Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 215-232, 1998

Neural Network Residual Kriging Application for Climatic Data

V. Demyanov
Nuclear Safety Institute (IBRAE), B. Tulskaya 52, Moscow, 113191 Russia

M. Kanevsky
Nuclear Safety Institute (IBRAE), B. Tulskaya 52, Moscow, 113191 Russia
m_kanevski@ibrae.ac.ru

S. Chernov
Nuclear Safety Institute (IBRAE), B. Tulskaya 52, Moscow, 113191 Russia

E. Savelieva
Nuclear Safety Institute (IBRAE), B. Tulskaya 52, Moscow, 113191 Russia

V. Timonin
Nuclear Safety Institute (IBRAE), B. Tulskaya 52, Moscow, 113191 Russia

ABSTRACT **Direct Neural Network Residual Kriging (DNNRK) is a two step algorithm (Kanevsky *et al.* 1995). The first step includes estimating large scale structures by using artificial neural networks (ANN) with simple sum of squares error function. ANN, being universal approximators, model overall non-linear spatial pattern fairly well. ANN are model free estimators and depend only on their architecture and the data used for training. The second step is the analysis of residuals, when geostatistical methodology is applied to model local spatial correlation. Ordinary kriging of the stationary residuals provides accurate final estimates. Final estimates are produced as a sum of ANN estimates and ordinary kriging (OK) estimates of residuals. Another version of NNRK — Iterative NNRK (INNRK), is an iterated procedure when the covariance function of the obtained residuals are used to improve error function, by taking into account correlated residuals and to specify residuals followed by ANN modelling, etc. INNRK allows reducing bias in the covariance function of the residuals. However, INNRK is not the subject of this paper. The present work deals with the application of DNNRK model. NNRK models have proved their successful application for different environmental data (Kanevsky *et al.* 1995; Kanevsky *et al.* 1997a, 1997b, 1997c)**

KEYWORDS: **artificial neural networks (ANN), ordinary kriging, variography, residuals, non-stationarity, trend, clustering.**
ACKNOWLEDGMENTS The work was supported in part by INTAS 94-2361 and INTAS 96-1957

Contents

1. Introduction

Artificial neural networks (ANN) are frequently applied nowadays for analysing spatial data as well as for prediction mapping. Usually, when a spatial trend exists in the data, ANN is applied to residuals, remaining after a trend removal procedure. In this case one has to estimate the trend with, for instance, a polynomial, or another model.

The method suggested in the present paper is based on the idea of using ANN to estimate a non-linear drift and application of a geostatistical predictor (kriging) to the residuals. This alternative approach to ANN use brings high quality results described in the paper.

2. Neural Network Residual Kriging Algorithm

Data preparation The data are prepared and analysed by using descriptive statistics. Special attention is paid to outliers and data magnitude and variability.

All available data are split into training and validation data sets, according to the conditions of the SIC'97 project. The data set is used for training the ANN. The validation set is used only as an independent ("additional sampling") data set for testing.

The spatial structure of the data is explored using variography along with trend analysis. Usually, the description of spatial continuity using spatial correlation functions helps in understanding the phenomenon under study, its anisotropic structure, variability at different scales, possible relations between 'stochastic' and 'deterministic' parts, etc. (Isaaks and Srivastava 1989). ANN, being a data-driven approach, depends highly on quality and quantity of data. An important question is: how representative are the data in

respect to clustering of monitoring networks, described by spatial and dimensional (fractal) resolutions. Some aspects of these problems are discussed in short below.

Designing network architecture A workhorse of ANN — feedforward neural network (FFNN) — multilayer perceptron is used. FFNN consists of an input layer, an output layer and several hidden layers (Figure 1). No more than two hidden layers were never chosen. The number of neurons in the hidden layers can vary. Choosing the appropriate number of hidden neurons is very important. Using too few will starve the network of the resources required to solve the problem. Using too many will increase the training time and may cause a problem called overtraining or overfitting (network will have too much information processing capabilities and will learn insignificant aspects of the training set, which are irrelevant to the general population). In this case it was important to investigate the behaviour of the predictions and residuals with a variable number of hidden layers and neurons. According to the NNRK method, we used as few hidden neurons as possible to catch the large-scale structure only. Spatial correlation on the smaller scale remaining in the residuals are modelled within a geostatistical framework.

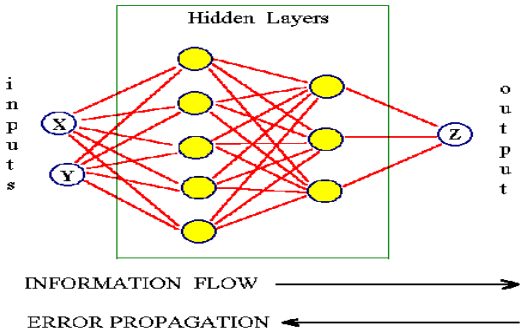

Figure 1 Four layer feedforward neural network (2-5-3-1) with two hidden layers

Training of the network The training data set was used for supervised learning. A classical vanilla backpropagation algorithm was applied with a few essential modifications: initial weights are selected with the help of genetic optimisation algorithms; conjugate gradients are used for the efficient local minimum search of error function; simulated annealing is used in order to escape from local minima.

Evaluating performance of the network Different tools can be used for the evaluation, like cross-validation accuracy test. The so called accuracy test is ANN estimation function values in the points used for training. The accuracy test provides the residuals at the training points, which are used in further geostatistical analysis. Neural Network Residual Kriging (NNRK) is a two step algorithm (Kanevsky *et al.* 1995). The first step includes estimating global (trend) structure by using ANN. The second step is residuals analysis, where geostatistical methodology is applied to model local spatial correlation. The final estimate is produced as a sum of ANN estimates of the samples and ordinary kriging (OK) estimates of residuals. Scatter plots – estimated versus real data – describe how well the FFNN capture the correlation between locations and contamination. Results of the accuracy tests for the different networks are presented as scatter plots.

ANN testing Testing is a process of estimating the FFNN ability to generalise, that delivers a correct response to the inputs ANN has never been exposed to before. At this phase a validation data set was used.

Operation phase: prediction mapping Co-ordinates on a regular grid are given as the input to the ANN, which produce gridded predictions of the output.

