New Method of Environmental Assessment Based on the Methods of Self-Organization of Mathematical Models.

A.N. Timoshevski, V.I. Yeremin, S.A.Kalkuta
Institute of Magnetism, 36-b Vernadsky St., 03680 Kiev, Ukraine, e-mail: tim@ukron.kiev.ua

Abstract: The method of generation of mathematical models of the observations defined on an irregular grid of observation data has been proposed. The method is based on self-organization of mathematical models and allows for the reconstruction and forecast (on a regular grid) of the observations fields of a different origin with controlling accuracy. This method could be used to any data field that is presented as a set of values in some region. A space distribution of geophysical (geochemical) parameters in a real geographic region could be considered as the data field of interest. In this study, the method has been used for the generation of the maps of the distribution of the ^{137}Cs and ^{90}Sr radioactive isotopes over the Zhytomir region of Ukraine (consequences of Chernobyl disaster). It is shown that this method allows to improve an accuracy when building the maps of the radioactive isotopes distribution for the cases of the fields defined on a small number of irregularly distributed points.

***Keywords**:* interpolation; approximation; self-organization of mathematical model; isotope distribution.

1. INTRODUCTION

Environmental science researches are growing in importance and scope with a major emphasis on translating the scientific basis into techniques and strategies that address both the needs of human societies and the requirement of natural systems. Emergency situation appearing due to a man-caused disaster requires a prompt appraisal of environmental issues and impact and forecast of its potential further evolution. For carrying on such appraisal and reaching full prognosis accuracy it is necessary to be based on unambiguous rigorous mathematical criteria of accuracy of the prognosis results. Decision making in these situations requires excluding or minimising personalistic decisions, which could be avoided only if we generate an optimal mathematical models that describe adequately, the observations defined on an irregular grid of observation data, as a rule. Using these models, we can reconstruct the data fields of potentially different origin and forecast their space distribution with a controlled accuracy. Presently, there exists a number of Geographic Information Systems (GIS) (Surfer, Er-Mapper, ArcView, ArcINFO, and MapInfo), which are highly flexible and devoted for the statistical spatial analysis of irregular cartographic databases. Their applicability for tackling a wide range of problems is proven, but a great number of modelling tasks are outside of their functionality. The GIS software is built using the following main methods. *Polynomial Regression Method* [Draper, N., and Smith, H., 1981] that is not a real interpolation method because local details of initial data have been lost at the map generation. *Triangular with Linear Interpolation* [Guibas, L., and J. Stolfi., 1985] and *Nearest Neighbor, Inverse Distance to a Power* [Davis, John, 1986] и *Natural Neighbor* [Watson, Dave, 1994] methods generate good surfaces if a tight grid of initial data is defined, but they are ineffective at the absence of initial data and therefore could not be used for extrapolation tasks. *Minimum Curvature* [Smith, W.H.F., and Wessel, P., 1990] and *Modified Shepard's Method* [Renka, R.J, 1988] methods are applicable for a majority of initial data sets, but could generate bad values at the regions where the initial data are absent. Surface smoothing is controlled by the parameter that is set manually. The method of weighted averaging (*Kriging)* [Cressie,N.A.C., 1990]) demonstrates many advantageous but it requires a great number of observations particularly for anisotropic fields. Besides, the most of algorithms use a regular grid of observation. *Radial Basis Functions* [Carlson, R.E., and Foley, T.A., 1991] are similar to the Kriging variogrammes, and the results of these two methods also are similar. The choice of the parameters is the same complicated as in *Kriging*. Thus, it is highly necessary to develop an approximation method that is "self-regulating» with respect to the initial data and real noise level. In this work, we propose such modelling method that is based on the methods of self-organization of mathematical models.

