POLISH ACADEMY OF SCIENCES
Systems Research Institute

ECO – INFO AND SYSTEMS RESEARCH

Editors:
Jan Studzinski
Olgierd Hryniewicz

ECO – INFO AND SYSTEMS RESEARCH

Polish Academy of Sciences • Systems Research Institute

Series: SYSTEMS RESEARCH
Vol. 52

Series Editor:

Prof. Jakub Gutenbaum

Warsaw 2006

ECO – INFO AND SYSTEMS RESEARCH

Editors:

Jan Studzinski
Olgierd Hryniewicz

The purpose of this publication is to present the information technology (IT) tools and techniques that have been developed at the Systems Research Institute of Polish Academy of Sciences in Warsaw (IBS PAN) and at the German Institute for Landscape System Analysis in Müncheberg (ZALF) in the area of applications of informatics in environmental engineering and environment protection. The papers published in this book were presented in the form of extended summaries during a special workshop organized by IBS PAN in Szczecin in September 2006 together with the conference BOS'2006 organized jointly by IBS PAN, University of Szczecin, and the Polish Society of Operational and Systems Research. In the papers the problems of mathematical modeling, approximation and visualization of environmental variables are described. Moreover, some questions concerning the environmental economy are also presented.

Papers Reviewers:
Prof. Andrzej Straszak
Dr. Lucyna Bogdan

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 83-894-7509-X
9788389475091
ISSN 0208-8029

CHAPTER 1

Specialized Software

KRIGING APPROXIMATION: ALGORITHMS, PROGRAM AND CALCULATION RESULTS

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

Abstract: *In the paper the development of the computer program KRIPOS for approximating such meteorological data as temperature, rain and snow falls and soil erosivity is described. The calculation algorithms and mathematical models implemented into the program are presented and their usefulness and effectiveness by approximating the yearly rain falls for the whole area of Poland are tested.*

Keywords: Kriging algorithms, approximation of meteorological data, environmental monitoring.

1. Introduction

The kriging algorithms have been developed and are commonly used to approximate in space and time environmental variables such as rain and snow falls, soil composition, soil erosivity etc. as well as to develop contour line maps of the above variables. In current situation there is impossible to develop such the maps for Poland with the data which are got only from the measurement points located at the national monitoring stations. The network of these measurements points is too rare and irregularly divided to get the maps with sufficient good accuracy and because of that the approximation of the values of these variables must be done in many new points.

The aim of this work is to check the usefulness of kriging approximation for solving this problem and to do it an own computer program has been developed and tested on some real measurement data.

2. The algorithm

Kriging approximation means the estimation of unknown values of a variable in some selected points of an area (called calculation points) on the base of the known values of this variable which have been collected in other points of this area (called measurement points). Investigation of kriging algorithms belongs to the geostatistics research area.

The main function of geostatistics is the following variogram function

$$D^2[Z(x+h) - Z(x)] = E[Z(x+h) - Z(x)]^2 = 2\,\square(h)$$

with: $Z(x)$ – the variable investigated, D, E – variance and the average value of $Z(x)$, $2\square(h)$ – variogram function, $\square(h)$ – semivariogramm function and $Z(x)$, $Z(x+h)$ – variable values in points x and $(x+h)$ for $h \in [0,+\infty]$.

The calculation algorithm of the kriging approximation consists of 4 following steps:

Step 1: Calculation of the semivariogram on the base of the measurement data.

$$\gamma(h) = \frac{1}{2n_h} \sum_{i=1}^{n_h} (z_{h+1} - z_i)^2$$

where z_i, z_{i+h} are the values of the variable investigated in the points which are outlying up to the distance h from each other and n_h is the number of the pairs of points outlying up to h from each other.

Step 2: Modeling of the semivariogram with different analytical functions using the methods of static optimization.

Step 3: Calculation of the weight coefficients on the base of the conditions of the unbiaseness and maximal effectiveness of the variable estimator.

$$E(z_i - m) = 0$$

$$\sigma^2 = E[(z_i - m)^2] = \min$$

As the result of the conditions the equations system follows

$$\gamma(x_j, x_0) = \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) + \mu$$

for j=1,2,...,N, and in the matrix form

$$\begin{vmatrix} y_{11} & \cdots & y_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ y_{N1} & \cdots & y_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix} \cdot \begin{vmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{vmatrix} = \begin{vmatrix} y_{10} \\ \vdots \\ y_{N0} \\ 1 \end{vmatrix}$$

from which the weight coefficients can be calculated.

Step 4: Calculation of the value of the variable investigated in the new selected point.

$$z_o = \sum_{i=1}^{N} \lambda_i z_i$$

where B_i are the weight coefficients and z_i are the variable values in the known measurement points.

3. The computer program

On the base on the above calculation algorithm the computer program Kripos has been developed in the computer language C++. Into the program 8 analytic functions for semivariogram modeling have been implemented. The identification of the coefficient values in the modeling functions occurs with the Marquardt algorithm of static optimization.

The following modeling functions have been implemented:

1. Exponential function: $\gamma(h) = C(1 - \exp(-Bh))$
2. Gauss function: $\gamma(h) = C(1 - \exp(-B^2 h^2))$
3. Square function: $\gamma(h) = CB^2 h^2 (1 + B^2 h^2)$
4. Wave function: $\gamma(h) = C(1 - \sin(Bh) / Bh)$
5. Spherical function: $\gamma(h) = CBh(1{,}5 - 0{,}5 B^2 h^2))$
6. Power function 1: $\gamma(h) = C(Bh)^{1/4}$
7. Power function 2: $\gamma(h) = C(Bh)^{1/2}$
8. Linear function: $\gamma(h) = CBh$

4. Calculation results

For checking and performing the Kripos program the rain fall data of 1980 from 106 measurement stations in Poland have been taken and the approximation of the variable for the whole area of Poland has been done.

Different calculation options of the program have been used and discussed, such as:

- Omnidirectional kriging
- Directional kriging
- Distance calculation between the measurement points in geographical degrees and kilometers
- Kriging approximation in 2D and 3D space
- Kriging approximation in unlimited and limited areas.

The results of applying the program for the approximation tasks with the data mentioning are shown on the several following figures.

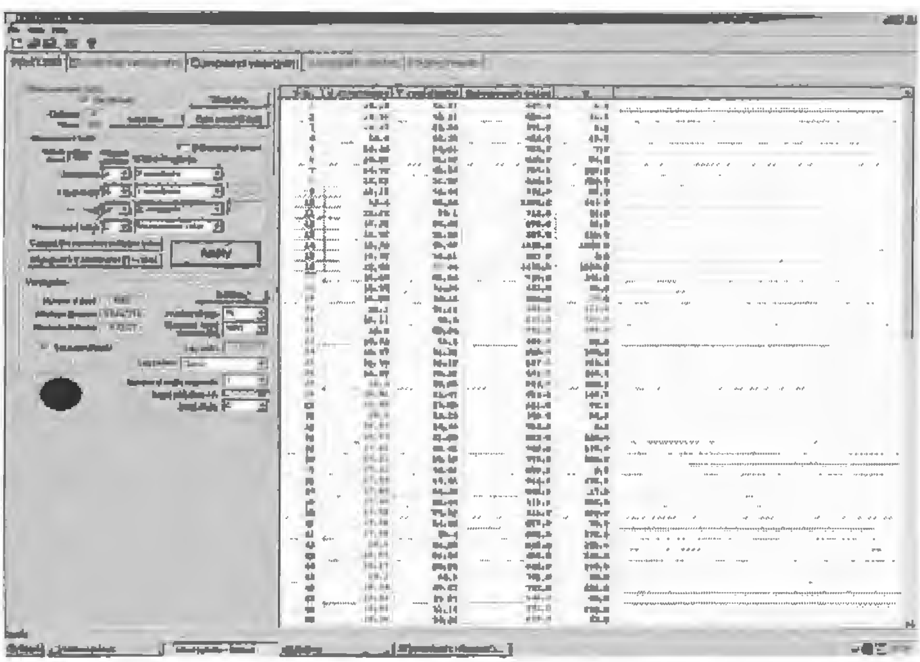

Figure 1. Measurement data from all measurement points for omnidirectional kriging with distances in degrees and in 2D space (example 1).