Analysis of the residuals, structural analysis and modelling, kriging Residuals, obtained after the learning phase, were analysed with the help of exploratory variography. Two kinds of residuals behaviour are possible: 1) network was able to learn the data, and the residuals are not correlated (neural network regression model); 2) network was able to catch only large scale structure, and the residuals are spatially correlated. Usually, residuals, unlike the original data, display stationarity and well-behaved semivariograms. Neural networks feature robust behaviour, while finding large-scale structures, leaving local peculiarities for more sensitive tools, like a variogram. Much more powerful networks should be designed, trained, and validated, in order to learn small-scale variations.

Prediction mapping A developed variogram model was used for kriging prediction mapping. After kriging predicted residuals were added to the results of neural network predictions (NNRK – neural network residual kriging). Comparing ANN and NNRK predictions, it is seen that NNRK provides better treatment of the original samples, because OK procedure estimates the function exactly at the sampling points, unlike ANN, which gives close but not the exact values. This feature allows discovery of more contamination spots and local peculiarities of the contamination pattern.

Final validation Calculating final NNRK predictions at validation points and comparison with the true values.

3. Monitoring Network Analysis

Before analysing spatial distribution function, network analysis is monitored for better understanding of the spatial pattern and its cluster structure. The following networks are considered: training network (100 samples initially available for analysis), validation network (367 points to be estimated), all data network (467 sample points of from both networks). Network cluster structure is shown with the help of Delaunay triangulation (Figure 2). Monitoring network analysis is one of the most important parts of the analysis of spatially distributed data, because all currently known interpolation technologies are affected by the spatial distribution of sample points, and even the statistical characteristics of data (histograms, mean and others) are dependent on the presence of cluster structures. Several different approaches can be used for network analysis: spatial/geometrical, (e.g., a histogram of squares of areas of influence), statistical, (e.g., Morishita diagram), and analysis of the network (dimensional resolution – all kinds of fractal network analysis). A histogram of these areas of influence (also called Dirichlet cells, Voronoi polygons) can characterise the spatial resolution and homogeneity of the monitoring network (Figure 4). The presence of several peaks implies that there is the contagious structure of monitoring network. The presence of a long right tail means the presence of holes in the network. The presentation of the Voronoi map helps one to discover the regions with prevailing network disturbances and the values (Figure 3). The average distance between the samples characterises effective size of the smallest spatial structure to be discovered by the network (see Table1).

Table 1 Average distance between the samples.			
Monitoring networks	**Training Net** (100 samples)	**Validation Net** (367 samples)	**All data Net** (467 samples)
Average distance (meters)	19870.0	11010.0	9920.0

Statistical approach to network analysis is realised with the help of a Morishita diagram (Morishita 1959). To estimate the Morishita index (I_δ) the whole given area is divided into small rectangular cells of equal size (d). Then,

$$I_\delta = Q\frac{\sum_{i=1}^{Q} n_i(n_i-1)}{N(N-1)} \qquad (1)$$

where n_i *(1=1,2...Q)* is the number of points in the *i*-th cell.

A Morishita diagram measures the dependence between *Id* and the cell size (δ). It provides various information about the structure and the clustering size of the network. The diagram has three characteristic types of behaviour. If the distribution of points is uniform, the Morishita index *Id* increases to 1 with the growth of the cell size. If the distribution is contagious (with cluster structures) the Morishita index *Id* decreases with the growth of the cell size. The characteristic of the cluster size is the point on the Morishita diagram where its behaviour is changing. *Id* =1 for a random distribution of points and *Id* >>1 for self-similar fractal clustering.

The Morishita diagram in Figure 5 leads to the following conclusions. The training monitoring network (100 samples) has a cluster structure up to 30 km in size. Beyond this range clusters can be considered as separate points. The validation network (367 samples) and the network for all data (467 samples) are similar in their cluster characteristics. The effective cluster size is about 40 km.

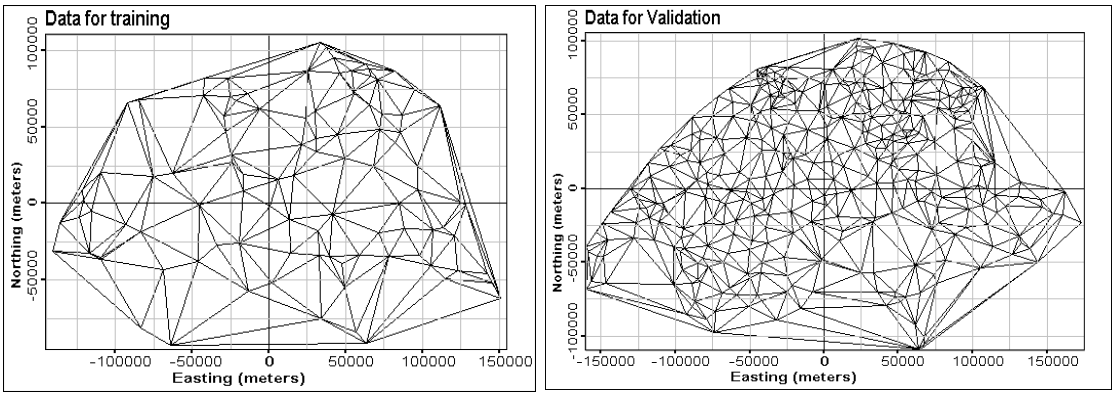

Figure 2 Triangulation for training (left) and validation (right) monitoring networks

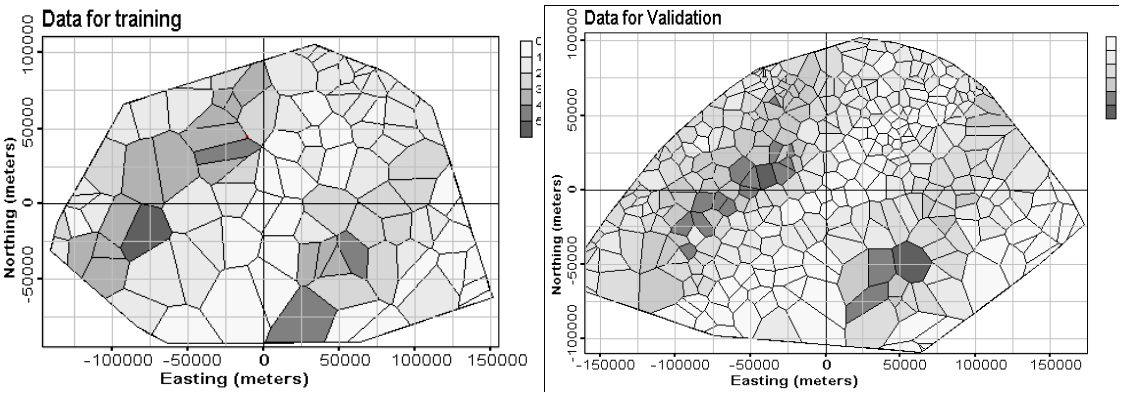

Figure 3 Voronoi polygons for training (left) and validation (right) monitoring networks