2. THREE- AND FOUR-DIMENSIONAL GLOBAL INTERPOLATION

The problem is to determine the values of the unknown function of many variables in any space point, basing on the known values of a function in a finite number of observation points. If the field to be recovered should coincide exactly with the initial function values, then we consider these tasks as interpolation; when recovering should be made with some deviations, such tasks are to be considered as approximation. Three stages for building algorithms of the field recovering are proposed. The first stage is the choice of the interpolation function form. At present, the function is being chosen depending on the information about the process under study. As a result, we have some function of many variables:

$y = F(\vec{x}, \vec{A})$, where $\vec{x} = \{x_1, x_2, ..., x_m\}$ – vector of current coordinates, $\vec{A} = \{a_1, a_2, ..., a_m\}$ – vector of unknown parameters. For global approximation, the values of the unknown parameters a_i $(i = 1, k)$ are determined, basing on the observation points and do not depend on the current coordinates, i.e. a_i =constant within the whole interpolation range. For local approximation, parameters a_i are determined at each interpolation site or in some local range, i.e. a_i depend on the current coordinates x_i. This interpolation should be used for the cluster type data. The method of the local interpolation is described in details elsewhere [A.N.Timoshevskii et al, 2000].

At the second stage, the metrics (p) for the interpolation functional are selected and minimized during the determination of the unknown $\overline{A}$ values. In the general case, metrics could be written as follows:

$$\left(\sum_{i=1}^{n} \left| y_i - F(\vec{x}_i, \vec{A}) \right|^p \right)^{1/p} \xrightarrow[A]{} \min$$

The choice of metrics is very important task only if the number of unknown parameters $A = \{a_1, a_2, ..., a_m\}$ is less than the number of initial parameters x_i *(i=1,…, n)*.

At the third stage, the functions of the functional weights are chosen. In such a case the weight c_i of the point i does not depend on the current coordinates of the global interpolation:

$$\left(\sum_{i=1}^{n} \left| y_i - F(\vec{x}_i, \vec{A}) \right|^p c_i \right)^{1/p} \xrightarrow[A]{} \min$$

And finally, the choice of a suitable interpolation scheme should be estimated using the specific numerical measurements that depend on the quality of interpolation in the prognostic points. The methods of self-organization of the models are used at all the stages of the process for making the best decision.

The essence of approach is modifying choice of the weights of initial points for the method of moving average:

$$F(x_t, y_t) = \frac{\sum_{i=1}^{n} c_i(x_t, y_t) f_i}{\sum_{i=1}^{n} c_i(x_t, y_t)}$$

Here and below for the purposes of writing simplicity, we have omitted dependence of the *F* function on the $\vec{A}$ vector of parameters. f_i (i=1,n) – is the value of parameter in initial point, x_t, y_t – the value of coordinate in the interpolation point, c_i (x_t, y_t) – weight of i-th point relative to the interpolation point (x_t, y_t).

We consider weighting function that has the following properties: firstly, at the sites, the function has the value equal to 1 while all other weights equal to zero, i.e.

$$c_i(x_j, y_j) = \delta_{ij} \text{ - Cronecer symbol}$$

Then, the weight of each interpolation point is considered to be linearly dependent on the whole set of initial points, i.e.

$$c_i(x_t, y_t) = \sum_{j=1}^{n} p_{ij} h_{jt}(\alpha)$$

here $h_{jt}(\alpha) = \sqrt{\Delta x_{jt}^2 + \Delta y_{jt}^2 + \alpha^2}$ - is the distance between the (x_t, y_t,) and (x_j, y_j,) points in three-dimensional space; α - is the constant parameter that controls the surface smoothing; p_{ij} – are the unknown values to be defined. In order to define the unknown p_{ij}, we consider the first condition for the weight functions, i.e.:

$$c_i(x_j, y_j) = \delta_{ij}$$

Then, after replacing c_i (x_j, y_j) with a second member of above equation, we could write:

$$\delta_{ij} = \sum_{k=1}^{n} p_{ik} h_{kj}(\alpha) \qquad \begin{cases} i = 1, n \\ j = 1, n \end{cases}$$

here δ_{ij} are the elements of matrix that is a product of matrices with p_{ik} and $h_{kj}(\alpha)$ elements. After doing some mathematical transformations, we will have two systems of linear algebraic equations, that are to be solved:

$$\sum_{i=1}^{n} h_{ij}\, a_i = f_j \qquad j = 1, n$$

$$\sum_{i=1}^{n} h_{ij} b_i = 1 \qquad j = 1, n$$

The solution of the system of equations relative to a_i and b_i , will result in writing a final formulae, as follows:

$$F(x_t, y_t) = \frac{\sum_{j=1}^{n} a_j h_{jt}}{\sum_{j=1}^{n} b_j h_{jt}}$$

or:

$$F(x_j, y_j, \alpha) = \frac{\sum_{i=1}^{n} a_i \sqrt{\Delta x_{ij}^2 + \Delta y_{ij}^2 + \alpha^2}}{\sum_{i=1}^{n} b_i \sqrt{\Delta x_{ij}^2 + \Delta y_{ij}^2 + \alpha^2}}$$