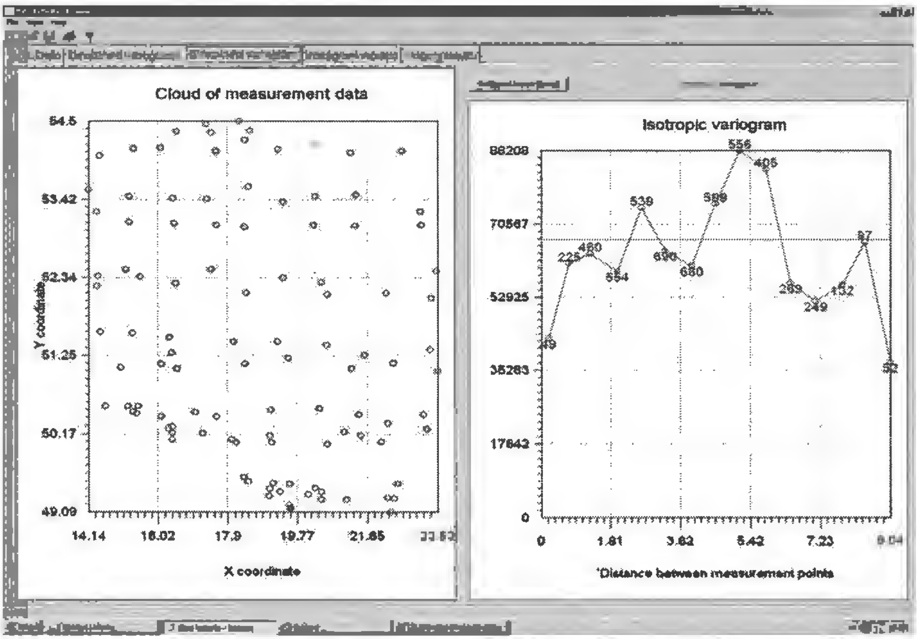

Figure 2. Variogram function for example 1.

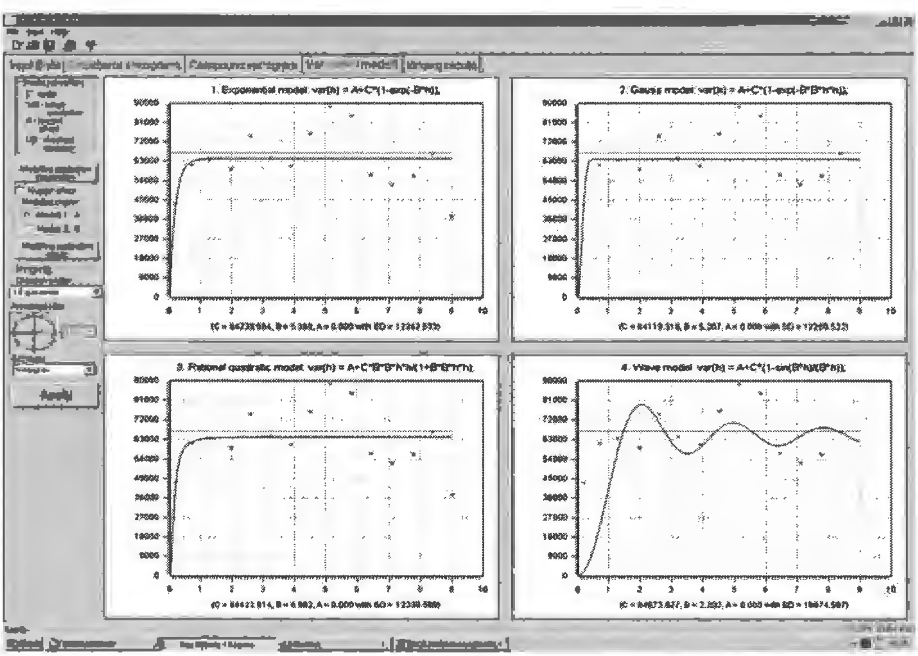

Figure 3. First 4 variogram models for example 1.

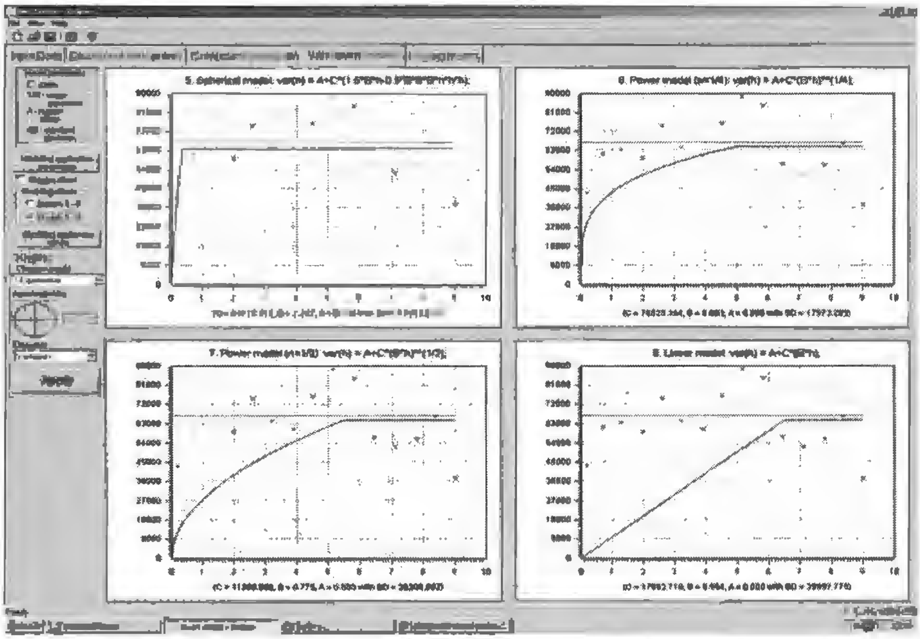

Figure 4. Second 4 variogram models for example 1.

