POLISH ACADEMY OF SCIENCES
Systems Research Institute

MODELLING CONCEPTS AND DECISION SUPPORT IN ENVIRONMENTTAL SYSTEMS

Editors:
Jan Studzinski
Olgierd Hryniewicz

Polish Academy of Sciences • Systems Research Institute

Series: SYSTEMS RESEARCH
Vol. 45

Series Editor:

Prof. Jakub Gutenbaum

Warsaw 2006

MODELLING CONCEPTS
AND DECISION SUPPORT
IN ENVIRONMENTAL SYSTEMS

MODELLING CONCEPTS AND DECISION SUPPORT IN ENVIRONMENTAL SYSTEMS

Editors:

Jan Studzinski

Olgierd Hryniewicz

The purpose of the present publication is to popularize information tools and applications of informatics in environmental engineering and environment protection that have been investigated and developed in Poland and Germany for the last few years. The papers published in this book were presented during the workshop organized by the Leibniz-Institute of Freshwater Ecology and Inland Fisheries in Berlin in February 2006. The problems described in the papers concern the mathematical modeling, development and application of computer aided decision making systems in such environmental areas as groundwater and soils, rivers and lakes, water management and regional pollution. The editors of the book hope that it will support the closer research cooperation between Poland and Germany and when this intend succeeds then also next publications of the similar kind will be published.

Papers Reviewers:
Prof. Olgierd Hryniewicz
Prof. Andrzej Straszak

Text Editor: Anna Gostynska

Systems Research Institute of Polish Academy of Science
Newelska 6, PL 01-447 Warsaw

Section of Scientific Information and Publications
e-mail: biblioteka@ibspan.waw.pl

ISBN-10: 83-894-7505-7
ISBN-13: 978-83-894750-5-3
ISSN 0208-8029

CHAPTER 4

Water management and Decision support

DEVELOPMENT OF KRIGING ALGORITHMS FOR APPROXIMATING ENVIRONMENTAL MEASUREMENTS DATA*

Lucyna BOGDAN
Systems Research Institute, Polish Academy of Sciences, Warsaw
<*bogdan@ibspan.waw.pl*>

Abstract: *In the System Research institute in the frame of a research project supported by the Polish Ministry of science the software for such environmental data as yearly rain and snow falls, the soil composition and the soil erosion components is developed. The end result of this investigation shall be the development of geostatic maps for the whole Poland or some polish regions. The problems which are to be solved by developing the kriging algorithms concern the choice of two- or three-dimensional approximation, the choice of directional or directionless approach, the formulation of analytical functions for mathematical modeling of the data. There are some computer programs for kriging approximation available on the market but our decision was to work out an own software because of the possibility of using the own algorithms developed in the institute. In the paper the theory of kriging algorithm with detailed proof explanations that has been developed into computer program to solve the above tasks of geostatistic approximation is presented and some calculation results are enclosed.*

Keywords: Space approximation, kriging algorithms, mathematical modeling, environmental measurements.

1. Introduction

In this paper we will look at ordinary kriging, a method that is often associated with the acronym B.L.U.E. for "best linear unbiased estimator". Ordinary kriging is linear because its estimates are weighted linear combinations of the available data; it is unbiased since it tries to have the mean error m_R equal to 0; it is best because it aims at minimizing σ_R^2, the variance of the errors. In ordinary kriging we do not know m_R and therefore cannot guarantee that it is exactly 0. Nor do we know σ_R^2; therefore we cannot minimize it.

* The paper is a result of the research project of the Polish Ministry of Education and Science No. 429/P04/2003/25

The best we can do is to build a model of the data we are studying and work with the average error and the error variance for the model. We use a model in which the bias and error variance are calculated and then chose weights for nearby samples that ensure that the average error for our model $\tilde{m}_R$ is exactly 0 and that our modeled error variance $\tilde{\sigma}_R^2$ is minimized.