The equation for the four-dimensional model could be written in a similar form:

$$\mathrm{F}(\mathrm{x_j}, \mathrm{y_j}, \mathrm{z_j}, \gamma, \alpha) = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{a_i} \sqrt{\Delta \mathrm{x}_{\mathrm{ij}}^2 + \Delta \mathrm{y}_{\mathrm{ij}}^2 + \gamma^2 \Delta \mathrm{z}_{\mathrm{ij}}^2 + \alpha^2}}{\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{b_i} \sqrt{\Delta \mathrm{x}_{\mathrm{ij}}^2 + \Delta \mathrm{y}_{\mathrm{ij}}^2 + \gamma^2 \Delta \mathrm{z}_{\mathrm{ij}}^2 + \alpha^2}}$$

here f_i (i=1,n) – is the value of parameter in the initial point, γ-anisotropy factor, α-smoothing factor. In such a case, one more coordinate **z** appears, which could influence on the parameter to be mapped. Since this coordinate **z** can have another dimensions, a scaling coefficient γ is introduced. In order to find an optimal value of γ coefficient, it is necessary to carry on a regularization, and for every γ we will have ***a*** and ***b*** values, so behaviour of model will differ for different γ values.

3. REGULARISATION OF THE INTERPOLATION SCHEMES

For the three-dimensional model, the regularization is much more simpler due to absence of the □ dependence, and in such a case one can calculate the determination coefficient and then a smoothing factor α. Therefore we are presenting regularization for the four-dimensional model as being more general case. When varying the values of the γ and α parameters, we have an infinite number of mathematical models because an infinite set of surfaces can be drawn through the initial data with and without discrepancies. Among these models, we should find an optimal one. Therefore, it is necessary to introduce a quantitative estimation of the quality of any prognostic properties of the model. Let us consider a scheme for the regularisation (choice) of γ coefficient using the technique of "creation of informational deficit" in the initial sample. Let us fix values for a certain γ=0 и α=0. Then, let us remove k point from the initial sample and, basing on the above algorithm, build the forecast into this point at the fixed parameters. Thus, one can calculate a deviation between the real and forecasting values. The deviation measure is introduced as follows:

$$\Delta^k(\gamma) = [f_k - F^k(x_k, y_k, z_k, \gamma, \alpha)]^2$$

When repeating such procedure for all of the n points, we will have a forecast dispersion for all the samples, i.e.

$$R^2(\gamma) = \sum_{k=1}^{n} \Delta^k(\gamma) = \frac{1}{n} \sum_{k=1}^{n} [f_k - F^k(x_k, y_k, z_k, \gamma, \alpha)]^2$$

Let us introduce a dimensionless measure of the quality of the model forecasting properties, based on the multiple determination coefficient:

$$D(\gamma) = \left\{ 1 - \frac{\frac{1}{n} \sum_{k=1}^{n} [f_k - F^k(x_k, y_k, z_k, \gamma, \alpha)]^2}{\frac{1}{n} \sum_{k=1}^{n} [f_k - \bar{f}]^2} \right\} 100\%$$

The determination coefficient will show the percentage of determining constituent in this field at this method. Now, we have some value of the determination coefficient and should solve an inverse problem, i.e. it is necessary to define the surface smoothing extent based on the known determination coefficient. Because this interpolation scheme can ”explain” only $D(\gamma^{opt}) \leq 100\%$ percentage, let us select the surface with the determination coefficient equal to $D(\gamma^{opt})$ in all the points, by the mean of varying a smoothing coefficient α.