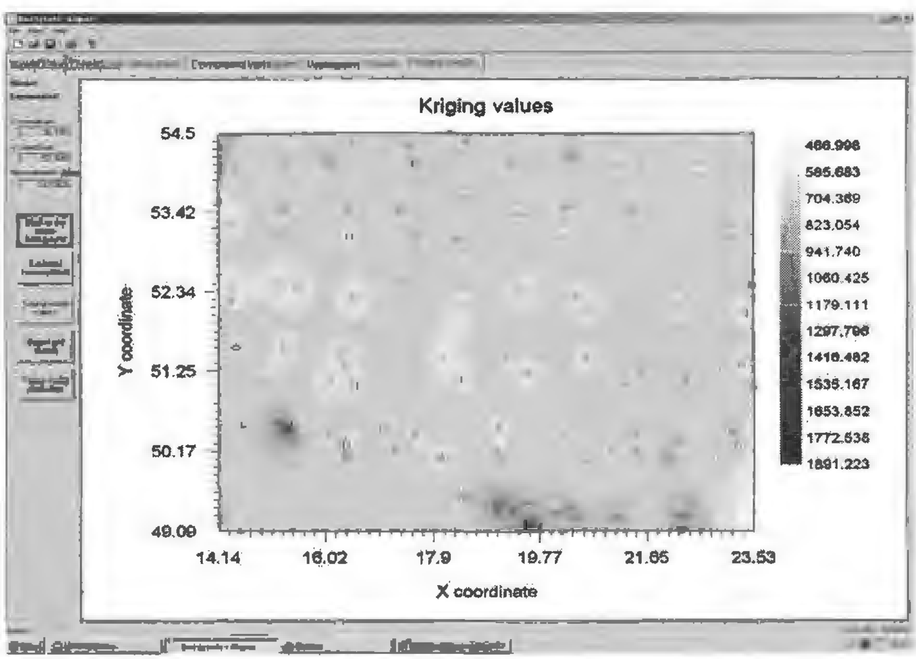

Figure 5. Contour lines for model 1 of example 1.

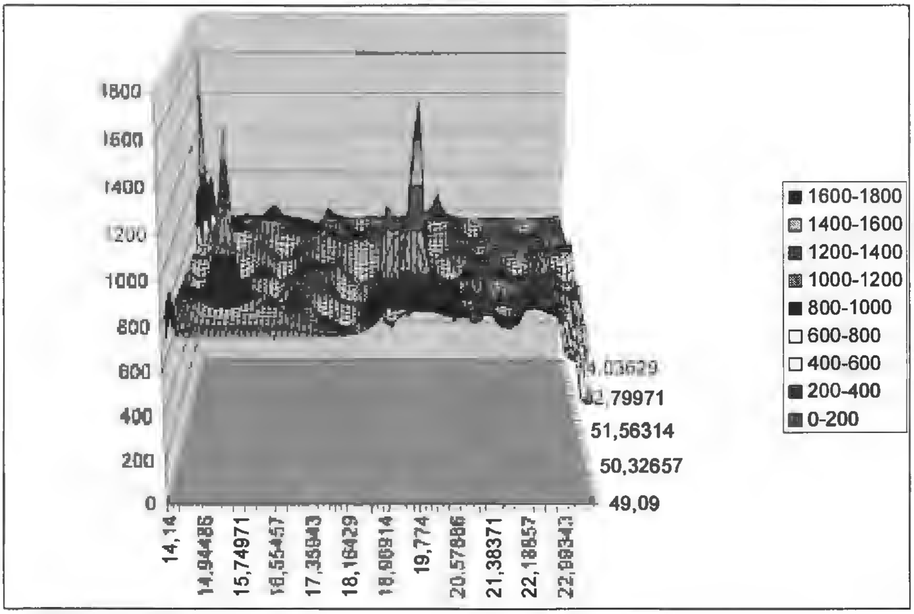

Figure 6. 3D results visualization for model 1 of example 1.

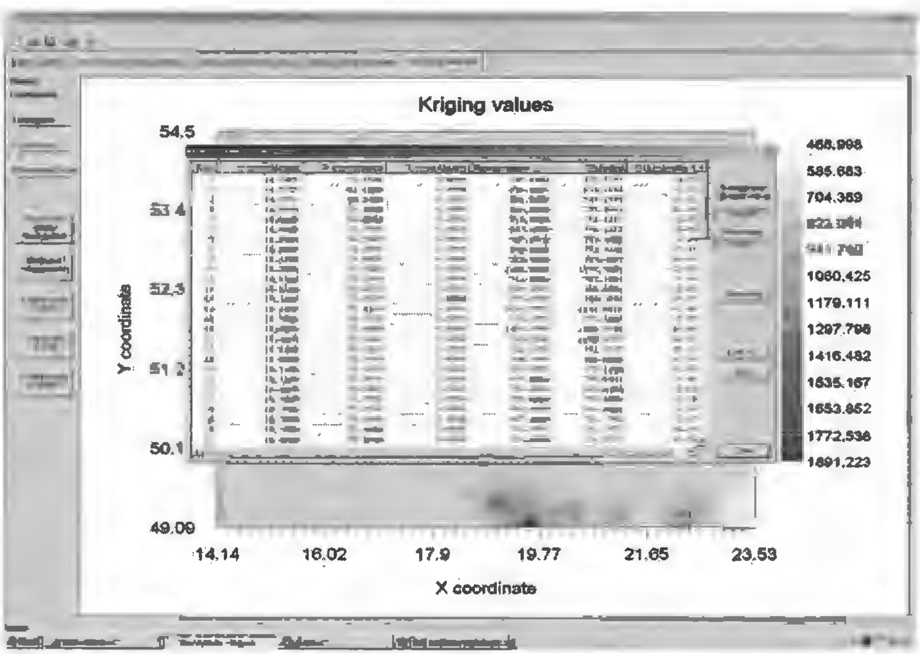

Figure 7. Numerical results of approximation for model 1 of example 1.

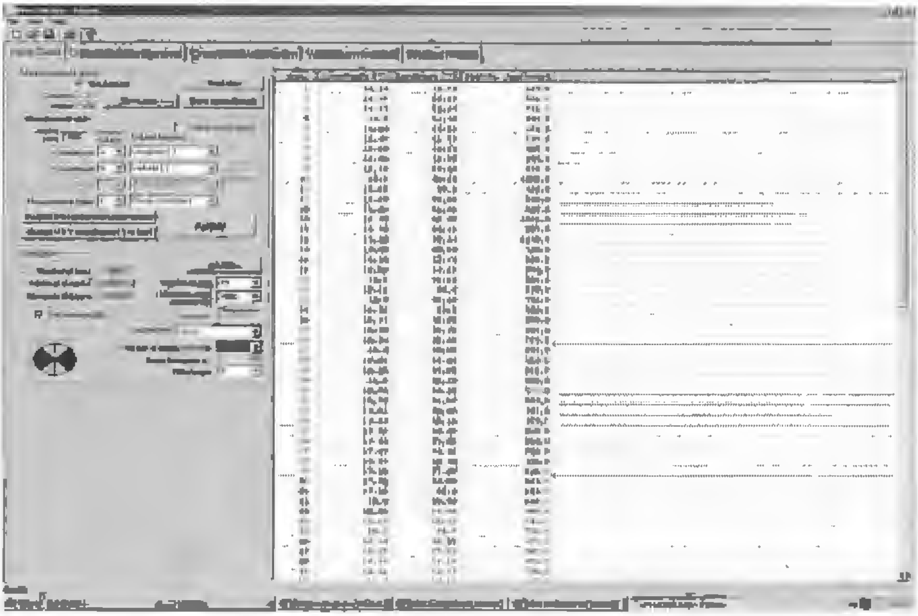

Figure 8. Measurement data for directional kriging for all measurement points with distances in degrees, 2 angle segments and in 2D space (example 2).