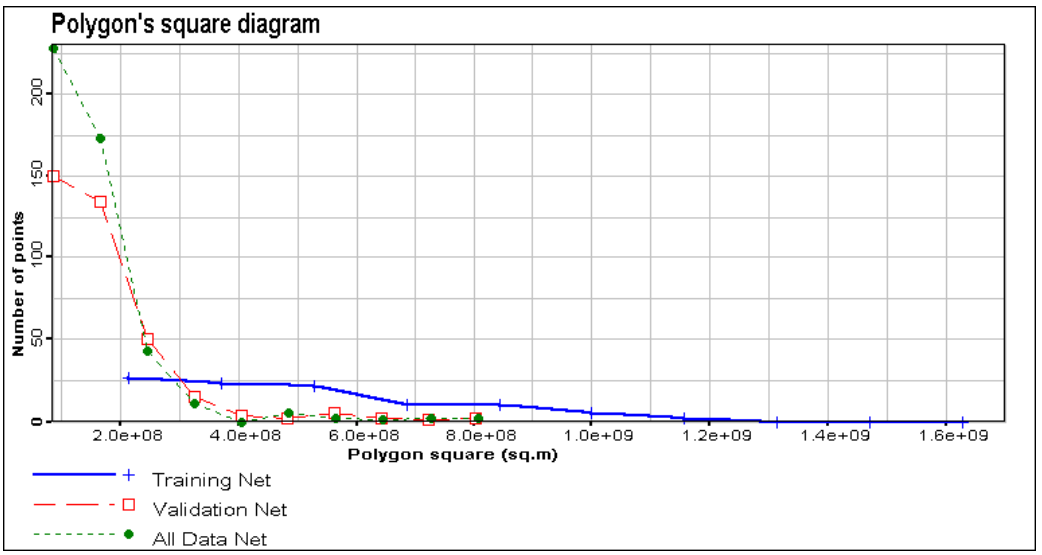

Figure 4 Diagram of Voronoi polygon squares for training, validation and both together monitoring networks

If a set of sampling points is self-similar (i.e., a magnified version of any small part of it is statistically similar to the whole set), it can be characterised by its fractal dimension. It is well known, that in order to detect the phenomena by a monitoring network, a network must have both sufficient spatial resolution and sufficient dimensional resolution. Whenever the fractal dimension *D(fract)* is less than the Euclidean dimension *D(Eucl)*, sparsely distributed phenomena with a dimension less than *D(Euclid) -D(fact)* cannot be observed, and the difficulties in interpolating low-dimensional sparse data onto two-dimensional regular grids is assumed (Korvin *et al.* 1990).

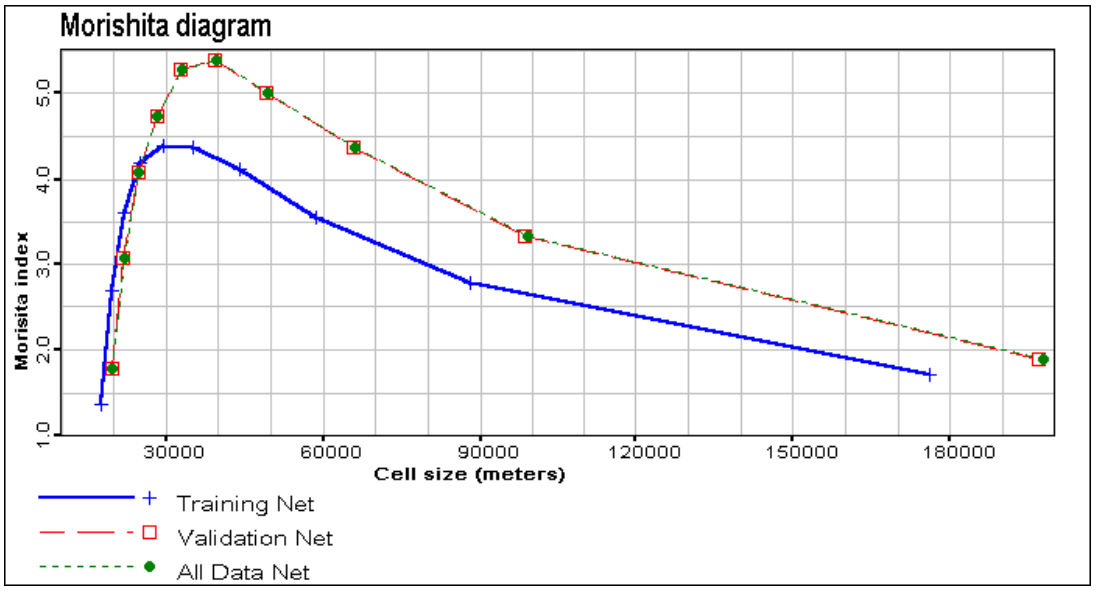

Figure 5 Morisita index diagram for training, validation and both together monitoring networks

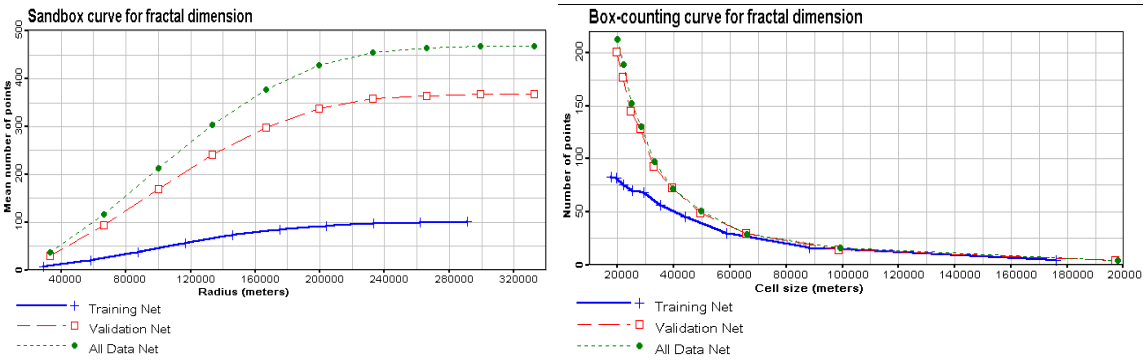

Figure 6 Fractal dimension curves for training, validation and both together monitoring networks: sand-box counting method (left), box counting method (right)