4. APPLICATION OF THE METHOD

Let us demonstrate the effectiveness of the method for the simple case of generation of two-dimensional model. As the function to be recovered we have chosen Y(x)=x+3sin(2x)+2 function. The control points have been selected randomly. In order to clarify whether the effectiveness of the method depends on the number of the control points, we have randomly chosen the number of the control points, as follows: 50, 100, 150, 200 and 250. When generating models at these sets of the points, one can calculate the determination coefficient D that defines "quality" of each model. The results of these calculations are shown in Fig. 1 (curve 1). One can conclude, that already for 100 points we succeed in reaching a good approximation results. The results of the generation of the model at the set of 200 points in a comparison with the test function Y(x) are shown in Fig. 2a. In order to examine the availability of the method for the noisy observation data, we have generated 10% of noise for the Y(x) test function. From the Fig.1 (curve 2) it is shown that 10% of noise decreases the determination coefficient but the dependence on the control points does not change practically. When using twice increase of noise (20%), the

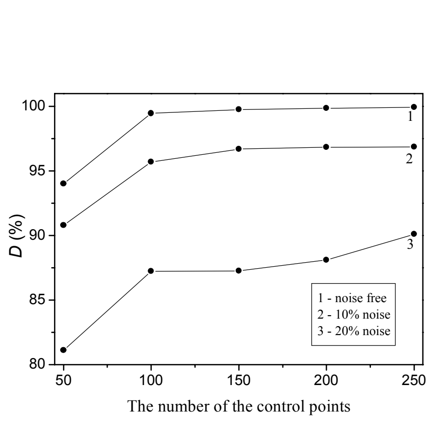

Fig.1. Dependence of the determination coefficient on the number of the control points

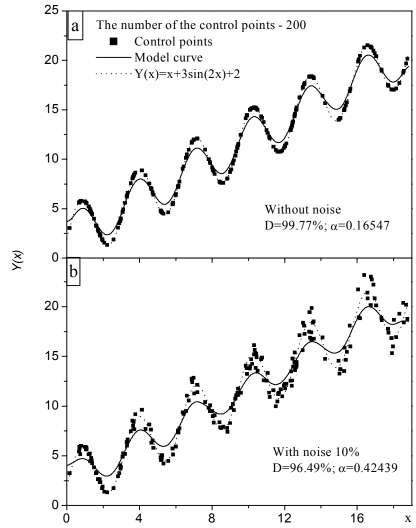

Fig.2 Reconstruction of the test dependence of Y(x) function using the 2D model

determination coefficient decreases hardly and, as a consequence, the prognostic properties of optimal model become worse. In Fig.2b, the results of the model generation for the data with 10% of noise are shown. When comparing with the results shown in Fig. 2a, we could see that the model accuracy decreases. Nevertheless, prognostic properties of our method remain stable. This allow us expecting successful resolution of the more complex tasks, for example, building the map of distribution of the radioactive isotope or man-caused contamination of the regions affected due to the ecocatastrophes.

We have used our method for building the maps of the ^{137}Cs and ^{90}Sr radioactive isotope distribution over the some regions of Ukraine, which have been contaminated due to the Chernobyl disaster The observation data on the ^{137}Cs isotope (940 control points) and ^{90}Sr isotope distribution differ significantly both for the number of the points and the noisy level. The experimental data set was on 25% smaller than for the ^{137}Cs dataset, and these data have a higher noise level. Thus, the quality of the optimal model for the ^{90}Sr isotope distribution becomes worse and prognostic accuracy of the map of the ^{90}Sr isotope distribution also decreases.

Indeed, at the three-dimensional modelling the determination coefficients for the ^{137}Cs isotope distribution is 89.54% (α=0.78), and for the ^{90}Sr isotope distribution it is 53.51% (α=2.11). So small determination coefficient for the ^{137}Cs isotope distribution could be generated if we use

