THE METHOD OF SELF-ORGANISATION OF MATHEMATICAL MODELS AND ITS APPLICATION TO THE POPULATION ANALYSIS AND ENVIRONMENTAL MONITORING

A.Timoshevskii, V.Yeremin, S.Kalkuta
Institute of Magnetism of National Academy of Sciences, 36-b Vernadsky St., 03142 Kiev, Ukraine,
e-mail: tim@ukron.kiev.ua

The method of self-organisation of mathematical models that the “Spacer” software suit is based upon has been applied to the treatment of measurements data on contamination of some regions of Ukraine with radioactive isotopes. We have built the maps of isotopes distribution, and determined accuracy of these maps generation. It was processed the data of the people death-rate on the territory of Ukraine in different years. It is determined that these data are not homogeneous. The maps of the people death-rate over the territory of Ukraine in different years are built. An accuracy of these maps generation was estimated.

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. Thus, it is highly necessary to develop an approximation method that is "self-regulating" with respect to the initial data and real noise level. We propose such modelling method that is based on the methods of self-organization of mathematical models. The authors of this research work propose required methodology that is based on the principles of self-organization. We have developed the algorithms and generated the software suit (***“SPACER”***) that is based on the methods of self-organization of multidimensional mathematical models and enables definition of deterministic and noise components of data defined on a strongly irregular grid of noised observation data. In contrast to all presently existing methods, we can control accuracy of optimal model being able to calculate prognostic errors over a total region of existence of measured data, and also determination coefficient that characterizes quality of the model chosen.

Method

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.

Four stages for building algorithms of the field recovering are proposed.

The **first stage** is the choice of the interpolation function form. At present, this procedure has not been formalized and therefore 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, $\vec{A} = \{a_1, a_2, ..., a_k\}$ - vector of unknown parameters.

At the **second stage**, the function definition boundary is determined. There are two possibilities, as follows: 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. 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. These cases correspond to the global approximation (1) and local approximation. In this work we will use local approximation. The essence of this method is stated below.

At the **third stage**, the functions of the functional weights are chosen. In such a case the weight S_i of the point *i* does not depend on the current coordinates of the global interpolation. When carrying out a local interpolation, some negative and non-increasing function $u(\vec{x}_t, \vec{x}_i, \vec{B})$ that depends on the initial points and interpolation cites is introduced. In general case, the weight function can include the constants. The values of these constants are to be determined during the minimization of the function, or are to be taken basing on the physical considerations.

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

where: $\vec{x}_t = \{x_{t_1}, x_{t_2}, ..., x_{t_m}\}$ is a current interpolation point; $\vec{x}_i = \{x_{i_1}, x_{i_2}, ..., x_{i_m}\}$ is the i-th point of initial data; $\vec{B} = \{b_1, b_2, ..., b_l\}$ are the parameters that determine the properties of the weight function.

And finally, the choice of suitable interpolation scheme should be estimated using specific numerical measurements that depend on the quality on interpolation in prognostic points. Methods of the model self-organizations are used at the all of stages for making the best decision.

The scheme of the four-dimensional polynomial local interpolation is shown below. Some functional for an arbitrary point $\{x_t, y_t, z_t\}$ could be written as follows:

$$j\left(x_t, y_t, z_t, \vec{A}\right) = \sum_{i=1}^{n} \left(f_i - p^k(x_i, y_i, z_t, \vec{A}) \right)^2 w\left(x_t, y_t, z_t, x_i, y_i, z_i, a, b\right)$$

where: $\{x_i, y_i, z_i, f_i\}^n$ - initial points (data); $\{x_t, y_t, z_t\}$ - current interpolation point;

$p^k\left(x_{i,} y_i, z_i, \vec{A}\right)$ - k-power polynomial of three-variables.