4. Structural analysis of the raw data

The initial data set (100 samples) was explored for spatial correlation structures. The main objective of this study is to decide which model of spatial continuity to use: stationary or non-stationary. Drift analysis was performed along with conventional variography. The variogram rose (Figure 7) shows periodical structures with different ranges in all directions. The number of pairs in directional variograms ranges from 50 to 200 depending on the lag and direction. Along with periodicity, geometric anisotropy can be seen in SW-NE direction on a smaller scale (30-60 km). The drift rose in Figure 7 demonstrates both positive and negative drift depending on the direction. On a smaller scale (40 km), the drift module is lower and ranges around zero. The presence of drift and complex anisotropic correlation structures requires complicated non-stationary prediction models or the use of trend-removing procedures before applying stationary models. The original rose display for variograms and drift was developed with the help of VarRose software (http://www.ibrae.ac.ru/~mkanev/).

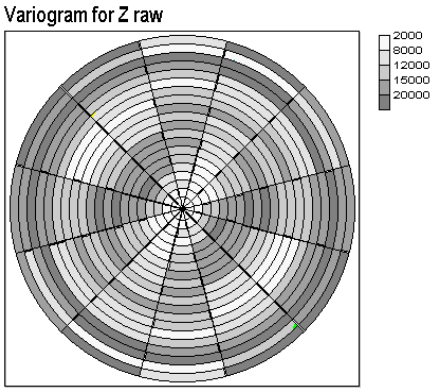

Figure 7a Variogram for Z raw

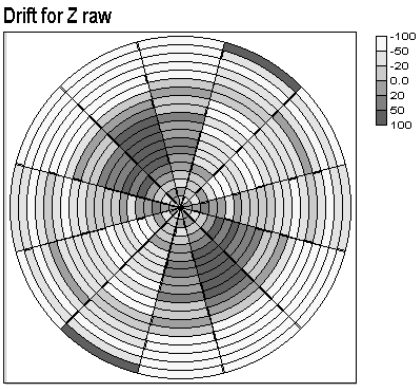

Figure 7b Drift for Z raw

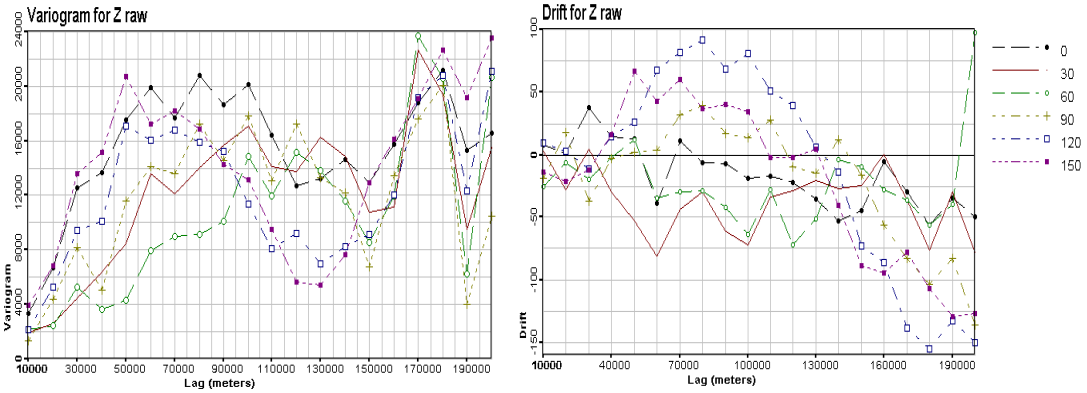

Figure 8 Directional raw variograms (left) and directional drift (right)

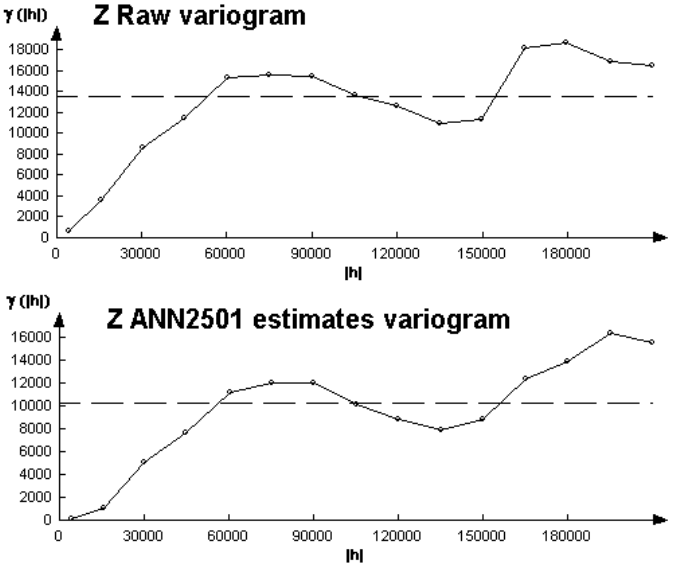

Figure 9 Variograms for raw data (top) and for ANN estimates (bottom)

5. NNRK Prediction of Climatic Data

For the climatic data on Switzerland NNRK was applied in order to model multi-scale spatial structures.

5.1 ANN Estimation

ANN was chosen to estimate large-scale correlation structures, including periodical structures discovered by the variogram (see Figure 9 top). One hundred samples were used for training and testing the network. ANN of different architecture were tested. [2-5-0-1] ANN showed the best performance, with 1 hidden layer containing 5 neurons. An accuracy test was used to qualify the ANN training (see Figure 10, Table 2). ANN [2-5-0-1] show good overall performance on the large scale (Figure 14 left), and fair correlation (0.851) with the measurements on the validation set (367 samples). ANN estimates perfectly reflect the large-scale correlation structure on 80000 m and periodical effects (Figure 9).