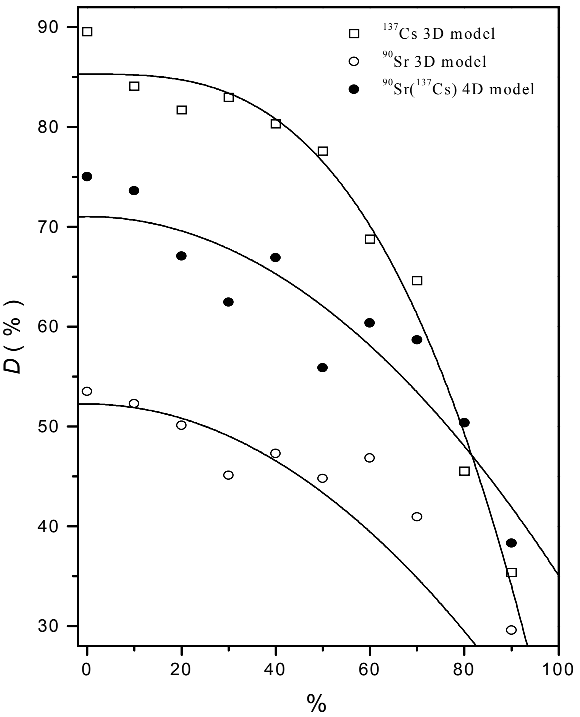

Fig.3. The dependence of the determination coefficient of the number of the points discarded

only 20% of the control points. This results from the calculations that we have carried out in order to find the determination coefficients for generation of three-dimensional models of the ^{137}Cs and ^{90}Sr isotope distribution for a different number of random control points. In Figure 3, the results of these calculations are shown. We can see a high stability of our calculation method when decreasing the number of the control points. From Figure 3 it is shown that even when building model on 50% of the control points, for the ^{137}Cs isotope the determination coefficient decreases from 89,5% to 80% and for ^{90}Sr isotope it decreases from 53.51% to 45%. Our investigations show that the method that we propose is a highly effective for the cases when it needed to recover the distribution of the fields

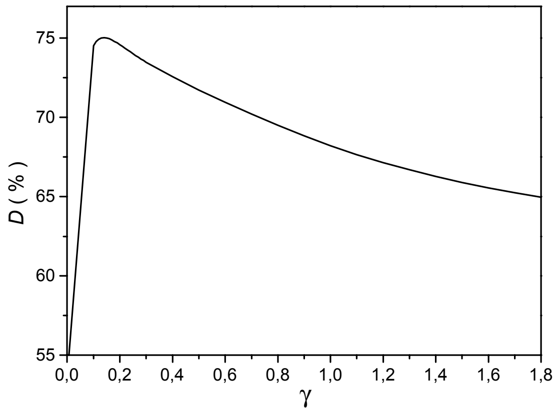

Fig.4. Regularization plot over the γ for the four-dimensional model of the ^{90}Sr(^{137}Cs).

defined on a small number of irregularly distributed points. So, we have shown that the map of the ^{90}Sr isotope distribution has been built with a less accuracy. The corresponding map of the prognostic deviations verifies this also. So, the question arises: whether it is possible to improve an accuracy of the map of the ^{90}Sr isotope distribution without carrying out additional measurements? The method, that we propose, allows doing that. Taking the information of the ^{137}Cs isotope distribution into account, a more reliable map of the ^{90}Sr isotope distribution can be calculated. This is because the experimental data of the ^{137}Cs distribution are more accurate and, for another thing, because the ^{137}Cs and ^{90}Sr isotopes emission has occurred simultaneously and over the same territory. In our method of multi-dimensional mathematical modelling, it is enough to come from building the three-dimensional model to the four-dimensional one. For this purpose, we could write the ^{90}Sr isotope distribution as the function with three coordinates ^{90}Sr=F(x, y, z (^{137}Cs)). In such a case we will have two parameters, as follows: anisotropy coefficient γ and smoothing coefficient α. To define an optimal four-dimensional mathematical model it is necessary to carry on the regularization over

