



Jul 1st, 12:00 AM

The Method of Self-Organisation of Mathematical Models and its Application to the Population Analysis and Environmental Monitoring

L. Tymoshevska

A. N. Timoshevski

V. Yeremin

S. A. Kalkuta

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Tymoshevska, L.; Timoshevski, A. N.; Yeremin, V.; and Kalkuta, S. A., "The Method of Self-Organisation of Mathematical Models and its Application to the Population Analysis and Environmental Monitoring" (2004). *International Congress on Environmental Modelling and Software*. 38.

<https://scholarsarchive.byu.edu/iemssconference/2004/all/38>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

The Method of Self-Organisation of Mathematical Models and its Application to the Population Analysis and Environmental Monitoring

L.Tymoshevska, A.Timoshevskii, V.Yeremin, S.Kalkuta

Institute of Magnetism of National Academy of Sciences, Kiev, Ukraine, (liliya@ukron.kiev.ua)

Abstract: We propose new methodology for the treatment of highly noised statistical data that are measured and defined on a strongly irregular grid of observations. Using this methodology, we have processed and analyzed statistical data on congenital heart diseases and oncological diseases over three Ukrainian regions which were influenced by the Chernobyl disaster. The data appeared to be strongly inhomogeneous. We have built the maps of the people morbidity spatial distribution and the maps of the man caused contaminations in these regions. We have used Cs-137 and Sr-90 radioactive isotopes' and pesticide pollutions as the factors that characterize the rate of environmental contamination over these regions of Ukraine. 81 points have presented the information about each kind of disease and pollution. Geometrically, these points correspond to the location of regional capital cities. On the base of these results, we have built the maps of correlation coefficients of morbidity versus pollution spatial distribution.

Keywords: approximation; self-organization; environmental contamination

1. INTRODUCTION

In ecology and medicine, geology and seismology and many other different fields of science and technology there exists the problem that concerns processing of large statistical data sets defined on a strongly irregular grid of observation data. Very often these data are highly noised, and the number of measurements is limited. Processing and analysis of these data imply extraction of maximum information with a controlled accuracy and require excluding or minimising human decisions. Thus we concluded that we need to have methodology that would be the base for the generation of self-adjusted methods for data processing, which could be applied to any measured data with any noise content. Self-adjustment implies generation of optimal mathematical models that describe adequately the observations defined on an irregular grid of observation data and exclude an influence of researchers' human decisions. Using these models, we can reconstruct the data fields of potentially different origin and forecast their space distribution with a controlled accuracy. Another constituent of the above mentioned methodology is the method for data analysis which should enable searching implicit rules that might characterize measured data. Important information can be

hidden due to a high noise content, strongly irregular grid of observation and limited number of data themselves.

2. METHOD

The authors of this research work propose 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 the data that are defined on a strongly irregular grid of noised observation data. In contrast to previous methods, with our method we can control accuracy of optimal model and calculate prognostic errors over a total region of existence of measured data; we are also able to calculate determination coefficient that characterizes quality of the model chosen.

The problem is to determine the values of unknown function of many variables in any space point being based on known function values 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, these tasks are to be considered as approximation.

Four stages for building algorithms of the field recovering have been proposed.

The **first stage** is the choice of the form of interpolation function. 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 $\vec{x} = \{x_1, x_2, \dots, x_m\}$ - vector of current coordinates, $\vec{A} = \{a_1, a_2, \dots, a_k\}$ - 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 from the observation points and do not depend on the current coordinates, i.e. $a_i = \text{constant}$ within the

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 (2).

At the **third stage**, metrics (p) for interpolation functional are chosen, which is minimized at the determination of unknown \vec{A} values.

At the **fourth stage**, the functions of the functional weights are chosen. In such a case the weight S_i of 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 $v(\vec{x}_t, \vec{x}_i, \vec{B})$ is introduced that depends on the initial points and interpolation sites. 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 from the physical considerations.

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

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

And finally, the choice of suitable interpolation scheme should be estimated using specific numerical measurements that depend on quality of interpolation in prognostic points. The methods of

the model self-organisations are used at all of the stages for making the best decision.

For analysis of measured data and search of implicit rules we have generated algorithms and programs that realise 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 a some vector of values, which defines the form of distribution. The problem could be stated as follows: we have observation data X_1, \dots, X_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 density function; P_i – portion of i group (group probability).