Table 2 Accuracy test results: correlation between ANN estimates and measurements

ANN estimates	**[2-5-0-1]**	**[2-10-0-1]**
Correlation coefficient	0.866	0.860

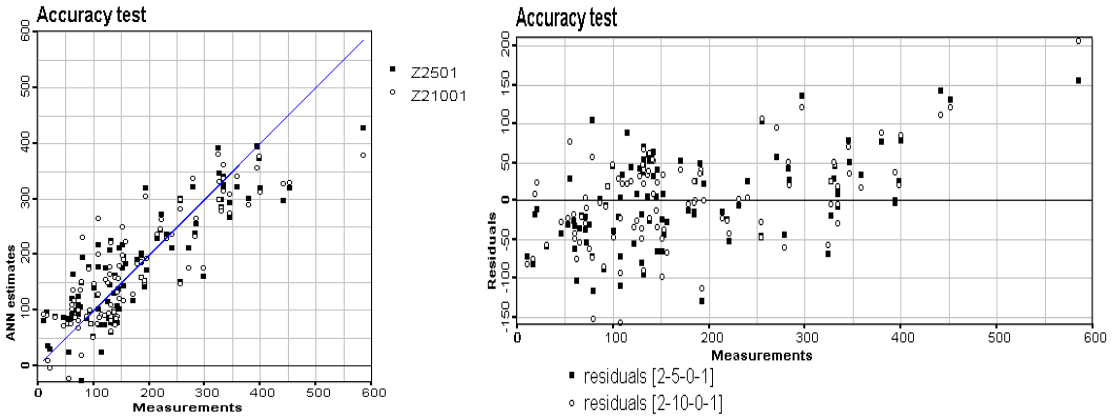

Figure 10 Accuracy test: Scatter plot of ANN estimates for [2-5-0-1] and [2-10-0-1] (left), scatter plot of residuals (right)

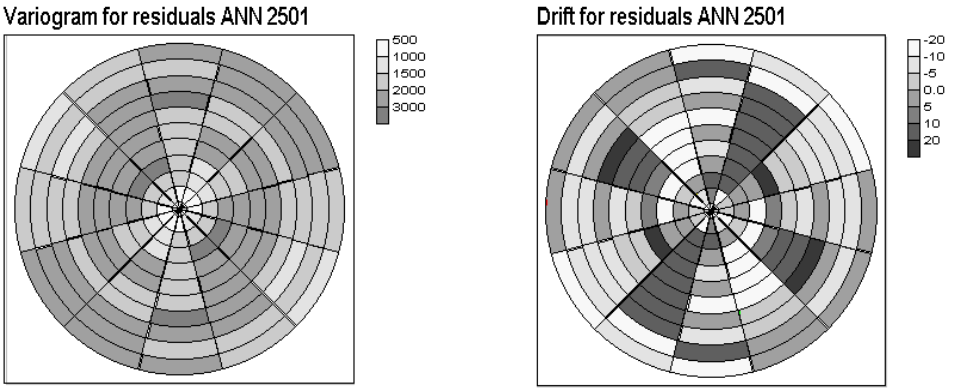

Figure 11 Variogram rose (left) and drift rose (right) for residuals from ANN [2-5-0-1]

5.2 Estimates of Residuals

Now, let us study ANN *residuals*=[*measurements - ANN estimates*]. There appears to be significant correlation with the measurements – 0.494 (see Figure 10 right). This is a sign of remaining correlation, which was not captured by ANN. ANN has removed most of the periodicity and large scale trend, which is shown in the variogram and drift roses (Figure 11). Drift fluctuates around zero and does not appear any straight tendency in most directions (Figure 12). The spatial structure of residuals is on a small-scale (30000 m), features stationary behaviour and is easily modelled (Figure 13). There still is some geometrical anisotropy, which can be modelled within stationary variogram model.

Estimation of residuals is made with ordinary kriging (OK). Prediction of residuals is shown in Figure 14.

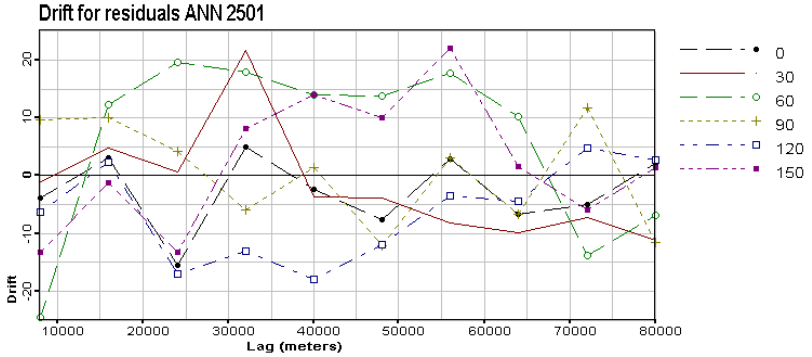

Figure 12 Drift in 6 directions for residuals from ANN [2-5-0-1]

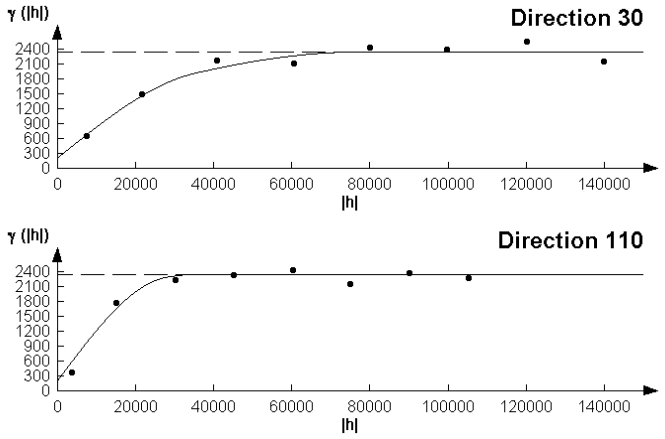

Figure 13 Anisotropic variogram model for residuals from ANN [2-5-0-1]

