are statistically reliable, hence the model is correctly constructed. The plot of the theoretical curve corresponding to this model is shown in Fig. 5.
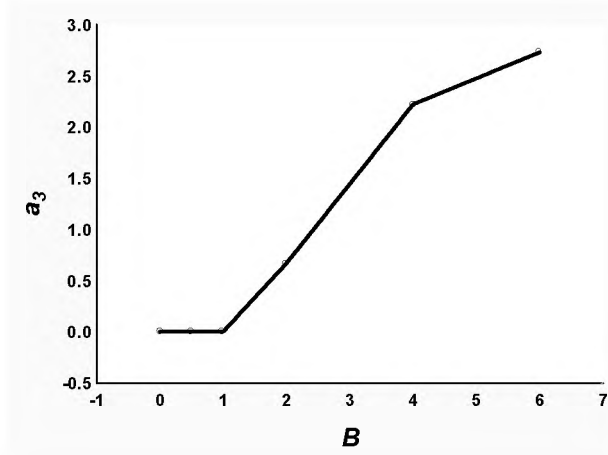
Fig. 6. The original graph of $a_3(B)$ changes

Table 7. **Statistical characteristics of model parameters $a_3(B)$**

Parameters	Statistical characteristics			
	average	S_b	t_b	p_b
b_6	2.47	0.195	12.72	0.006
b_7	1.62	0.492	3.30	0.081
b_8	0.99	0.067	14.80	0.005

Source graph of shift a_3 parameter depending on DF concentration (B) is presented in Fig. 6.

The most appropriate function for describing this kind of dependency is the exponential function with a shift and restriction of growth:

$$a_3(B) = b_6(1 - exp(-b_7(B + b_8))).$$

The results of calculating the model parameters are given in Table 7. The function parameters are statistically reliable or tend to be reliable, therefore the model can be considered adequate.

Thus, all parameters of the model "DF concentration — time — effect" to describe the germination of wheat were obtained, and the model has the following general form:

$$\hat{y} = b_1/ ((B + b_2)^{\wedge}b_3)*\{1 - exp[- (b_4 / (B + b_5)) *$$
$$* (A - (b_6* (1 - exp(- b_7(B +b_8)))))]\}.$$

We determine the error of the model $d(\%)$ according to the formula of relative standard deviation:

$$d\% = \sqrt{\frac{\sum_{k=1}^{K}[(y_k - \hat{y}_k)/ \hat{y}_k]^2}{K-1}}100.$$

The error of the model is $d = 14.3\%$, which indicates a rather accurate description of the initial data with the help of the developed mathematical model.
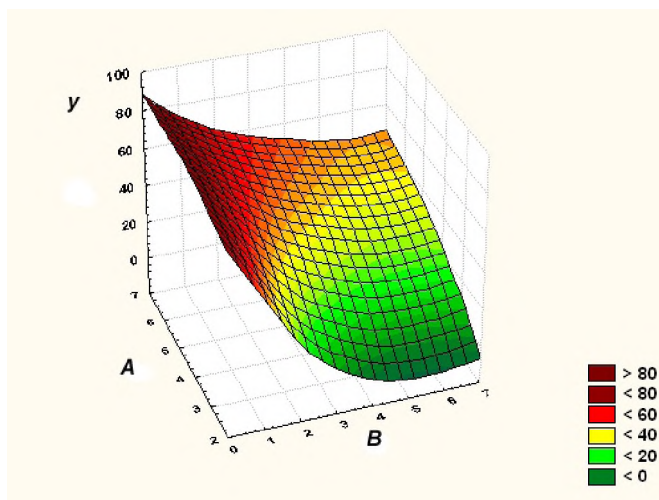
Fig. 7. Plot of dependencies "time (A) — concentration (B) — effect ($\hat{y}$)" of germination of wheat under influence of DF

The chart of a model function in 3D is shown in Fig. 7.

According to the guidelines (MP No. 2609-82) the minimum active concentration of PP should be considered such a concentration that causes growth inhibition of test seedlings plants by at least 20% compared to control [31].

With the help of the model we have calculated a threshold concentration of DF, which is equal to 1.2 g/kg. The theoretical value was higher than the experimental concentration of 1.0 g/kg, where test germination did not differ from control for 20%, and less than 1.5 g/kg, where this effect has already been observed.

Thus, using the proposed algorithm for the initial data, we constructed a multifactorial model of nonlinear regression, which quite accurately describes the results of the experimental study. Using this model, we calculated the threshold concentration of DF, which was an intermediate value between experimentally active and inactive concentrations.