Thus, the problem of a mixture separation and classification of the results by groups comes to the task of unknown parameters $M, \{A_i, P_i\}$ ($i=1, M$) evaluation.

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

3. RESULTS

In this work, we have demonstrated an application of “*SPACER*” suit for processing and analysis of statistical data for two kinds of congenital diseases, as follows: cancers and heart diseases over three regions of Ukraine (Zhytomyr, Kiev and Tchernigov regions). Using the methods that we have developed, we can determine an influence of different environmental contaminations on these diseases. As the factors that characterize the rate of environmental contamination over three regions of Ukraine we have used their pollution by radioactive Cs-137 and Sr-90 isotopes and pesticide pollution. The information about each kind of disease and pollution is shown by 81 points. Geometrically, these points correspond to the location of regional capital cities. Thus, the information in each point is integral and characterises morbidity and ecological situation of a whole region.

On the first stage of data processing, we have synthesised optimal three-dimensional models and

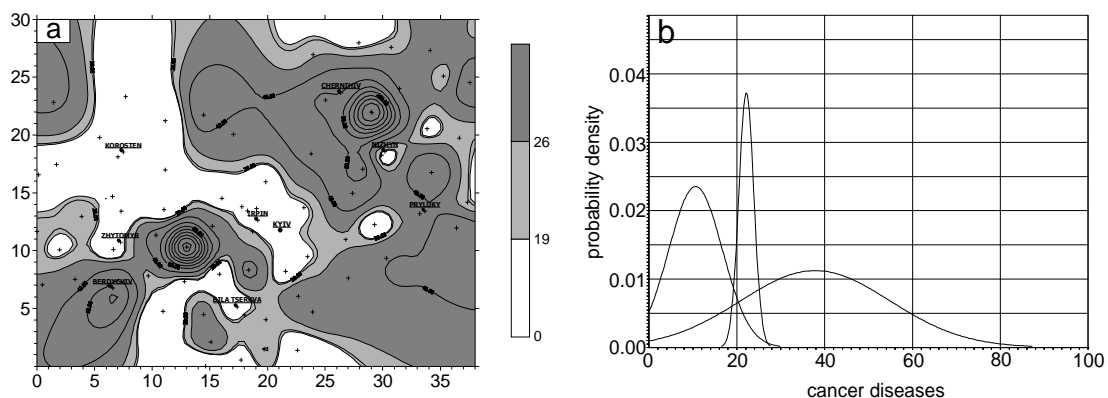


Figure 1: (a) - The map of cancer diseases; (b) - Mixture separation for cancer diseases.

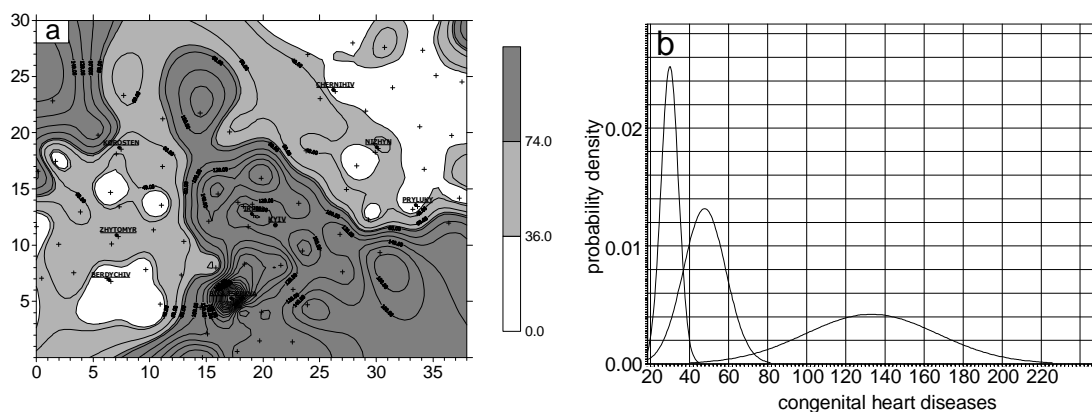


Figure 2: (a) - The map of congenital heart diseases; (b) - Mixture separation for congenital heart diseases.

built the maps of distribution of 2 kinds of morbidity (Fig.1-2) and the maps of Cs and Sr isotopes space distribution, total radiation doze (Fig.3a,4a,5a) and pesticide distribution (Fig.6a). Accuracy of the maps is within 70-80%, which shows high reliability of the results obtained.