Figure 9. Directional variograms with anisotropy ratio 2,5 for example 2.

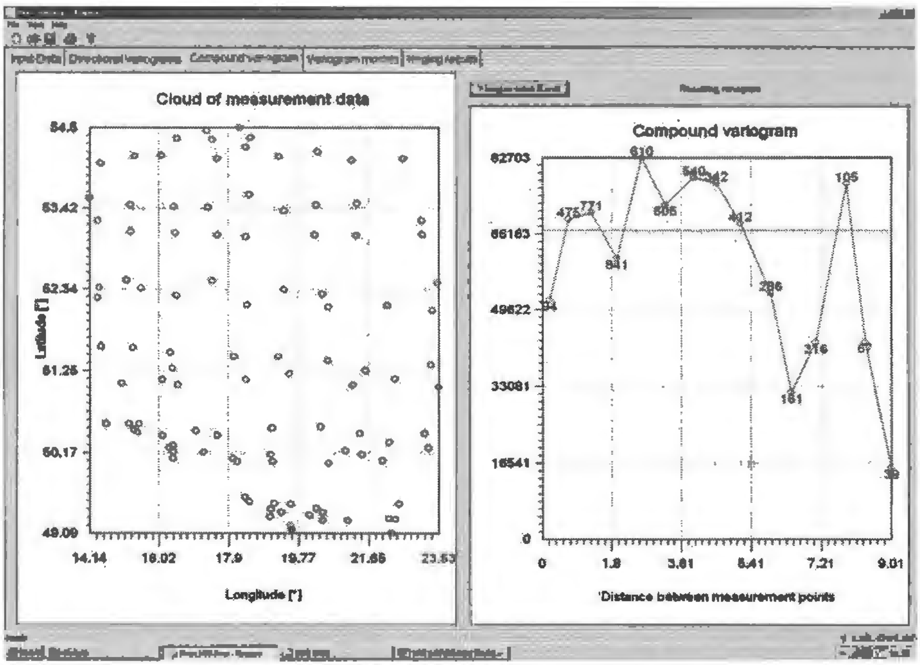

Figure 10. Compound variogram for example 2.

On the Figures 1-10 some selected windows of the developed program KRIPOS and also the results of its using for the approximation of the rain falls variable are presented. The data used for the calculation are the yearly rain falls values that have been collected in 1980 at 106 monitoring stations divided across whole Poland (see Figures 1 and 2). The aim of the calculation was to test the program with its all algorithms included and to verify the usefulness of the kriging algorithms for the rain falls approximation.

The Figures 1-7 show the results of approximation with the use of the omnidirectional (or: directionless) kriging algorithm to calculate the discrete variogram function and of the exponential function (model 1) to get the continuous variogram model (example 1). By the modeling the discrete variogram (see Figure 2) eight analytical functions are used (see Figures 3 and 4) and the choice of only one model for farther calculation occurs generally on the base of residuum values got for all models: the model with the lowest residuum value is considered then as the best one. One can see in Table 1 that model 1 with the residuum value of 12.242 fits best the discrete variogram points and because of that the following rain falls approximation was made with the use of it. The approximation results can be visualized on the 2D or 3D space after a discrete net to cover the investigated area has been defined and the kriging approximation in the node points of this net has been realized (see Figures 5 and 6).

The verification of the quality of kriging approximation regarding the rain falls variable was made by sequential removing of single measurement points from the whole data table containing the data from 106 monitoring stations and by subsequent calculation of the rain falls value at the points removed. For the right values were known at the points removed then getting the approximated values for the same points the quality of approximation could be done by the simple comparison of measured and calculated values (see Figure 7). After such calculation is made the quality of approximation realized can be assessed by computing the average absolute error and the average percentage error computed for the variable values at all measurement points (see Table 1).

Unfortunately such the assessment of the approximation is not univocal in every case of calculation what is to see in Table 1 by comparing the results got for model 1 and model 3 (square function): the approximation made with the use of model 3 is better than this one made with model 1 regarding their average percentage errors but at the same time it is worse regarding their average absolute errors. However there is consoling that the differences between the both approximation runs are not big and practically negligible.

On Figures 8-10 and also in the bottom part of Table 1 the results got from the directional kriging approximation made for two angle segments and with the use of model 2 (Gauss function) are presented (example 2). While combining two directional variograms into one combined variogram the anisotropy ratio of 2,5 was settled (see Figures 9 and 10). This means that the directional variogram oriented horizontally weighs essentially more by creating the compound variogram than the second directional variogram oriented orthogonally.

The results of calculation received with the directional approximation are worse than the previous ones and this conclusion is obvious now: both assessment criteria computed for the best model 2 are worse than in the case of the omnidirectional approximation realized with the best model 1 or with the second best model 3 (see Table 1). These results mean that the directional kriging approximation gives no better results than the directionless one by approximating the rain falls variable but this conclusion can not be generalized for other kinds of environmental variables.

Table 1. Comparison results for examples 1 and 2 of kriging approximation.

No.	Model	Residuum	Absolute error	Percentage error
Example 1: Omnidirectional kriging				
1	**Model 1**	**12.242**	**129,351**	**16,041**
2	Model 2	12.259		
3	**Model 3**	**12.239**	**129,527**	**15,939**
4	Model 4	18.674		
5	Model 5	12.259		
6	Model 6	17.573		
7	Model 7	25.208		
8	Model 8	35.997		
Example 2: Directional kriging				
9	Model 1	18.781		
10	**Model 2**	**18.779**	**143,962**	**17,881**
11	Model 3	18.819		
12	Model 4	25.285		
13	Model 5	18.779		
14	Model 6	26.074		
15	Model 7	33.307		
16	Model 8	42.702		

On Figures 11 and 12 the structures of the measurement data files for the approximation runs made in 2D and 3D spaces are shown (examples 3 and 4). The calculations have been made with the data transformed and the coordinates of the measurement points are given now in kilometers and not in degrees as before. The calculation results are presented in Table 2.

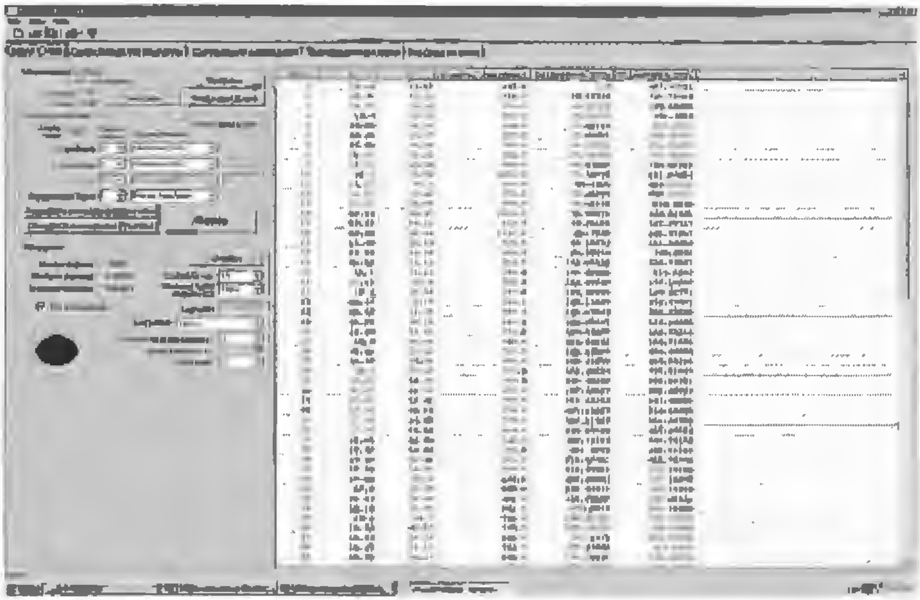

Figure 11. Measurement data for omnidirectional kriging for all measurement points with distances in kilometers and in 2D space (example 3).