CONCLUSIONS

It was shown that the methodological difficulties in justifying threshold levels of one of priority pollutants of soil — petrol and its derivatives can be overcome by using mathematical simulation of results in the experimental study.

Offered algorithm that consists of a hierarchical procedure of graphical analysis, the choice of the class of functions and numerical calculation of parameters enables the design of the most complex type of regression models — multifactor nonlinear function.

According to the proposed algorithm, a mathematical model was constructed to describe the results of an experimental study of the effect of DF on the germination of wheat.

On the basis of the constructed model, the critical (threshold) concentration of DF (1.2 g / kg) was calculated, that does not negatively affect the germination of plants as compared to the control.

REFERENCES

1. Tyuleneva V.A., Solyanik V.A., Vaskina I.V., Shalugin V.S. Regarding oil filtration in soils. *Bull. of the KSPU*. Iss. 2. Part 2. 2006. P. 110–112. (in Russian).
2. Rogozina E.A. Topical issues on cleaning of oil-contaminated soils. *Oil and Gas Geology. Theory and practice (1)*. 2006. P. 1–10. (in Russian).
3. Bilonenko G.M. Changes in soil fertility under hydrocarbon contamination. *Bull. of Agrarian Science*. 2002. No. 10. P. 52–54. (in Ukrainian).
4. Abramov Yu.A., Grinchenko E.N., Kirochkin A.Yu. Emergency monitoring. Kharkiv, 2005. 530 p. (in Ukrainian).
5. Boychenko S.V. Rational use of hydrocarbon fuels. Monograph. Kyiv, 2001. 216 p. (in Ukrainian).
6. Oborin A.A. Oil-contaminated biogeocinoses. Monograph. Ural Branch of the RAS. Perm, 2008. 511 p. (in Russian).
7. Shevchenko A.A., Kulagin A.A. The Study of petroleum stability of Petroleum Products in black-soil. Materials of a Scientific-Practical Conference with International Participation "Preventive Medicine: Achieving the Present and Looking to the Future", (May 19–20, 2016, Dnipropetrovsk). Dnipropetrovsk, 2016. P. 189–190. (in Ukrainian).
8. Solntseva N.P. Oil production and geochemistry of natural landscapes. Moskow, 1998. 405 p. (in Russian).
9. On approval of maximum permissible concentrations of petroleum products in land (including soils) for various land categories. Ministry of Health of the Republic of Belarus, 2012. No. 17/1. (in Russian).
10. Rekultivatsiya zemel, zagryaznennykh neftyu i nefteproduktami https://ecodelo.org/9935- 71_rekultivatsiya_zemel_zagryaznennykh_neftyu_i_neftepro-duktami-rekultivatsiya_zemel (in Ukrainian).
11. Procedure for determining the size of damage from pollution of land with chemical substances (approved by Roskomzem on November 10, 1993 and the Ministry of Natural Resources of the Russian Federation on November 18, 1993). (in Russian).
12. Methodology of damage determination caused by pollution and clogging of land resources in consequence of violations of environmental legislation. Ministry of Environmental Protection and Nuclear Safety, Kyiv, 1998. (in Ukrainian).
13. Goncharuk E.I., Sidorenko I.G. Hygienic valuation of chemical substances in soil. Moskow, 1986. 320 p. (in Russian).
14. Lu M., Tilley B.C. Use of odds ratio or relative risk to measure a treatment effect in clinical trials with multiple correlated binary outcomes: data from NINDS t-PA stroke trial. *Statist. Med.* 2001. Vol.20. P. 1891–1901.
15. Yang Jun-Yuan, Chen Yuming, Zhang Feng-Qin. Stability analysis and optimal control of a hand-foot-mouth disease (HFMD) model. *Journal of Applied Mathematics and Computing*. 2013. Vol. 41 T. 1-2. P. 99–117.
16. Xu Shihe, Chen Yinhui, Bai Meng. Analysis of a time-delayed mathematical model for solid avascular tumor growth under the action of external inhibitors. *Journal of Applied Mathematics and Computing*. 2016. Vol. 52. T. P. 403–415.
17. Hubbard M., Jones P.F., Sleeman B.D. The Foundations of a Unified Approach to Mathematical Modeling of Angiogenesis. *International Journal of Advances in Engineering Sciences and Applied Mathematics*. 2009. 1: 43.
18. Prouvée L., Fernández-Cara E., Optimal control of mathematical models for radiotherapy of gliomas: the scalar case. *Computational and Applied Mathematics*. 2016. Vol. 69. P. 1–18.
19. Introduction to the Use of Regression Models in Epidemiology Ralf Bender, Cancer Epidemiology. v.1. 2009. P. 179–195.
20. Zhonghua Zhang, Yaohong Suo. Qualitative analysis of a SIR epidemic model with a saturated treatment rate. *Journal of Applied Mathematics and Computing*. 2010 (34). T. 1–2. P. 177–194.
21. Bull S.B.Regression models for multiple outcomes in large epidemiological studies. *Static Med.* 1998. Vol. 17. P. 2179–2197.