2. The Random Function Model and Error Variance

For any point at which we attempt to estimate the unknown value our model is a stationary random function that consists of n sample locations $V(x_1),\ldots,V(x_n)$ and one for the unknown value at the point we are trying to estimate, $V(x_0)$. Each of this random variables has the same probability law; at all locations the expected value of the random variable is $E\{V\}$.

Every value in this model is seen as the outcome of a random variable; the samples are outcomes of random variables as well as the unknown true value. Our estimate is a weighted linear combination on the random variables at the available sample points:

$$\tilde{V}(x_0) = \sum_{i=1}^{n} w_i \cdot V(x_i)$$

Similarly the estimation error

$$R(x_0) = \tilde{V}(x_0) - V(x_0)$$

is also a random value as the difference between the estimate and the random value modeling the true value. Substituting the previous equation we obtain:

$$R(x_0) = \sum_{i=1}^{n} w_i \cdot V(x_i) - V(x_0)$$

We can ensure that the error at any particular location has an expected value 0 by applying the formula for the expected value to the above equation:

$$E\{R(x_0)\} = E\left\{\sum_{i=1}^{n} w_i \cdot V(x_i) - V(x_0)\right\} =$$

$$= \sum_{i=1}^{n} w_i E\{V(x_i)\} - E\{V(x_0)\}$$

We have already assumed that the random function is stationary, which allows us to express all the expected values on the right-hand side as $E\{V\}$:

$$E\{R(x_0)\} = \sum_{i=1}^{n} w_i \cdot E\{V\} - E\{V\}$$

Settling the expected value to 0 to ensure unbiased ness results we obtain:

$$E\{R(x_0)\} = 0 = E\{V\}\sum_{i-1}^{n} w_i - E\{V\}$$

$$E\{V\}\sum_{i=1}^{n} w_i = E\{V\}$$

$$\sum_{i-1}^{n} w_i = 1$$

So it is a condition of unbiased ness.

Now we will try to present a set of estimates for which the variance of the errors is minimum.

The error variance σ_R^2 of a set of k estimates can be written as

$$\sigma_R^2 = \frac{1}{k}\sum_{i=1}^{k}(r_i - m_R)^2$$

where $r_i = \tilde{v}_i - v_i$ is the difference between the estimated value and the true value at the same location and the average error is:

$$m_R = \frac{1}{k}\sum_{i=1}^{k} r_i = \frac{1}{k}\sum_{i=1}^{k}(\tilde{v}_i - v_i)$$

If we are willing to assume that we have a mean error m_R equal 0 then we can simplify this equation to the formula:

$$\sigma_R^2 = \frac{1}{k}\sum_{i=1}^{k}(r_i - 0)^2 = \frac{1}{k}\sum_{i=1}^{k}(\tilde{v}_i - v_i)^2$$

We cannot get very far with this equation for the error variance because it calls for knowledge of the true values.

To get out of this unfortunate situation we will again turn to random function models. To form our estimate we use a weighted linear combination of available samples:

$$\tilde{V}(x_0) = \sum_{i=1}^{n} w_i \cdot V(x_i)$$

Our error will be the difference between the estimate and the corresponding true value:

$$R(x_0) = \tilde{V}(x_0) - V(x_0)$$

We will minimize the variance of our modeled error $R(x_0)$ by finding the expression for the modeled error variance $\tilde{\sigma}_R^2$ and setting to zero the partial derivatives of this expression.