Without loss of generality, we consider a local polynomial of the first order, i.e.:

$p^1\left(x_i, y_i, z_i, \vec{A}\right) = a_1 + a_2 x_i + a_3 y_i + a_4 z_i$

$w_{ti}(a, b) \equiv w(x_t, y_t, z_t, x_i, y_i, z_i, a, b)$ is the weighting function, i.e. it is the weight of i-th initial point relative to the interpolation site $\{x_t, y_t, z_t\}$. In our case, the weighting function reads as follows:

$$w_{ti}(a, b) = \frac{[(x_t - x_i)^2 + (y_t - y_i)^2 + (z_t - z_i)^2 + b^2]^{-a}}{\sum_{j=1}^{n} [(x_t - x_j)^2 + (y_t - y_j)^2 (z_t - z_j)^2 + b^2]^{-a}}$$

Thereby we define a weighting function as monotone decreasing function depending on the distance between the t points and ranging between 0 and 1. Function has value 1, when $b = 0$, when the points coincide, and 0 - when they are infinitely far from one another. $a \geq 0$ - parameter that controls the rate of the weighting function decreasing; b - parameter that controls surface smoothing .

At the $b = 0$, surface comes through the initial points, and at the $b = \infty$, the surface degenerate into the k-th order polynomial. Thus, at the fixed in value a and b, in arbitrary point $\{x_t y_t\}$ we can find unknown coefficients $\vec{A}$ using the mean square root method. For this aim, we take partial derivatives $\{a_i\}$ $(i = 1, k)$ and equate them to zero:

$$\frac{\partial j(x_t, y_t, z_t, \vec{A})}{\partial a_i} = 0 \qquad (i = 1, k)$$

Solution of the system of normal linear (relative to a_i) algebraic equations will result in getting unknown polynomial coefficients in $\{x_t y_t\}$ point. Then after the calculation of polynomial in this point we will have interpolating value of function. The details of the method are given elsewhere (2).

Mixture separation

For analysis of measured data and search of implicit rules we have generate algorithms and programs that realize the method of mixture separation (the second part of the ***"SPACER"*** software suit). The procedure of a mixture separation belongs to parametric methods of the estimation of probability density function. It should be used in those cases when it is necessary to extract homogeneous groups of data and classify the results of observations. The premise for the method of a mixture separation is the fact that every homogeneous group could be presented by its probability density function *f(X,A)*, where *A* parameter is some vector of values, which defines the form of distribution. The problem could be stated as follows: Let we have observation data $?_1,\ldots, ?_N$, which should be classified. For that, the equation of a finite mixture of distribution density functions could be written as:

$$h(x) = \sum_{i=1}^{M} P_i f(X, A)$$

where: M – is a number of homogeneous groups in summary sampling; A_i –parameters of i-th density function; P_i – portion of i-th group (group probability).
Thus, the problem of a mixture separation and classification of the results by groups comes to the task of evaluation of unknown parameters $M, \{A_i, P_i\}$ (i=1, M).

There exist several methods of these parameters evaluation. The most prevailing is the method of maximum of plausibility.

Application of the software suit for analysing fields of radioactive contamination resulted from the Chernobyl disaster