22. Waleed M., Imran M., Khan A. Stochastic Analysis of an Influenza Epidemic Model. *International Journal of Applied and Computational Mathematics*. 2017 Vol. 3.T.2. P. 425–443.

23. Boutayeb A., Chetouani A., Achouyab A., Twizell E.H. A nonlinear population model of diabetes mellitus. *Journal of Applied Mathematics and Computing*. 2006 Vol. 21. T 1–2. P. 127–139.

24. Biostatistics for Medical and Biomedical Practitioners, 1st Edition. Academic Press, 2015. 770p.

25. Biomedical Informatics. Computer Application in Health Care and Biomedicine, 4th edition. Editors: E. H. Shortliffe, J. J. Cimino. NY, 2014. 965 p.

26. James S.L., Zongyuan M., Xie G.Y. Computational and Statistical Methods for Analyzing Big Data with Applications in 1st Edition. Academic Press, 2015. 206 p.

27. Bhattacharya P.K., Burman P. Theory and Methods of Statistics .1st Edition. Academic Press. 2016. 544 p.

28. Antonomov M.Yu. Mathematical processing and analysis of medical and biological data. Kiev, 2006. 588 p. (in Russian).

29. Pomerantsev A.L. Methods of nonlinear regression analysis for modeling the kinetics of chemical and physical processes Thesis for the degree of Doctor of Physical and Mathematical Sciences. Moscow: 2003. 304 p. (in Russian).

30. Ayvazyan S.A., Yenyukov IS, Meshalkin LD Applied statistics: Investigation of dependencies. Moscow, 1985. 487p. (in Russian).

31. Methodological recommendations for the hygienic substantiation of the MPC of chemical substances in soil No. 2609-82. (Effective from 05.08.1982). Moscow, 1982. 57 p. (in Russian).

ЛИТЕРАТУРА

1. Тюленева В.А., Соляник В.А., Шалугин В.С. К вопросу исследования фильтрации нефти в почвах. *Вісн. КДПУ*. Вип. 2. Ч. 2. 2006. С. 110–112.

2. Рогозина Е.А. Актуальные вопросы проблемы очистки нефтезагрязненных почв. *Нефтегазовая геология. Теория и практика (1)*. 2006. С. 1–10.

3. Білоненко Г.М. Зміни родючості ґрунту при вуглеводневому забрудненні. *Вісн. аграрної науки*. 2002. №10. С. 52–54.

4. Абрамов Ю.О., Грінченко Є.М., Кірочкін О.Ю. Моніторинг надзвичайних ситуацій. Харьків, 2005. С. 530.

5. Бойченко С.В. Раціональне використання вуглеводневих палив. Монографія. Киев, 2001. С. 216.

6. Оборин А.А. Нефтезагрязненные биогеоцинозы. Монография. УрО РАН. Пермь, 2008. С. 511.

7. Шевченко О.А. Кулагін О.О. Вивчення стабільності нафтопродуктів у чорноземі типовому. Матеріали науково-практичної конференції з міжнародною участю «Профілактична медицина: здобутки сьогодення та погляд у майбутнє»19–20 травня, 2016, Дніпропетровськ). Дніпропетровськ, 2016 р. С. 189–190.

8. Солнцева Н.П. Добыча нефти и геохимия природных ландшафтов. Москва, 1998. С. 405.

9. Об утверждении предельно допустимых концентраций нефтепродуктов в землях (включая почвы) для различных категорий земель. Министерство здравоохранения Республики Беларусь. 2012. №17/1.

10. Заходи з рекультивації ґрунтів, забруднених нафтою та нафтопродуктами, необхідно призначити з урахуванням санітарно-гігієнічних норм та оцінки екологічної обстановки території. URL: http://ekologprom.com/osnovi-prirodooblashtuvannya-ta-zaxistu-navkolishnogo-seredovishha/92-zaxodi-z-rekultivacii-gruntiv-zabrudnenix-naftoyu.html.