We can express the variance of the error as:

$$\begin{aligned}
&Var\{R(x_0)\} = \\
&= Cov\left\{\tilde{V}(x_0)\tilde{V}(x_0)\right\} - Cov\left\{\tilde{V}(x_0)V(x_0)\right\} - \\
&- Cov\left\{\tilde{V}(x_0)V(x_0)\right\} + Cov\left\{V(x_0)V(x_0)\right\} = \\
&= Cov\left\{\tilde{V}(x_0)\tilde{V}(x_0)\right\} - 2Cov\left\{\tilde{V}(x_0)V(x_0)\right\} + \\
&+ Cov\left\{V(x_0)V(x_0)\right\}
\end{aligned}$$

The above formula is a sum of three terms. Using the another formula for the variance of a weighted linear combination

$$Var\left\{\sum_{i=1}^{n} w_i V_i\right\} = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \cdot w_j \cdot Cov\left\{V_i\ V_j)\right\}$$

we obtain the first term of $Var\{R(x_0)\}$ as

$$\begin{aligned}
&Cov\left\{\tilde{V}(x_0)\tilde{V}(x_0)\right\} = Var\left\{\tilde{V}(x_0)\right\} = \\
&= Var\left\{\sum_{i=1}^{n} w_i V_i\right\} = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i\, w_j\, \tilde{C}_{ij}
\end{aligned}$$

The third term of $Var\{R(x_0)\}$ is $Cov\left\{V(x_0)V(x_0)\right\}$, which is the variance of $V(x_0)$. If we assume that all of our random variables have the same variance $\tilde{\sigma}^2$ this third term can be expressed as

$$Cov\left\{V(x_0)V(x_0)\right\} = \tilde{\sigma}^2$$

Remembering, that

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

the second term of $Var\{R(x_0)\}$ can be written as

$$2Cov\left\{\tilde{V}(x_0)V(x_0)\right\} = 2Cov\left\{\left(\sum_{i=1}^{n} w_i \cdot V_i\right)V_0\right\} =$$

$$= 2E\left\{\sum_{i=1}^{n} w_i V_i \cdot V_0\right\} - 2E\left\{\sum_{i=1}^{n} w_i V_i\right\} \cdot E\{V_0\} =$$

$$= 2\sum_{i=1}^{n} w_i \cdot E\left\{V_i \cdot V_0\right\} - 2\sum_{i=1}^{n} w_i \cdot E\{V_i\} \cdot E\{V_0\} =$$

$$= 2\sum_{i=1}^{n} w_i \cdot Cov\left\{V_i, V_0\right\} = 2\sum_{i=1}^{n} w_i \cdot \tilde{C}_{i0}$$

Combining these three terms again we have the following expression for the error variance $Var\{R(x_0)\} = \tilde{\sigma}_R^2$:

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \tilde{C}_{ij} - 2\sum_{j=1}^{n} w_i \cdot \tilde{C}_{i0}$$

If we try to tackle the minimization of $\tilde{\sigma}_R^2$ as an unconstrained problem, we run into difficulties. Setting the n partial first derivatives of $\tilde{\sigma}_R^2$ to 0 will produce n equations and n unknowns. The unbiased ness condition will add another equation without adding any more unknowns. This leaves us with a system of $n+1$ equations and only n unknowns. To avoid a problem we introduce another unknown into our equation for $\tilde{\sigma}_R^2$. The new variable is called μ and it is the Lagrange parameter. The technique of Lagrange parameters is a procedure for converting a constrained minimization problem into an unconstrained one.

Remembering the condition of unbiasedness $\sum_{j=1}^{n} w_i = 1$ we can write

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \tilde{C}_{ij} -$$

$$-2\sum_{i=1}^{n} w_i \cdot \tilde{C}_{i0} + 2\mu\left(\sum_{j=1}^{n} w_i - 1\right)$$

where $2\mu\left(\sum_{j=1}^{n} w_i - 1\right) = 0$.

The addition of this new term, which does not affect the equality, is all we need to convert our constrained minimization problem into an unconstrained one. The error variance is now a function of $n+1$ variables, the n weights and the one Lagrange parameter. By setting the $n+1$ partial first derivatives to 0 with respect to each of these variables, we will have a system of $n+1$ equations and $n+1$ unknowns. Setting the partial first derivative to 0 with respect to μ will produce our unbiased ness condition. Since the solution of those $n+1$ equations will produce the set of weights that minimizes $\tilde{\sigma}_R^2$ under the constraint that the weights sum to 1. A value μ, as we see later, is useful for calculating the resulting minimized error variance.