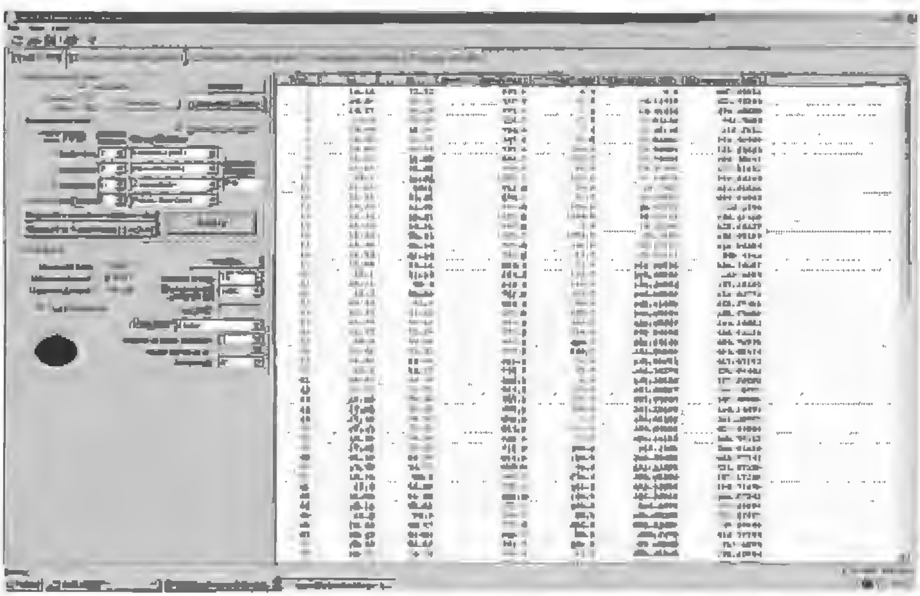

Figure 12. Measurement data for omnidirectional kriging for all measurement points with distances in kilometers and in 3D space; multiplier value for *x* and *y* coordinates: 0,01 (example 4).

By the 3D approximation another kind of data transformation is made consisting in multiplying the x and y coordinates of all points considered in the area investigated by a multiplier value in order to normalize all components of distances between these points. In our case the x and y coordinates are measured in kilometers while the z coordinate is measured in meters and this resulted in settling the multiplier value equal to 0,01.

By comparing the appropriate numbers in Tables 1 and 2 one can see that the kriging approximation made with distances measured in kilometers gives better results than this one with distances measured in degrees (see upper parts of Tables 1 and 2). On the other side there are no big differences between the approximation results while making the calculation in 2D or 3D space (see the upper and bottom parts of Table 2). The reason for this phenomenon can be the fact that the distances between the measurements points along the x and y coordinates are much bigger than these ones along the z coordinate and there is rather small and negligible influence of z on the calculation of distances between the points considered.

On Figures 13-17 and in Table 3 the results of rain falls approximation made for separated and limited regions of the area investigated are shown (example 5). For the whole Poland three regions have been selected concerning North, Middle and South Poland. The reason for such the selection was that the height differences of the ground levels are quite large between these regions and they are quite small inside of them: the area of North Poland is the coast of the Baltic See and in the area of South Poland several mountain chains are located. After the closed regions are defined using the computer mouse the approximation runs are made using only the measurement points which are located inside these regions.

Comparing the results obtained for three regions big differences in the quality of approximation can be observed. The best results have been got for Middle Poland where the ground is almost flat and the measurement points are located on quite the same height level. These results are nearly two times better than these ones when the approximation was made for the whole Poland (see the Tables 2 and 3). A little worse results were obtained for North Poland but nevertheless they are better than the results of the whole Poland approximation. The worst results of approximation have been got for South Poland where the measurement point heights are very differentiated and these results are not only worse than the previous two ones but also they are worse than the got for the whole Poland.

While realizing the calculation for the Middle Poland region some troubles with the choice of the best model for the continuous kriging approximation appear once again as they were noticed while discussing the results of example 1.

Table 2. Comparison results for examples 3 and 4 of kriging approximation.

No.	Model	Residuum	Absolute error	Percentage error
Example 3: 2D space calculation				
1	Model 1	14.309		
2	Model 2	14.346		
3	**Model 3**	**14.197**	**125,390**	**15,487**
4	Model 4	21.551		
5	Model 5	14.346		
6	**Model 6**	**13.737**	**108.203**	**12.769**
7	Model 7	19.828		
8	Model 8	30.419		
Example 4: 3D space calculation				
9	Model 1	13.185		
10	Model 2	13.269		
11	**Model 3**	**13.036**	**120.495**	**15.054**
12	Model 4	20.732		
13	Model 5	13.269		
14	**Model 6**	**12.089**	**109.689**	**13.098**
15	Model 7	17.881		
16	Model 8	28.883		

Under six models shown in Table 3 model 4 is best regarding its residuum value and in spite of that its approximation results are worse than these ones got with the use of models 1, 2 and 3.

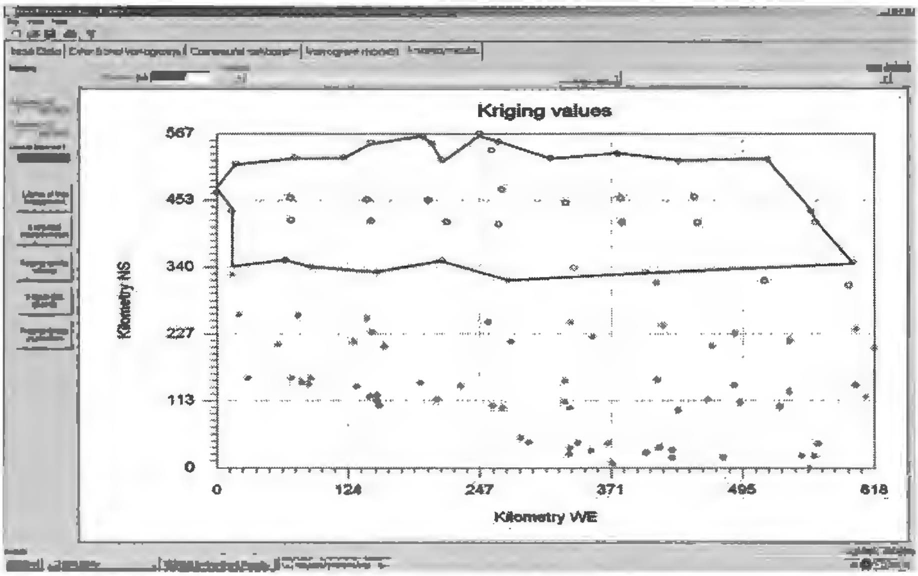

Figure 13. Omnidirectional kriging for selected measurement points in North Poland: distances in kilometers, 3D space calculation, multiplier value 0,01 (example 5).