11. Порядок определения размеров ущерба от загрязнения земель химическими веществами (утв. Роскомземом 10ноября 1993 г. и Минприроды РФ 18 ноября 1993 г.).

12. Методика визначення збитку, обумовленого забрудненням і засміченням земельних ресурсів в результаті порушення природоохоронного законодавства. Міністерство охорони навколишнього природного середовища і ядерної безпеки, Київ, 1998.
13. Гончарук Е.И., Сидоренко И.Г. Гигиеническое нормирование химических веществ в почве. Москва, 1986. 320 с.
14. M. Lu, B. C. Tilley. Use of odds ratio or relative risk to measure a treatment effect in clinical trials with multiple correlated binary outcomes: data from the NINDS t-PA stroke trial. *Statist. Med.* 2001. Vol. 20. P. 1891–1901.
15. Jun-Yuan Yang, Yuming Chen, Feng-Qin Zhang. Stability analysis and optimal control of a hand-foot-mouth disease (HFMD) model. *Journal of Applied Mathematics and Computing.* 2013. Vol. 41. T. 1–2. P. 99–117.
16. Xu Shihe, Chen Yinhui, Bai Meng . Analysis of a time-delayed mathematical model for solid avascular tumor growth under the action of external inhibitors. *Journal of Applied Mathematics and Computing.* 2016. Vol. 52. T. P. 403–415.
17. Hubbard M., Jones P.F., Sleeman B.D. The foundations of a unified approach to mathematical modeling of angiogenesis. *International Journal of Advances in Engineering Sciences and Applied Mathematics.* 2009. 1:43.
18. Fernández-Cara E., Prouvée L. Optimal control of mathematical models for the radiotherapy of gliomas: the scalar case. *Computational and Applied Mathematics.* 2016. Vol. 69. P. 1–18.
19. Introduction to the Use of Regression Models in Epidemiology Ralf Bender, Cancer Epidemiology. V. 1. 2009. P. 179–195.
20. Zhonghua Zhang, Yaohong Suo. Qualitative analysis of a SIR epidemic model with saturated treatment rate. *Journal of Applied Mathematics and Computing.* 2010. Vol. 34, T. 1–2. P. 177–194.
21. Bull S.B. Regression models for multiple outcomes in large epidemiologic studies. *Statist. Med.* 1998. Vol.17. P. 2179–2197.
22. Waleed M., Imran M., Khan A. Stochastic Analysis of an Influenza Epidemic Model. *International Journal of Applied and Computational Mathematics.* 2017. Vol. 3. T.2. P. 425–443.
23. Boutayeb A., Chetouani A., Achouyab A., Twizell E.H. A non-linear population model of diabetes mellitus. *Journal of Applied Mathematics and Computing.* 2006. Vol. 21. T. 1–2. P. 127–139.
24. Hoffman J. Biostatistics for Medical and Biomedical Practitioners ,1st Edition. Academic Press, 2015. 770 p.
25. Biomedical Informatics. Computer Application in Health Care and Biomedicine, 4th edition. Editors: E. H. Shortliffe, J. J. Cimino. NY, 2014. 965 p.
26. James S.L., Zongyuan M., Xie G.Y. Computational and Statistical Methods for Analysing Big Data with Applications 1st Edition. Academic Press, 2015. 206 p.
27. Bhattacharya P.K., Burman P. Theory and Methods of Statistics .1st Edition. Academic Press, 2016. 544 p.
28. Антомонов М.Ю. Математическая обработка и анализ медико-биологических данных. Киев, 2006. 558с.
29. Померанцев А.Л. Методы нелинейного регрессионного анализа для моделирования кинетики химических и физических процес сов. Диссертация на соискание ученой степени доктора физико-математических наук. Москва, 2003г. 304 с.
30. Айвазян С.А, Енюков И.С., Мешалкин Л.Д Прикладная статистика: Исследование зависимостей. Москва, 1985. 487с.
31. Методические рекомендации по гигиеническому обоснованию ПДК химических веществ в почве №2609-82. (Действующие от 05.08.1982). Москва, 1982. 57 с.