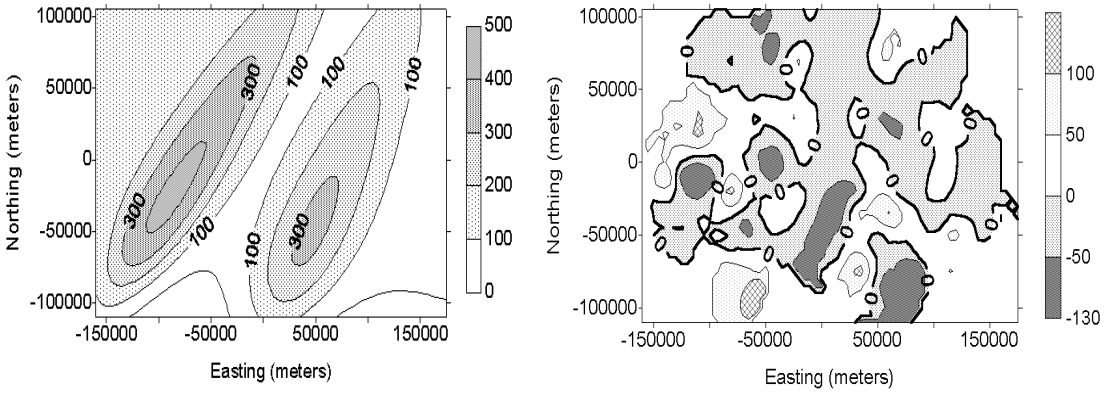

Figure 14 Maps of ANN [2-5-0-1] estimates (left) and OK estimates of residuals (right)

5.3 NNRK Final Predictions

Final NNRK predictions are obtained by summarizing ANN and OK estimates in Figure 14. Results of girded predictions are presented in Figure 15 with the 10 maximum and 10 minimum sample values from the whole data set (467 samples). The ten minimum and ten maximum samples are situated in the corresponding areas of maximum values and minimum values and show good fit to the isolines of the estimates. As far as NNRK exactly honours the initial data, 10 maximum and 10 minimum measurement values from the whole data set (467 samples) are fairly well reproduced. Six out of ten minimum samples fall into the 10 minimum NNRK estimates, and four out of ten maximum samples fall into the 10 maximum NNRK estimates. Underestimating some peculiarities in the hot spots, which were not reproduced by the training data, causes a lower fraction for maximum values. NNRK significantly improved pure ANN estimates; correlation with the validation measurements has increased as well (see Table 3).

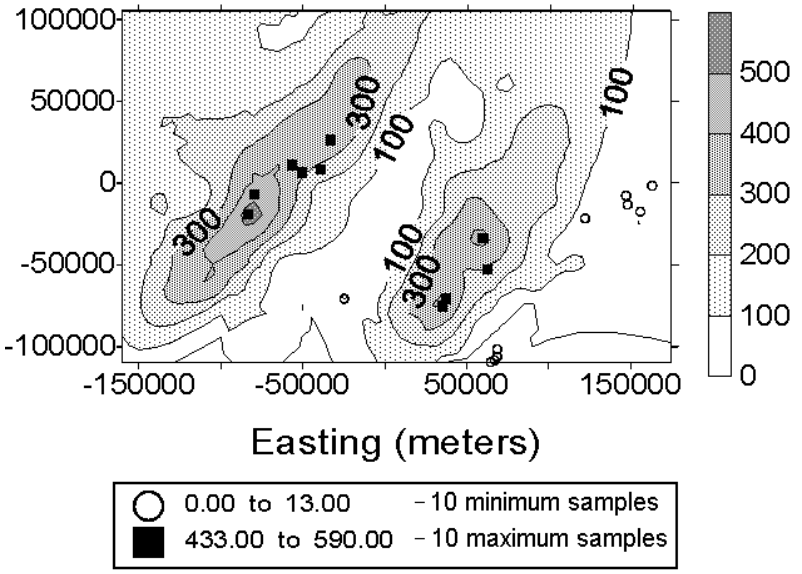

Figure 15 NNRK predictions with 10 maximum and 10 minimum measurement values from the whole data set

Table 3 Validation results: correlation between true measurements and NNRK estimates, ANN estimates		
Estimates	**NNRK**	**ANN [2-5-0-1]**
Correlation coefficient	0.866	0.851

6. Analysis of the Results

6.1 Validation Estimates

Quality of NNRK predictions were analysed with the help of the validation data set (367 samples), which was not used for training ANN or OK estimates. NNRK estimates appear to have good correlation with the validation samples (Figure 16 left). Correlation between validation errors (*estimated-measured*) and the measurements is -0.356, which means some underestimation especially for the high values (see Figure 16 right). The univariate distribution of NNRK predictions in the validation points perfectly reproduces the distribution of the measurements (see Table 4 and a histogram of NNRK estimates compared with the one for validation measurements in Figure 17). NNRK estimates and the validation measurements can be visually compared with the help of Voronoi polygons (Dirichlet cells), which represent areas of influence around the samples (see Figure 18). Along with good overall performance, there appeared be under estimation of the eastern hot spot. This may require deeper structural analysis of residuals or application of indicator methods for different cut-offs.

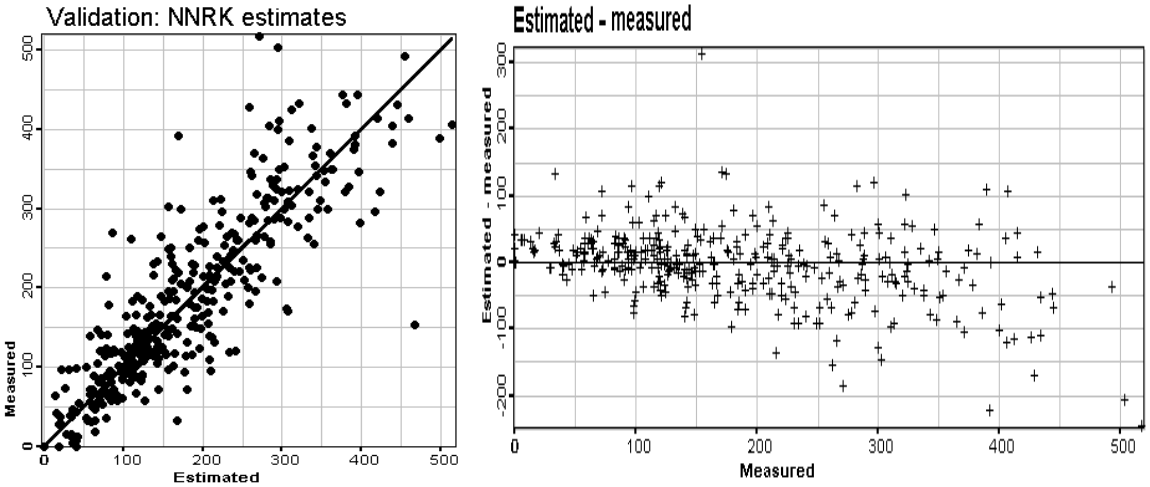

Figure 16 Validation estimates / measurements scatter plot (left), absolute error / measurements scatter plot (right)