The differentiation of $\tilde{\sigma}_R^2$ with respect to the weights and setting to 0 produces the following equations:

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_1} =$$

$$= 2\sum_{j=1}^{n} w_j\tilde{C}_{1j} - 2\tilde{C}_{10} + 2\mu = 0 \Rightarrow \sum_{j=1}^{n} w_j\tilde{C}_{1j} + \mu = \tilde{C}_{10}$$

....

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_i} =$$

$$= 2\sum_{j=1}^{n} w_j\tilde{C}_{ij} - 2\tilde{C}_{i0} + 2\mu = 0 \Rightarrow \sum_{j=1}^{n} w_j\tilde{C}_{ij} + \mu = \tilde{C}_{i0}$$

....

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_n} =$$

$$= 2\sum_{j=1}^{n} w_j\tilde{C}_{nj} - 2\tilde{C}_{n0} + 2\mu = 0 \Rightarrow \sum_{j=1}^{n} w_j\tilde{C}_{nj} + \mu = \tilde{C}_{n0}$$

$$\sum_{i=1}^{n} w_i = 1$$

This system of equations can be written in matrix notation as

$$\mathbf{C} \quad * \quad \mathbf{w} \quad = \quad \mathbf{D}$$

$$\begin{vmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix} * \begin{vmatrix} w \\ \vdots \\ w_n \\ \mu \end{vmatrix} = \begin{vmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{vmatrix}$$

$$(n+1)*(n+1) \qquad (n+1)*1 \qquad (n+1)*1$$

The set of equations has a form:

$$C * w = D$$

and

$$w = \boldsymbol{C}^{-1} \cdot \boldsymbol{D}$$

is the solution vector of weights.

Multiplying each of the n equations of the set by w_i produces the following result:

$$w_i\left(\sum_{j=1}^{n} w_j \tilde{C}_{ij} + \mu\right) = w_i \tilde{C}_{i0} \quad \text{for} \quad i = 1,\ldots, n$$

Summing these n equations leads to

$$\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \tilde{C}_{ij} + \sum_{i=1}^{n} w_i \mu = \sum_{i=1}^{n} w_i \tilde{C}_{i0}$$

This formula can be written as

$$\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \tilde{C}_{ij} = \sum_{i=1}^{n} w_i \tilde{C}_{i0} - \sum_{i=1}^{n} w_i \mu$$

Since the weights sum to 1, the last term is simply μ, which gives us

$$\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \tilde{C}_{ij} = \sum_{i=1}^{n} w_i \tilde{C}_{i0} - \mu$$

Substituting this into equation for $\tilde{\sigma}_R^2$ allows us to express the minimized error variance as

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^{n} w_i \tilde{C}_{i0} - \mu - 2\sum_{i=1}^{n} w_i \cdot \tilde{C}_{i0} =$$

$$= \tilde{\sigma}^2 - (\sum_{i=1}^{n} w_i \cdot \tilde{C}_{i0} + \mu)$$

We have assumed that all the random variables in our random function model had the same mean and variance. These two assumptions allow us to develop the following relationship between the model variogram and the model covariance:

$$\tilde{\gamma}_{ij} = \frac{1}{2} E\left\{ \left[V_i - V_j \right]^2 \right\} =$$

$$= \frac{1}{2} E\left\{V_i^2\right\} + \frac{1}{2} E\left\{V_i^2\right\} - E\left\{V_i \cdot V_j\right\} = E\left\{V_i^2\right\} - E\left\{V_i \cdot V_j\right\}$$

$$= E\left\{V_i^2\right\} - m^2 - \left(E\left\{V_i \cdot V_j\right\} - m^2 \right) = \tilde{\sigma}^2 - \tilde{C}_{ij}$$