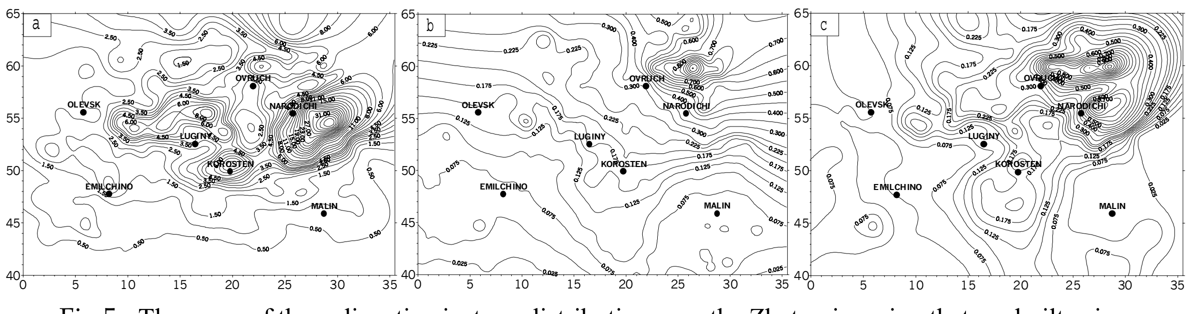

Fig.5. The maps of the radioactive isotope distribution over the Zhytomir region that are built using three-dimensional (*a* и *b*) and four-dimensional (c) modelling: *a* - ^{137}Cs; *b* - ^{90}Sr; *c* - ^{90}Sr(^{137}Cs)

Table 1: The results of the modelling of the space distribution of radioactive isotopes

Model dimensions	Isotope	α	D(%)
3D	^{137}Cs	0.78	89.54
	^{90}Sr	2.11	53.51
4D	$^{90}Sr(^{137}Cs)$	1.97 γ=0.14	75.03

the γ parameter. This regularization is shown in Fig. 4. For these values of α and γ parameters, we will have an optimal model that allows building the map of the ^{90}Sr isotope distribution. The determination coefficient in such a case is higher (D=75%) then for the three-dimensional modelling of the ^{90}Sr isotope distribution (D=53.51%). The calculation results are shown in Table 1. Figure 5 shows the maps of the ^{137}Cs and ^{90}Sr isotope distribution that has been obtained by the three-dimensional modelling and the map of the ^{90}Sr isotope distribution that has been obtained by the four-dimensional modelling taking the information of the ^{137}Cs isotope distribution into account. All calculated maps are shown in the relative coordinates and ci/km^2 units. The maps of the ^{90}Sr isotope distribution that has been obtained by the four-dimensional modelling differs significantly from that obtained using the three-dimensional modelling. Thus, the method of multi-dimensional modelling that is based on the principles of self-organization allowed us to build a more reliable map of the ^{90}Sr isotope distribution.

5. CONCLUSIONS

It is proposed a new effective method of multi-dimensional approximation, which is based on the idea of self-organization of mathematical models. Application of our method for the real experimental data on the contamination of the territory of Ukraine with radioactive ^{137}Cs and ^{90}Sr isotopes allowed us to build the more precise maps of the ^{90}Sr isotope distribution. Our investigations show that the method that we propose is a highly effective for the cases when it needed to recover the distribution of the fields defined on a small number of irregularly distributed points, which allow us expecting a successful resolution of the more complex tasks

6. ACKNOWLEDGEMENTS

The work was made under the financial support of the Science and Technology Center of Ukraine.

7. REFERENCES

Davis, John, *Statistics and data analysis in geology*, John Wiley and Sons, New York, 1986.

Cressie, N.A.C., The origin of kriging. *Mathenatical Geology*, *22*, 239-252, 1990.

Smith, W.H.F., and P. Wessel, Gridding with continuous curvature splines in tension, *Geophysics*, *55* (3), 293-305, 1990.

Renka, R.J., Multivariate interpolation of large sets of scattered data, *ACM Transaction on Mathematical Software*, *14*(2), 139-148, 1988.

Watson, Dave. *An implementation of natural neighbour interpolation*, David Watson, P.O.Box 734, Clarement, WA 6010, Australia, 1994.

Draper, N., and H. Smith, *Applied Regression Analysis*, second edition, Wiley-Interscience, 709 pp., 1981.

Carlson, R.E., and T.A. Foley, *Radial Basis Interpolation Methods on Track Data*, Lawrence Livermore National Laboratory, UCRL-JC-1074238, 1991.

Guibas, L., and J. Stolfi, Primitives for Manipulation of General Subdivisions and the Computation of Voronoi Diagrams, *ACM Transactions on Graphics*, *4*(2), 74-123, 1985.

Timoshevskii, A.N., V.I. Yeremin, and S.A. Kalkuta, New method for ecological monitoring based on the method of self-organising mathematical models, The proceedings Fifth International Symposium on Environmental Contamination in Central and Eastern Europe, Prague, Czech Republic, 2000.