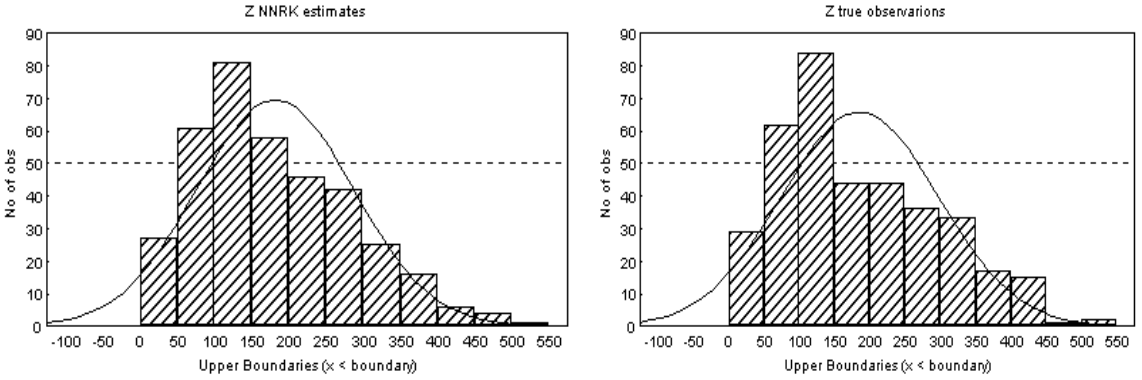

Figure 17 Histograms of NNRK estimates (left) and validation measurements (right)

Table 4 Validation: basic statistics for measured (true) and estimated (NNRK) values

Statistics	Minimum	Maximum	Mean	Median	Standard deviation
True values	0.0	517.0	185.36	162.0	111.167
NNRK Estimated values	0.0	514.1	181.45	160.5	105.220
ANN 2501 estimates	-3.53	425.9	186.28	178.4	101.3

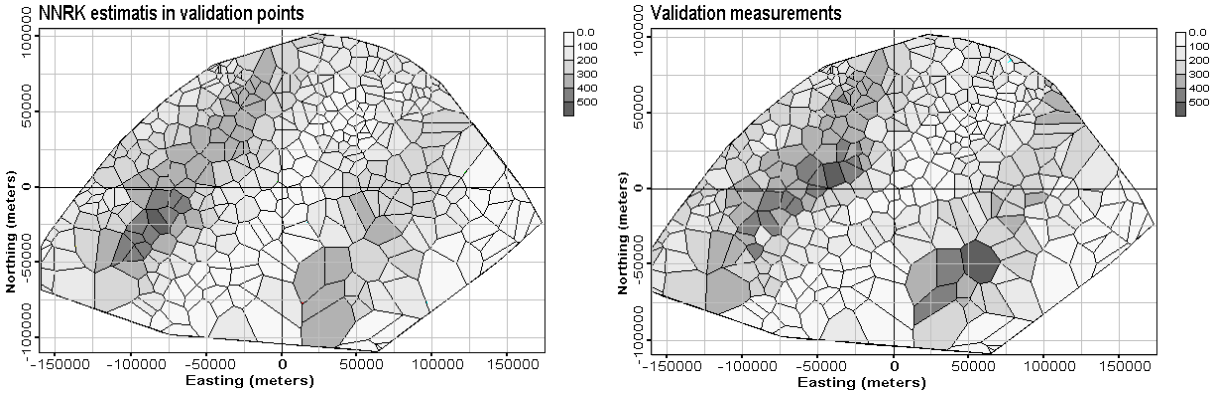

Figure 18 Voronoi polygons for validation results: NNRK estimates (left) and measurements (right)

Table 5 Bias analysis of validation errors of *NNRK estimates - measurements*

Statistics	Error *(Z*-Z)*	Absolute error *|Z*-Z|*	Relative error *(Z*-Z)/Z*
Mean	-3.9	39.38	0.099
Standard deviation	56.28	40.35	0.659
Minimum	-245.0	0	-1.0
Lower Quartile 25%	-30.86	12.47	-0.159
Median	0.0	27.70	0.0
Upper Quartile 75%	24.21	50.84	0.179
Maximum	313.24	313.24	6.697
Skewness	-0.154	2.40	5.55
Kurtosis	4.253	8.83	42.622

Table 6 Validation Error statistics of NNRK estimates (*estimated-measured*)	
Statistics	**NNRK estimates**
Root mean square error (RMSE)	56.342
Mean absolute error	39.38
Relative mean error	0.099

6.2 Validation Error

Validation errors (*estimated-measured*) characterize the quality of predictions. Ideally error mean and median should be 0.0, and distribution of errors should be symmetric around 0.0. Statistics for absolute errors *(Z*-Z)* and *(Z*-Z)/Z*, where *Z** is NNRK estimate and *Z* is true value, are presented in Tables 5 and 6. Absolute error distribution has got a slight negative bias of mean (< 1%) and no bias of median, which is considered acceptable. Moreover, absolute error distribution is very slightly skewed (-0.154). Along with almost symmetric quartiles, this fact supports the assumption of symmetry (see Figures 19 and 20). Relative error also has a zero median and symmetric quartiles. However, it is rather skewed because of positive outliers. These outliers also affect the mean, which is about 10% biased from 0.0. Four positive outliers can be distinguished in the Voronoi polygon map in Figure 21 – three of them are situated at the boundary.