On the second stage we have used the method of the mixture separation (second part of “SPACER”) for analysis of measured data. Our calculations have shown that the data for two kinds of diseases are not homogeneous and could be presented by 3 classes (Fig.1b,2b). The same calculations were carried out for classifying initial data on the heart diseases.

The results of classification and class spatial distribution are shown in Fig.2a and Fig.2b. From Fig.1a and Fig.2a it follows that the spatial distribution of oncological and heart diseases differs significantly. Application of the method of mixture separation to the investigation of initial data on the space distribution of isotopes and pesticides has shown that Sr-90 distribution contains 2 classes of data (Fig.3a); Cs-137 distribution, total radioactive doze and pesticide distributions contain 3 classes of data (Fig.4a,5a,6a).

Generation of the maps of space distribution of morbidity and different territory contaminations enables calculation of the maps of space distribution of coefficients of correlation between the morbidity and kind of pollution. The results of these calculations for oncological diseases are shown in Fig.3b, 4b, 5b, 6b. The same results for the heart diseases distribution are shown in Fig. 3c, 4c, 5c, 6c. From the results of our investigations we can conclude that environmental pollution has a stronger influence on congenital heart diseases than on oncological diseases. In order to clarify the results in details it is necessary to enlist the service of medical experts and to continue further joint studies using additional medical data.

4. REFERENCES

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.

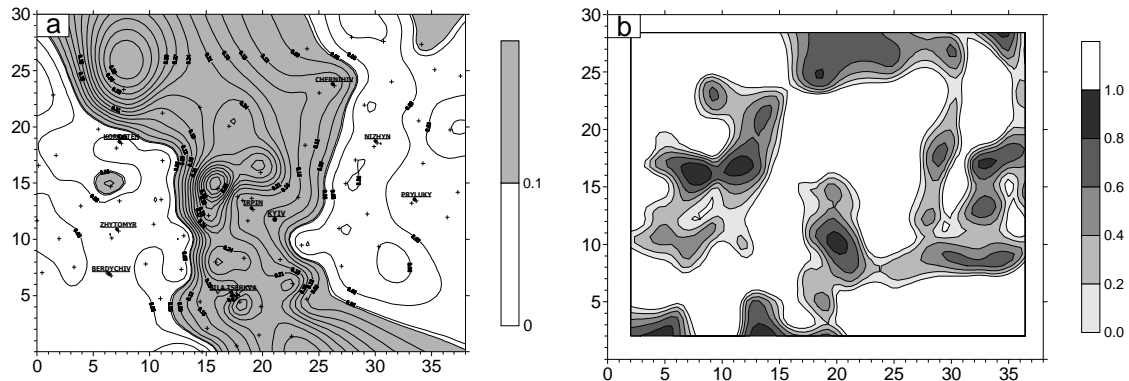


Figure 3:

- (a) The map of approx. average of soil contamination of populated lands with Sr-90 (Ci/km²), 1992.
- (b) The map of coefficient of correlation between cancer diseases and Sr-90.
- (c) The map of coefficient of correlation between heart diseases and Sr-90.

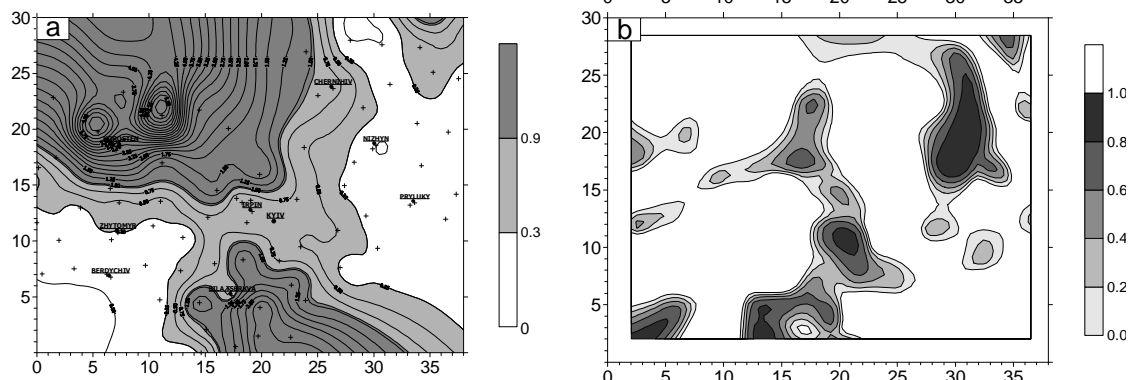


Figure 4:

- (a) The map of approx. average soil contamination of populated lands with Cs-137 (Ci/km²), 1992.
- (b) The map of coefficient of correlation between cancer diseases and Cs-137.
- (c) The map of coefficient of correlation between heart diseases and Cs-137.

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.

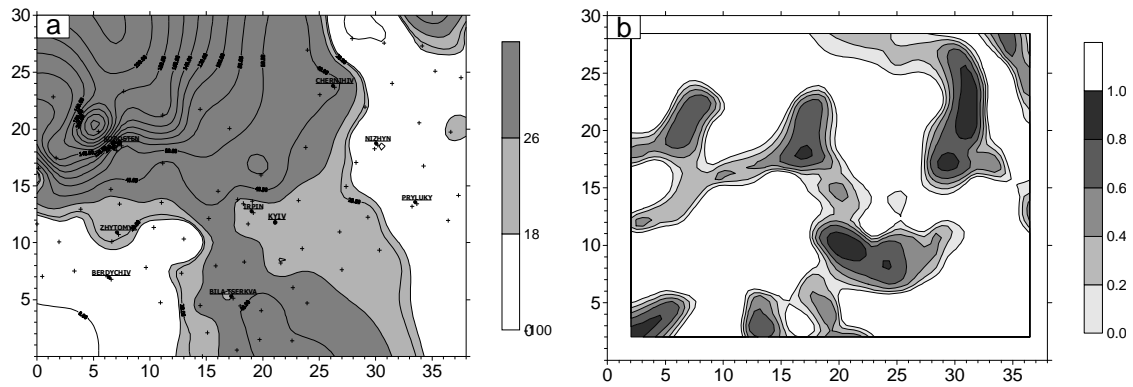


Figure 5:

- (a) The map of the average designed all year total dose of Chernobyl irradiation (mBer), 1992.
- (b) The map of coefficient of correlation between cancer diseases and total dose of Chernobyl irradiation.
- (c) The map of coefficient of correlation between heart diseases and total dose of Chernobyl irradiation.

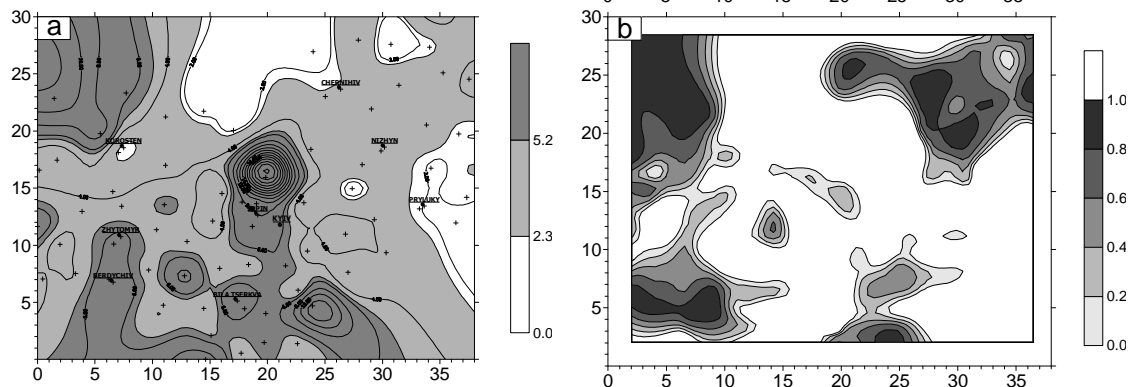


Figure 6:

- (a) The map of the all year loading of all type pesticides on arable lands (Kg/Hektar), 1988.
- (b) The map of coefficient of correlation between cancer diseases and all type pesticides on arable lands.
- (c) The map of coefficient of correlation between heart diseases and all type pesticides on arable lands.