Now we will show that the ordinary kriging system

$$\sum_{j=1}^{n} w_j \tilde{C}_{ij} + \mu = \tilde{C}_{i0} \quad \text{for} \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} w_i = 1$$

can be written in terms of variogram as

$$\sum_{j=1}^{n} w_j \tilde{\gamma}_{ij} - \mu = \tilde{\gamma}_{i0} \quad for \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} w_i = 1$$

with the modeled error variance given by:

$$\tilde{\sigma}_R^2 = \sum_{i=1}^{n} w_i \cdot \tilde{\gamma}_{i0} + \mu$$

So because:

$$\tilde{\gamma}_{ij} = \tilde{\sigma}^2 - \tilde{C}_{ij}$$

we can write:

$$\tilde{C}_{ij} = -\tilde{\gamma}_{ij} + \tilde{\sigma}^2$$

and from:

$$\sum_{j=1}^{n} w_j \tilde{C}_{ij} + \mu = \tilde{C}_{i0}$$

we obtain:

$$\sum_{j=1}^{n} w_j (-\tilde{\gamma}_{ij} + \tilde{\sigma}^2) + \mu = -\tilde{\gamma}_{i0} + \tilde{\sigma}^2$$

Then:

$$\sum_{j=1}^{n} ((-w_j \tilde{\gamma}_{ij}) + w_j \tilde{\sigma}^2) + \mu = -\tilde{\gamma}_{i0} + \tilde{\sigma}^2$$

and then:

$$-\sum_{j=1}^{n} (w_j \tilde{\gamma}_{ij}) + \tilde{\sigma}^2 \sum_{j=1}^{n} w_j) + \mu = -\tilde{\gamma}_{i0} + \tilde{\sigma}^2$$

Because $\sum_{j=1}^{n} w_j = 1$

hen we obtain:

$$-\sum_{j=1}^{n} w_j \tilde{\gamma}_{ij} + \tilde{\sigma}^2 + \mu = -\tilde{\gamma}_{i0} + \tilde{\sigma}^2$$

and from here:

$$\sum_{j=1}^{n} w_j \tilde{\gamma}_{ij} - \mu = \tilde{\gamma}_{i0}$$

what is the wanted formula.

3. The directional variogram

In the above picture we have the following relations:

$$\gamma_1(1) = \gamma_a(a)$$

$$\gamma_1(h/a) = \gamma_a(h)$$

If we calculate the model with range value 1 in the point h/a then we obtain the same value as for the model with range value a in the point h. So we can reduce the model with range value a to the equivalent model with range value 1 substituting the distance h by h/a.

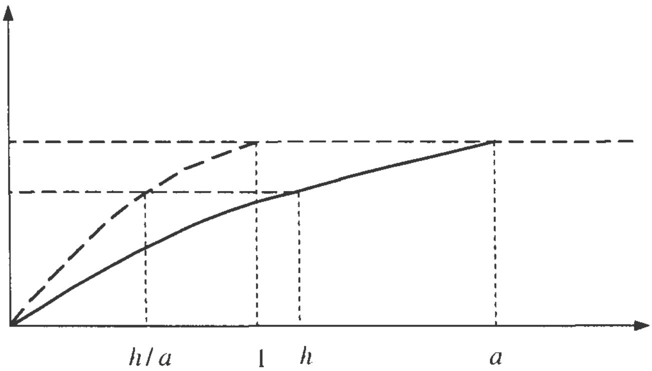

Figure 1. An example of the variogram.

The equivalence can be written as:

$$\gamma_1(h/a) = \gamma_a(h)$$

or

$$\gamma_1(h) = \gamma_a(ah)$$

If $h_1 = h/a$ then:

$$\gamma_1(h_1) = \gamma_a(h)$$

Concluding, each directional model with range value a can be reduced to standarized model with range value 1 substituting value h by reduced value h/a.