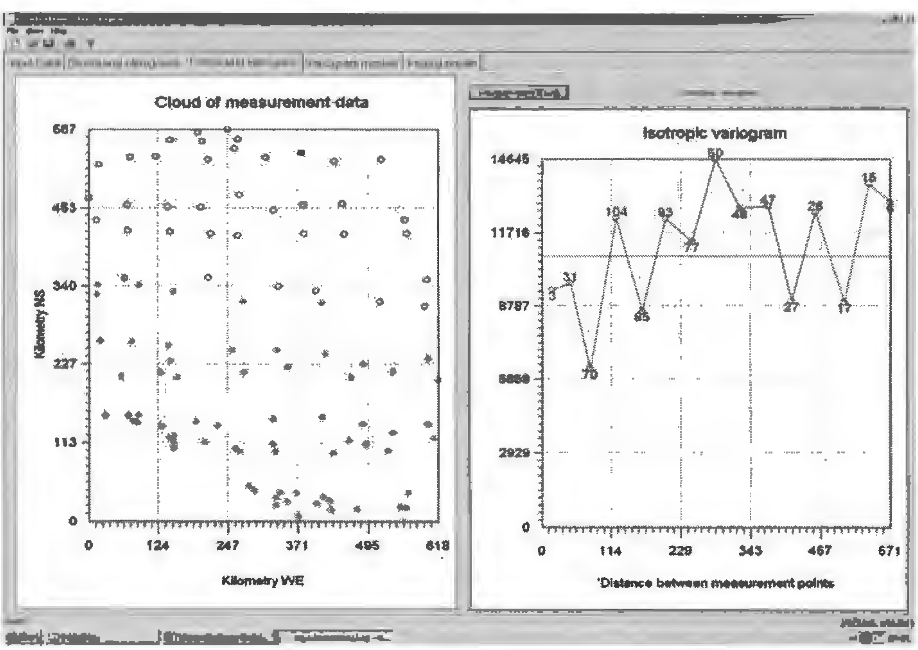

Figure 14. Variogram function for example 5.

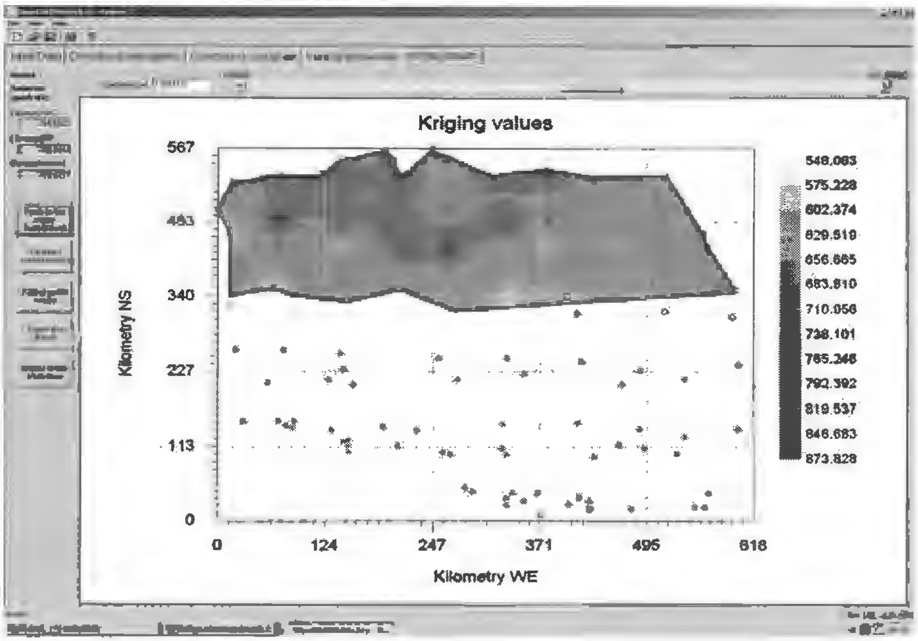

Figure 15. Contour lines for model 3 of example 5; drawn at the height level of 1 m for North Poland.

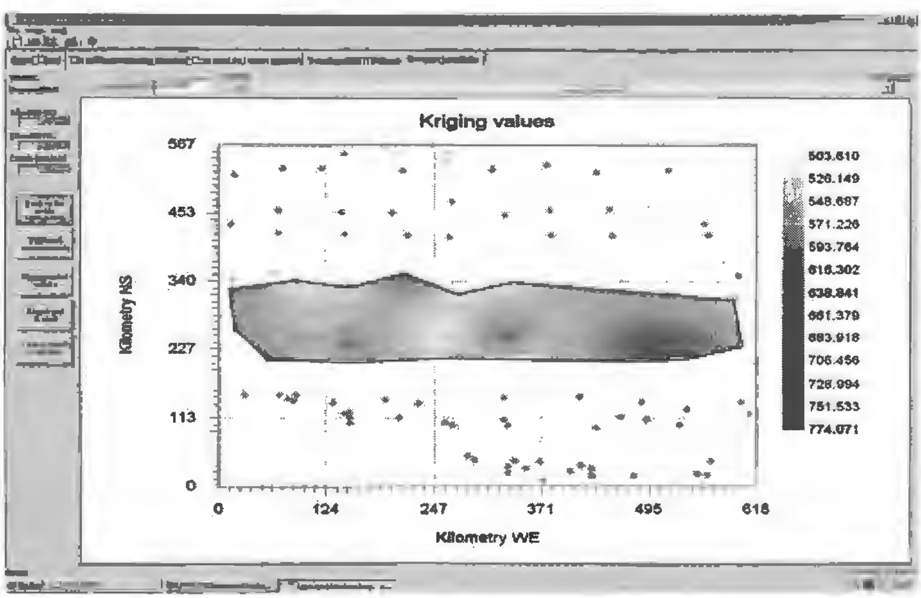

Figure 16. Contour lines for model 1 of example 5; drawn at the height level of 50 m for Middle Poland.

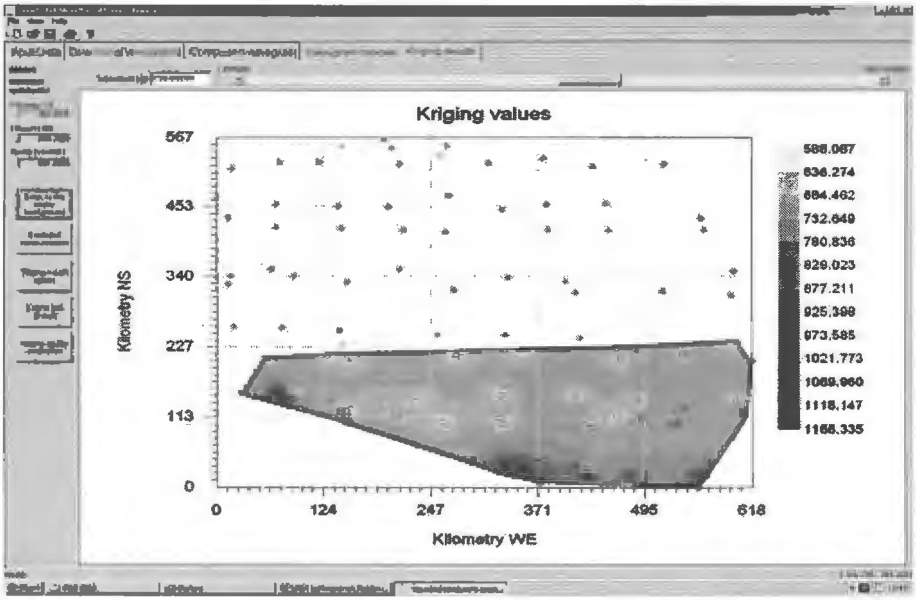

Figure 17. Contour lines for model 3 of example 5; drawn at the height level of 150 m for South Poland.