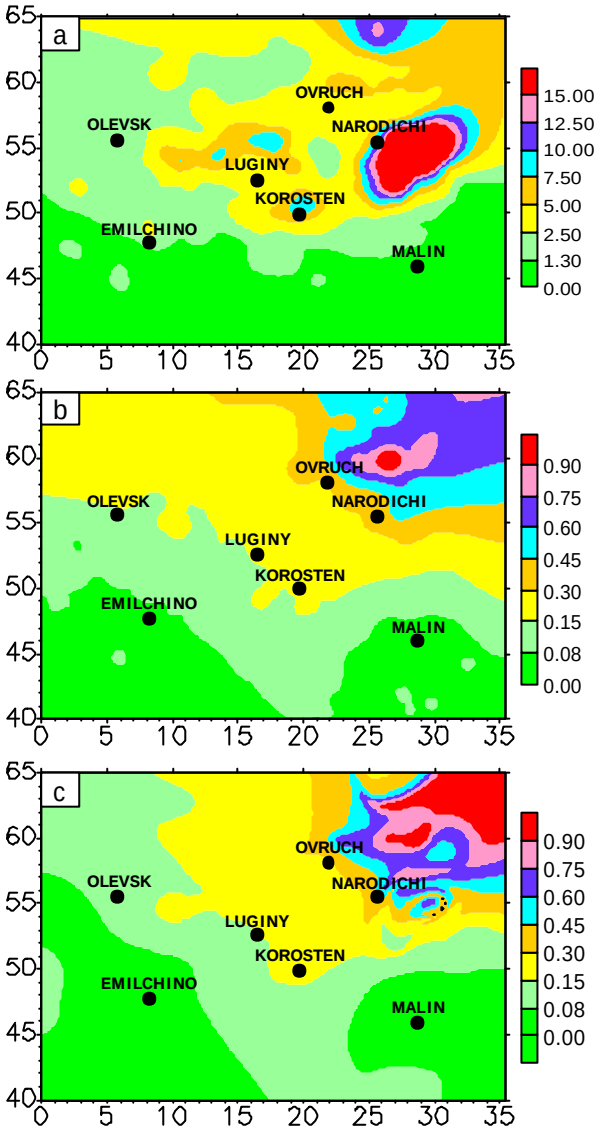
We have used software suit based on the described algorithms for building the maps of the ^{137}Cs and ^{90}Sr radioactive isotope distribution over the some regions of Ukraine, which have been contaminated due to the Chernobyl disaster. The observation data on the ^{137}Cs isotope (940 control points) and ^{90}Sr isotope (700) distribution differ significantly both for the number of the points and the noisy level. The calculation results are shown in Table 1. We have shown that the map of the ^{90}Sr isotope distribution has been built with a less accuracy. 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 because the ^{137}Cs and ^{90}Sr isotopes emission has occurred simultaneously. 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)). We have two parameters, as follows: non-linear coefficient α and smoothing coefficient β. To define an optimal four-dimensional mathematical model it is necessary to carry on the regularization over the α and β parameters. The determination coefficient in such a case is higher then for the three-dimensional modelling of the ^{90}Sr isotope distribution (Table 1).

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

Model dimension	**Isotope**	a	b	**D(%)**
3D	^{137}Cs	2.6	1.23	89.6
	^{90}Sr	0.9	0.64	59.9
4D	^{90}Sr(^{137}Cs)	2.6	3.8	71.7

Figure 1 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.

Figure 1: The maps of the radioactive isotope distribution over the Zhytomir region that are built using three-dimensional (*a* ? *b*) and four-dimensional (c) modelling: *a* - ^{137}Cs; *b* - ^{90}Sr; *c* - $^{90}Sr(^{137}Cs)$



Application of the software suit for analysing fields of death-rate of population of Ukraine.

We have used trhee-dimensional local polynomial model for building the map of the population death-rate in the Ukraine in 1995 (Fig. 2). The map is shown in relative units on a scale 1:1000000, the units are the number of deaths per 1000 persons. The number of control points equals to the number of regions of Ukraine (486). The death-rate is averaged over the region, and the control point coordinates are specified in the center of mass over the region population. The calculation of the model regularization over nonlinearity factor α enabled finding an optimal value α=1.3, which corresponds to D=61.7%. Due to a huge territory described by the map, this map demonstrates only the main tendency of the growing death-rate in the Ukraine with accuracy of 65.3%. Based on the value of the death-rate in 1990, we have built the map of the rate of the death-rate increase in the Ukraine during five years, namely, from 1990 till 1995 (Fig.3).

Figure 2: The maps of the death-rate population of Ukraine that are built using three-dimensional modelling: *a* – 1979, D=52.8%; *b* – 1985, D=57.5%; *c* – 1990, D=59.1%; *d*– 1995, D=61.7%.