The idea of balanced model can be extended for two dimensions. If a_x is a range value in x direction and a_y is a range value in y direction, then the anisotropic model of variogram can be written as:

$$\gamma(h) = \gamma(h_x, h_y) = \gamma_1(h_1)$$

where

$$h_1 = \sqrt{\left(\frac{h_x}{a_x}\right)^2 + \left(\frac{h_y}{a_y}\right)^2}$$

where h_x is the component of h along x axis and h_y is the component of h along y axis.

Similarly, the anisotropic variogram model in three dimensions, with ranges a_x, a_y and a_z can be expressed as:

$$\gamma(h) = \gamma(h_x, h_y, h_z) = \gamma_1(h_1)$$

and the reduced distance h_1 is given by:

$$h_1 = \sqrt{\left(\frac{h_x}{a_x}\right)^2 + \left(\frac{h_y}{a_y}\right)^2 + \left(\frac{h_z}{a_z}\right)^2}$$

where h_x, h_y and h_z are the components of h in x, y and z directions of the anisotropy axes and $\gamma_1(h_1)$ is the equivalent model with a standardized range of 1.

4. Variogram models

The empirical variogram is calculated as a half of average squared difference between the paired data values:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(ij)|h_{ij}=h} (v_i - v_j)^2$$

This empirical variogram is modeled using the following most popular models (see Fig. 2):

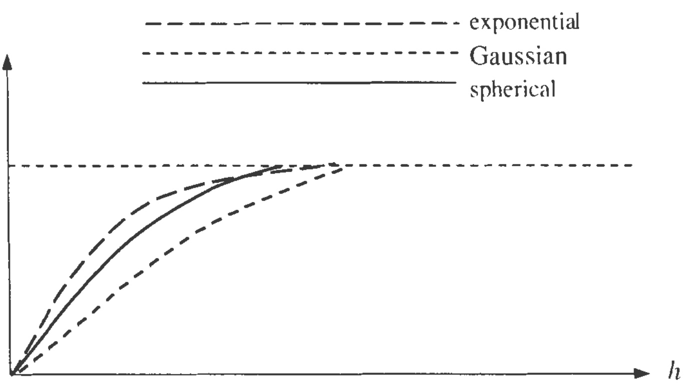

Figure 2. Three exemplary functions for modeling the variograms.

- ***Spherical model***

$$\gamma(h) = \begin{cases} 1{,}5h/a - 0{,}5(h/a)^3 & \text{for } h \le a \\ 1 & \text{otherwise} \end{cases}$$

- ***Exponential model***

$$\gamma(h) = 1 - \exp(-\frac{3h}{a})$$

- ***The Gaussian Model***

$$\gamma(h) = 1 - \exp(-\frac{3h^2}{a^2})$$

5. Computer calculations of kriging approximation

For the above algorithm of kriging approximation a computer program has been developed and some calculation of the approximation of yearly rainfall values for the area of Poland were performed.