М.Ю. Антомонов[1], д-р.біол. наук, проф.,
голов. наук. співроб. лаб. епідеміологічних досліджень
і медичної інформатики,
e-mail: antomonov@gmail.com
О.А. Шевченко[2], д-р мед. наук, проф.,
зав. Наукового центру Державного університету
«Дніпропетровська медична академія МОЗ України»
e-mail: toxysan@ukr.net
О.О. Кулагін[3], канд.мед. наук,
проф. кафедри гігієни і екології,
e-mail: kulagin111188@ukr.net
[1] ДУ «Інститут громадського здоров'я
ім. А.Н. Марзєєва НАМН України »,
вул. Попудренко, 50, м. Київ, 02660, Україна
[2] Науковий центр Державного університету
«Дніпропетровська медична академія МОЗ України»,
вул. Вернадського, 9, м. Дніпро, 49027, Україна
[3] «Дніпропетровська медична академія МОЗ України»
Соборна пл., 4, м. Дніпро, 49027, Україна

АЛГОРИТМИ БАГАТОФАКТОРНОГО РЕГРЕСІЙНОГО МОДЕЛЮВАННЯ В ЕКОЛОГО-ГІГІЄНІЧНИХ ДОСЛІДЖЕННЯХ

Запропоновано алгоритм побудови багатофакторних нелінійних регресійних моделей, який складається з ієрархічно організованих етапів: графічного аналізу, вибору виду часткових математичних моделей, чисельного розрахунку їх параметрів і подальшої згортки в загальну математичну модель. Представлено результати експериментального дослідження впливу на схожість найважливішої сільськогосподарської культури — пшениці, одного з видів нафтопродуктів (дизельного палива) за умови його попадання у грунт. Показано, що математична модель, яка описує процес пригнічення росту пшениці, повинна бути багатофакторною функцією («час — вплив — ефект»), що має нелінійний характер. Розроблено та наведено алгоритм розрахунку параметрів для багатофакторної нелінійної функції. За експериментальними даними на основі запропонованого алгоритму було побудовано адекватну багатофакторну нелінійну математичну модель і розраховано всі її параметри. За її допомоги встановлено порогову величину концентрації дизельного палива, яка не впливає на схожість пшениці.

Ключові слова: нафтопродукти, забруднення навколишнього середовища, порогові концентрації, математичне моделювання, нелінійні багатофакторні регресійні моделі.

М.Ю. Антомонов[1], д-р биол. наук, проф.,
глав. науч. сотр. лаб. эпидемиологических исследований
и медицинской информатики,
e-mail: antomonov@gmail.com

А.А. Шевченко[2], д-р мед. наук, проф., зав. Научным
центром Государственного университета
«Днепропетровская медицинская академия МЗ Украины»,
e-mail: toxysan@ukr.net

А.А. Кулагин[3], канд.мед. наук,
проф. кафедры гигиены и экологии
e-mail: kulagin111188@ukr.net

[1] ГУ «Институт общественного здоровья
им. А.Н. Марзеева НАМН Украины»,
ул. Попудренко, 50, г. Киев, 02660, Украина

[2] Научный центр Государственного университета
«Днепропетровская медицинская академия
МЗ Украины», ул. Вернадского, 9, г. Днепр, 49027, Украина

[3] «Днепропетровская медицинская академия
МЗ Украины», Соборная площадь, 4, г. Днепр, 49027, Украина

АЛГОРИТМЫ МНОГОФАКТОРНОГО РЕГРЕССИОННОГО МОДЕЛИРОВАНИЯ В ЭКОЛОГО-ГИГИЕНИЧЕСКИХ ИССЛЕДОВАНИЯХ

Предложен алгоритм построения многофакторных нелинейных регрессионных моделей, состоящий из иерархически организованных этапов: графического анализа, выбора вида частных математических моделей, численного расчета их параметров и последующей свертки в общую математическую модель. Представлены результаты экспериментального исследования влияния одного из видов нефтепродуктов — дизельного топлива, при его попадании в почву на всхожесть важнейшей сельскохозяйственной культуры — пшеницы. Показано, что математическая модель, описывающая процесс угнетения роста пшеницы, должна представлять собой многофакторную функцию («время — воздействие — эффект»), имеющую нелинейный характер. Разработан и представлен алгоритм расчета параметров многофакторной нелинейной функции. По експериментальным данным на основе предложенного алгоритма была построена адекватная многофакторная нелинейная математическая модель и рассчитаны все ее параметры. С ее помощью установлена пороговая величина концентрации дизельного топлива, не влияющая на всхожесть пшеницы.

Ключевые слова: *нефтепродукты, загрязнение окружающей среды, пороговые концентрации, математическое моделирование, нелинейные многофакторные регрессионные модели.*