Also considering the results obtained for models 1, 2 and 3 one can see a small inconsequence by the results assessment: the best model under these three ones is model 2 but the best approximation is made with the use of model 1. Fortunately the result differences between them are not so large and they can be considered as negligible.

The result differences that have been got for models 1, 2 and 3 by the Middle Poland approximation are better visible on Figures 18-21. As it is to see in Table 3 in its rows 7 to 10 the best results of the approximation have been received using model 2 and the worst ones using model 4 while the results got with the models 1 and 3 are almost of the same quality. The differences in the approximation quality with model 4 and with models 1 to 3 measured with the average absolute and percentage errors are equal nearly to 100 % and these differences are also to notice on the Figures: the contour lines drawn on Figures 18 and 20 are very similar and these ones seen on Figure 19 are quite similar to them while the contour lines drawn on Figure 21 show a quite another and false picture.

The general conclusion coming out from the above observations is that there is rather risky to assess the approximation quality only on the base of the calculation made using the variogram model that have been chosen regarding only the model residuum value. The kriging approximation shall be done rather by using several models with similar residuum values and the assessment of approximation results shall be then done by comparing the error figures got for different approximation runs.

Table 3. Calculation results for example 5 of kriging approximation.

No.	Model	Residuum	Absolute error	Percentage error
Example 5: North Poland				
1	Model 1	2.200		
2	Model 2	2.204		
3	**Model 3**	**2.174**	**89,379**	**15,487**
4	Model 4	3.284		
5	Model 5	2.204		
6	Model 6	2.237		
Example 5: Middle Poland				
7	**Model 1**	**1.935**	**55,244**	**8,944**
8	**Model 2**	**1.909**	**55,486**	**8,992**
9	**Model 3**	**1.939**	**56,286**	**9,118**
10	**Model 4**	**1.801**	**93,675**	**15,558**
11	Model 5	2.053		
12	Model 6	2.230		
Example 5: South Poland				
13	Model 1	22.073		
14	Model 2	22.090		
15	**Model 3**	**22.022**	**169,667**	**18,930**
16	Model 4	34.388		
17	Model 5	22.090		
18	Model 6	29.634		

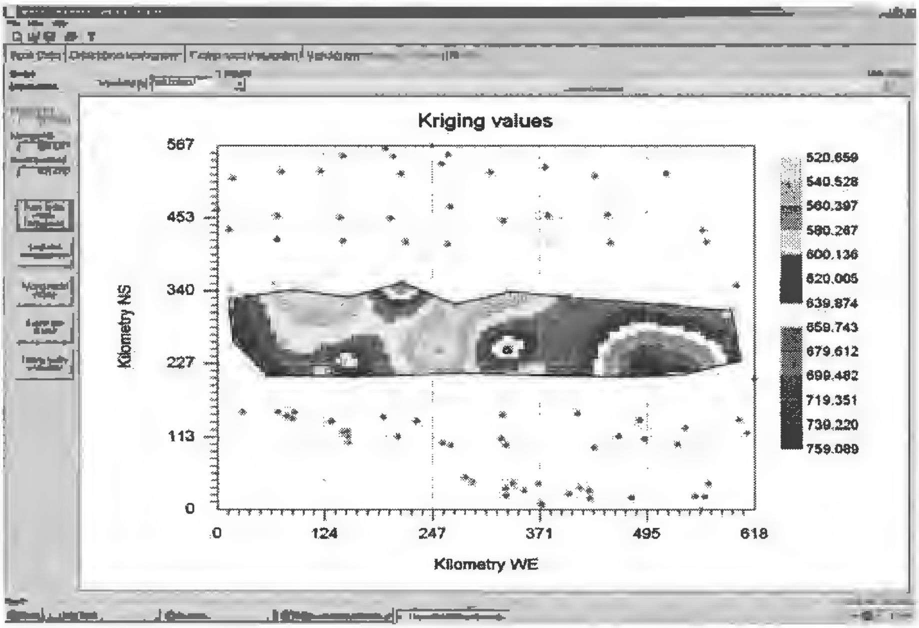

Figure 18. Contour lines for model 1 by Midlle Poland approximation of example 5.

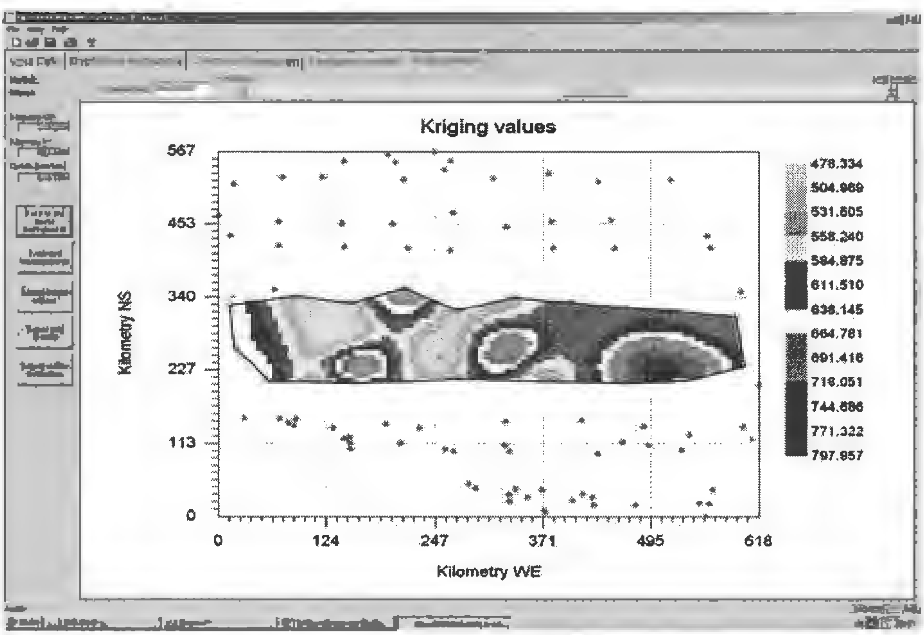

Figure 19. Contour lines for model 2 by Midlle Poland approximation of example 5.