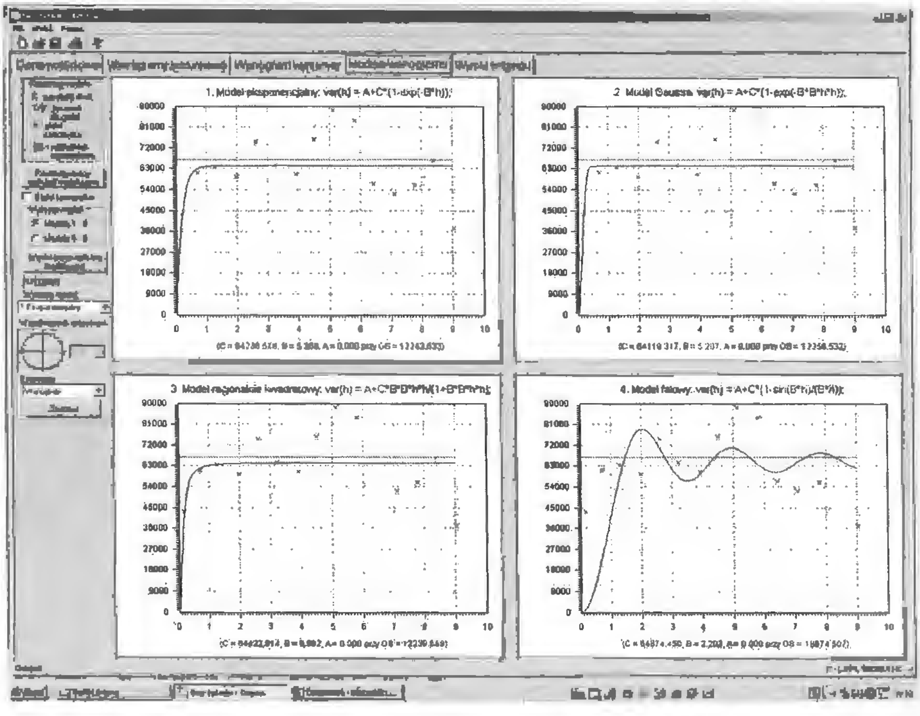

Figure 3. The program screen for model fitting with four different modeling functions (*from left to right and form up to down: exponential, Gaussian, quadratic and wave functions*).

An exemplary screen of the program developed with four functions of eight on the whole that are used to model the empirical variograms calculated with the measurement data is shown on Fig. 3.

Some calculation results for three measurement which were firstly collected and then approximated with the kriging algorithm are shown in Tables 1 and 2. The measurement values ranged from the minimal value to the maximal one in the whole area investigated.

Table 1. Data for making the kriging approximation.

Point of measurement	Coordinates		Parameter value	Parameter range
1	17,09	53,39	89	min.
2	19,59	49,14	2241	max.
3	19,24	53,15	759	medium

Table 2. Results of the Kriging approximation for rainfall data.

Model	Point 1 Error in %	Point 2 Error in %	Point 3 Error in %
Gaussian	806 (900 %)	1092 (51 %)	758 (-)
Quadratic	776 (870 %)	1141 (49 %)	776 (2 %)
Linear	755 (840 %)	1032 (53 %)	722 (4 %)
Spherical	719 (800 %)	1365 (39 %)	792 (4 %)
Exponential	760 (850 %)	1191 (46 %)	779 (2 %)

The conclusion resulted from these calculation is that the kriging algorithm can approximate quite well only the values which are similar to the average value of the measurements taking into consideration by the approximation. On the other side the kriging approximation is very deceptive by the calculation of the values which are extremely different from these average ones. This conclusion is justified at least by approximating the rainfall data or the data with very varying values.

References

Isaaks E.H. (1989) *Applied Geostatistics*. New York-Oxford, Oxford University Press.

Olea R.A. (1999) *Geostatics for Engineers and Earth Scientists*. Kluwer Academic Publishers.

Cressie Noel A.C. (1993) *Statistic for Spatial Data*. John Wiley and Sons, Inc. New York.

Jan Studzinski, Olgierd Hryniewicz (Editors)

MODELLING CONCEPTS AND DECISION SUPPORT IN ENYIRONMENTAL SYSTEMs

This book presents the papers that describe the most interesting results of the research that have been obtained during the last few years in the area of environmental engineering and environment protection at the Systems Research Institute of the Polish Academy of Sciences in Warsaw and the Leibniz-Institute of Freshwater Ecology and Inland Fisheries in Berlin (IGB). The papers were presented during the First Joint Workshop organized at the IGB in February 2006. They deal with mathematical modeling, development and application of computer aided decision making systems in the areas of the environmental engineering concerning groundwater and soil, rivers and lakes, water management and regional pollution.

ISBN-10: 83-894-7505-7
ISBN-13: 978-83-894750-5-3
ISSN 0208-8029