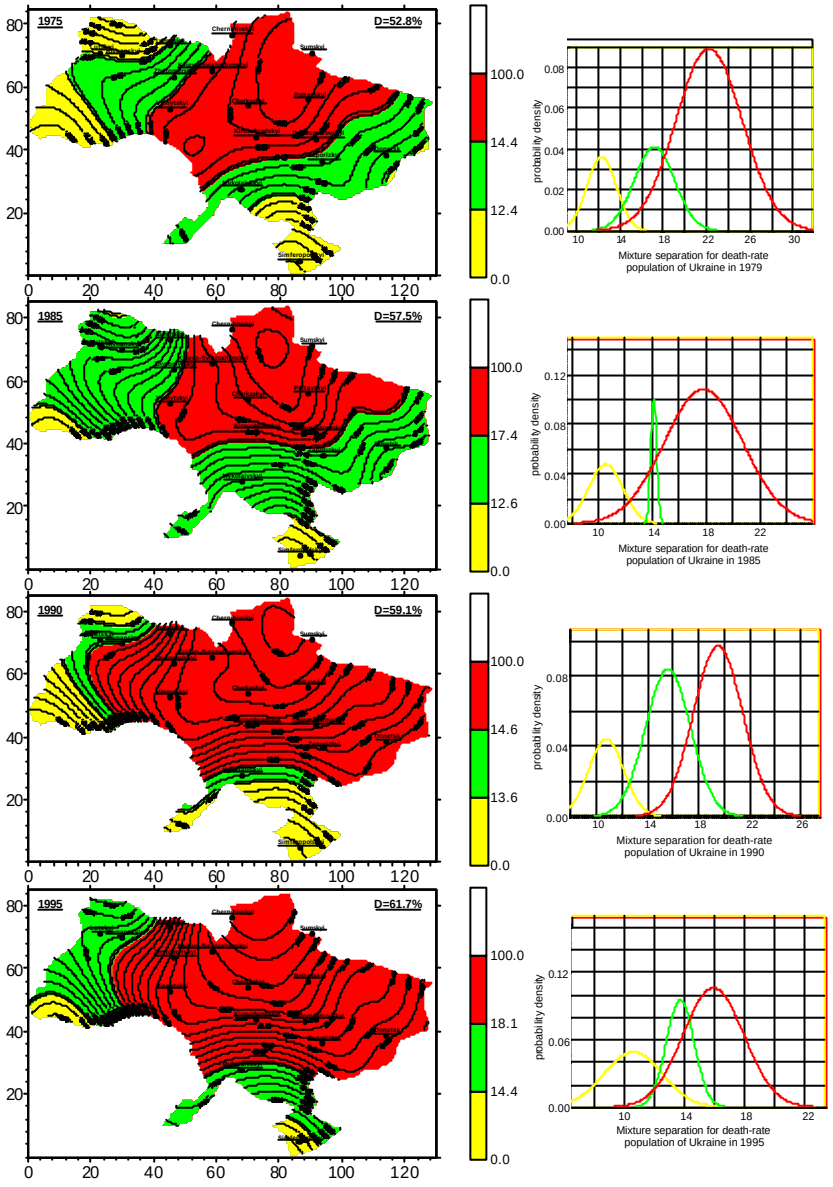
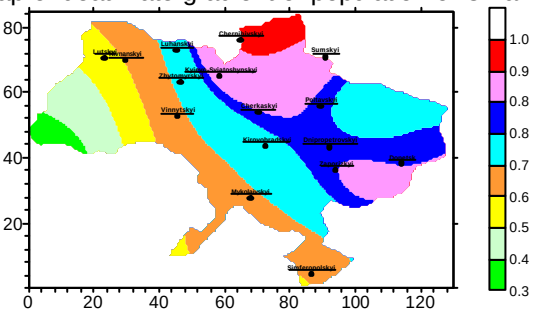


Figure 3: The map of death-rate gradient of population of Ukraine for 1990-1995.



References

(1) A. Timoshevski, V.I. Yeremin, S.A.Kalkuta "New Method of Environmental Assessment Based on the Methods of Self-Organization of Mathematical Models" The proceedings International Environmental Modeling and Software Society, Lugano, Switzerland, 24-27 June 2002, v.3, p.542-547.

(2) A. Timoshevskii, V. Yeremin, S. Kalkuta "New method for ecological monitoring based on the method of self-organising mathematical models" Ecological Modelling, 162, ? 1-2, 2003, p. 1-13