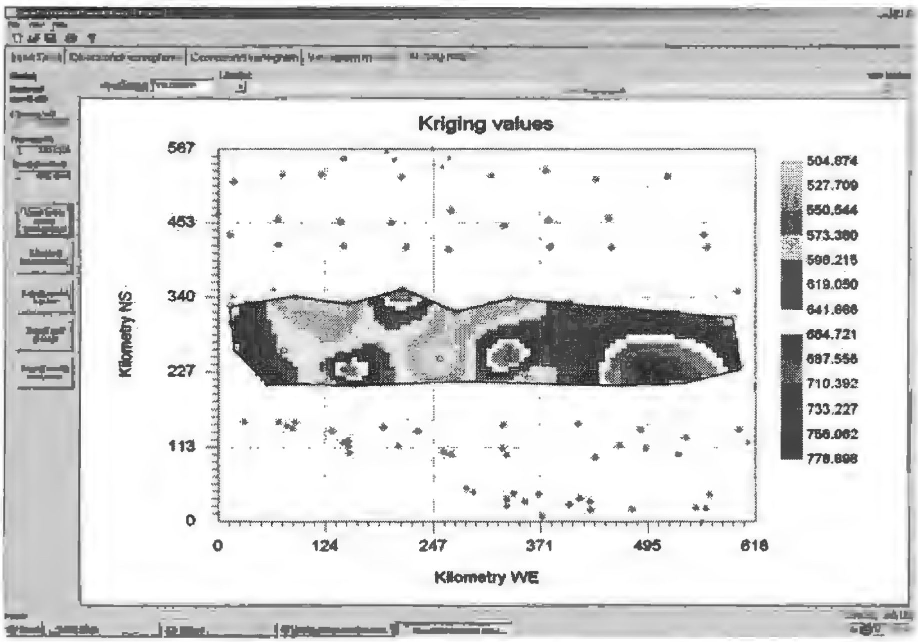

Figure 20. Contour lines for model 3 by Midlle Poland approximation of example 5.

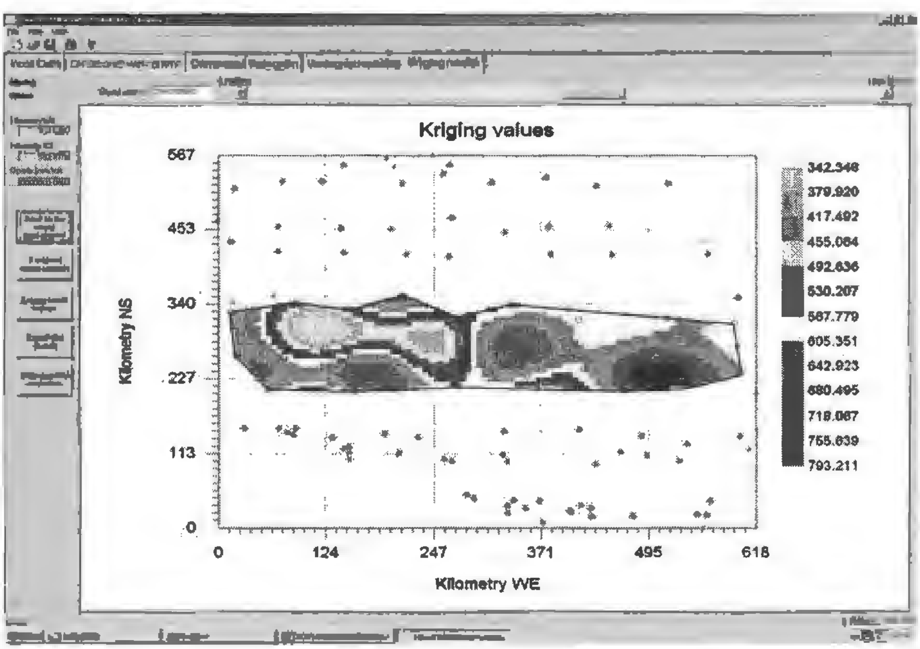

Figure 21. Contour lines for model 4 by Midlle Poland approximation of example 5.

There is also very helpful to assess the results of approximation qualitatively and not only quantitatively by comparison of the contour lines maps drawn for different approximation runs for qualitative evaluation is sometimes more reliable than this quantitative one.

The above observation has more general meaning and concerns not only the rain falls approximation and even not only the kriging calculation but it counts in the main for all cases of mathematical modeling and model identification of nonlinear systems. While modeling the kriging variograms mainly the nonlinear analytical functions are used and to identify their coefficients the Marquardt method of nonlinear gradient optimization is applied. This makes the modeling task really complicated and requires a big carefulness and also experience of the program operator by the evaluation of the obtained results.

5. Conclusions

After the test computation of the new developed KRIPOS program using the real rain falls data have been made some general conclusions concerning the kriging algorithms and also their applying to approximate environmental measurements could be done. They are as follows:

- There is reasonable to develop own programs for solving complex environmental tasks even if there are some ready tools on the market designed for the same problems.
- The own developed programs make possible to apply and test own numerical methods and algorithms what is not able with outside completed products.
- Kriging approximation is commonly useful to estimate the rain fall values.
- Kriging approximation is not quite right at the areas where the variables investigated have their values very differentiated and extreme.
- By very differentiated variable values there is useful to divide the whole investigated area into zones in which the variable values are more homogeneous.
- In the case of the rain falls approximation:
  - the omnidirectional kriging is more exact than the directional one,
  - the 3D kriging is more exact than the two dimensional one,

- the point coordinates given in kilometers are more suitable for calculations than these ones given in degrees.

- By the choice of the best variogram model for the subsequent variable approximation the residuum value is mostly a useful criterion but sometimes it is not.
- By the choice of the best approximation run there is always recommended to make the calculations using several variogram models and to make also a qualitative evaluation of results by comparing the diagrams drawn.

A more detailed description of the theory of kriging algorithms is presented in the literature given below.

References

Bogdan L., Studzinski J. (2006) *Anwendung der Krigingsapproximation zur Anfertigung der Regenfallkarten für Polen.* In: GI-Edition, Lectures Notes in Informatics (K.O. Wenkel, P. Wagner, M. Morgenstern, K. Luzi, P. Eisermannn, Hrsg.), Vol. P-78, Köllen Druck+Verlag, Bonn, 33-36.

Bogdan L., Studzinski J. (2006) *Entwicklung von Algorithmen zur Krigingsapproximation von Umweltdaten.* In: Simulation in Umwelt- und Geowissenschaften (J. Wittmann, Mike Müller, Hrsg.), Berichte aus der Umweltinformatik, Shaker Verlag, Aachen, 161-170.

Bogdan L., Studzinski J. (2006) *Mathematische Modellierung und Kriging-Algorithmen zur Approximation monatlicher und järlicher Regenereignisse.* In: Modellierung und Simulation von Ökosystemen (A. Gnauck, Hrsg.), Berichte aus der Umweltinformatik, Shaker Verlag, Aachen, 12-21.

Jan Studzinski, Olgierd Hryniewicz (Editors)

ECO – INFO AND SYSTEMS RESEARCH

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 applications of informatics in environmental engineering and environment protection at the Systems Research Institute of Polish Academy of Sciences in Warsaw (IBS PAN) and at the German Institute for Landscape System Analysis in Müncheberg (ZALF). The papers were presented in the form of extended summaries during a special workshop organized by IBS PAN in Szczecin in September 2006 together with the conference BOS'2006 dedicated to the applications of systems research in science, technology and economy and organized jointly by IBS PAN, University of Szczecin, and the Polish Society of Operational and Systems Research. They deal with mathematical modeling, approximation and visualization of environmental variables and with development of computer aided decision making systems in the area of environmental informatics.

ISBN 83-894-7509-X
9788389475091
ISSN 0208-8029