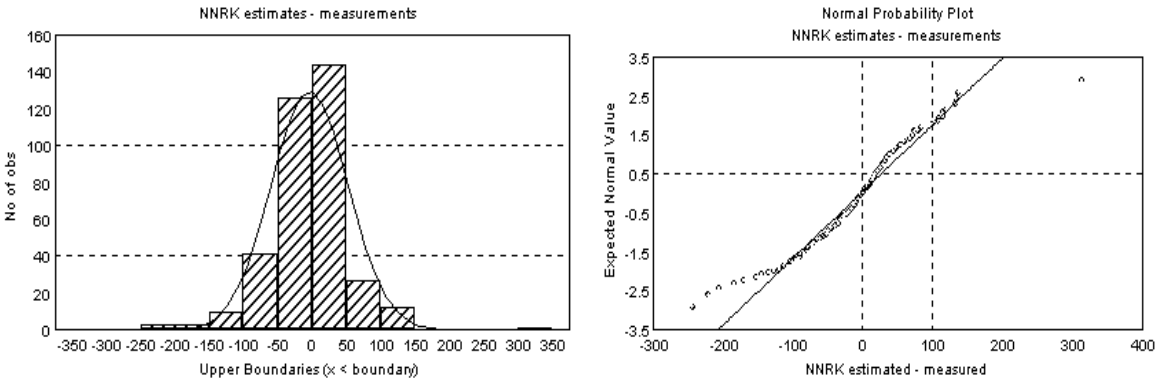

Figure 19 Validation: NNRK estimates - measurements, histogram (left) and normal probability plot (right)

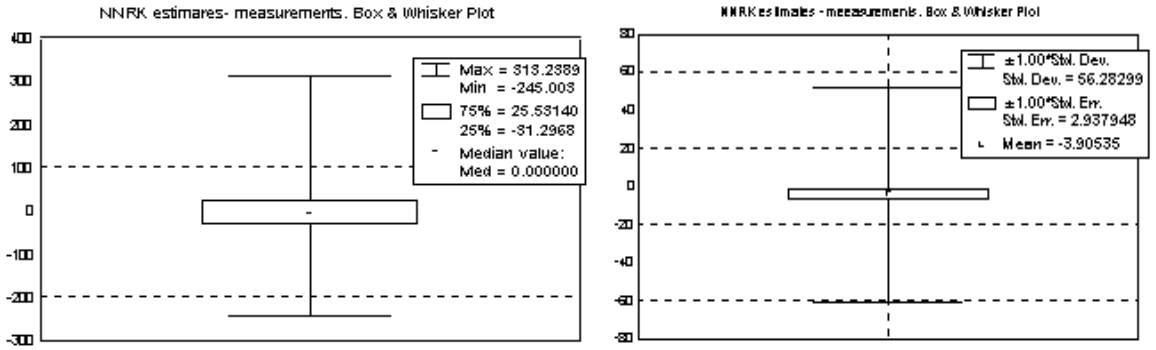

Figure 20 Box and whisker plots of mean and median for absolute errors: estimated-measured

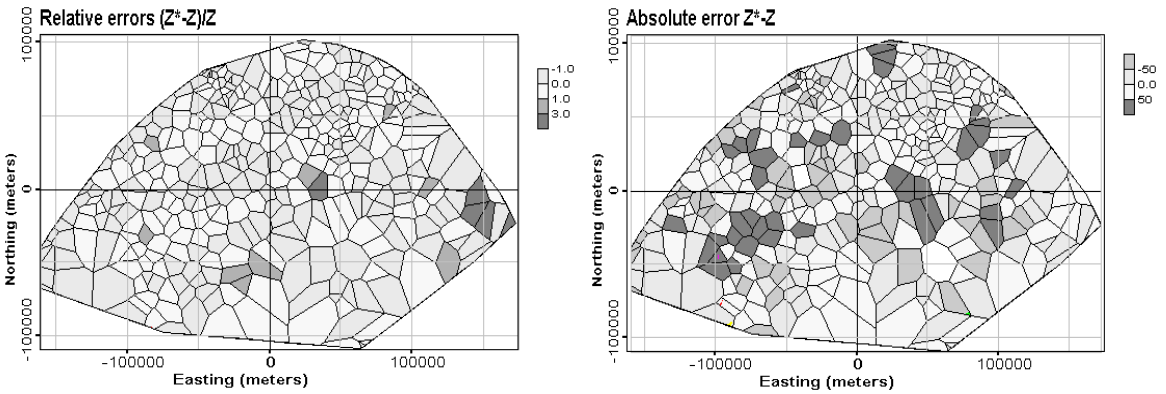

Figure 21 Voronoi polygons for validation results: relative error *(Z*-Z)/Z* (left), absolute error *Z*-Z* (right)

6.3 Accuracy of the Estimates

Accuracy of NNRK estimates is given by OK estimation error of residuals. The spatial distribution of the estimation error is shown in Figure 22. It is unconditional and represents the density of the monitoring network. Univariate distribution of the estimation error is quite close to normal (see Figure 23) and has no dependence on the validation measurements (Figure 24).

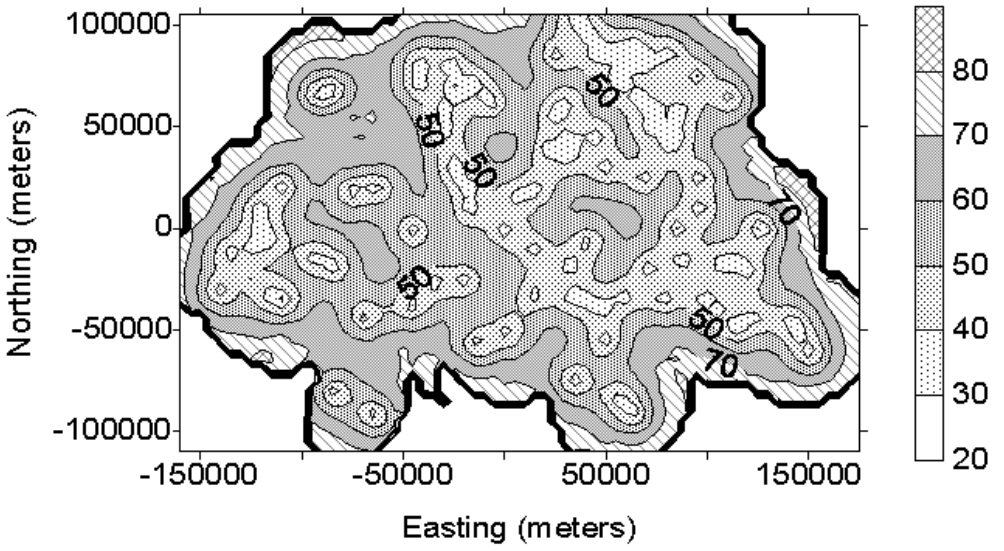

Figure 22 Map of NNRK estimation error

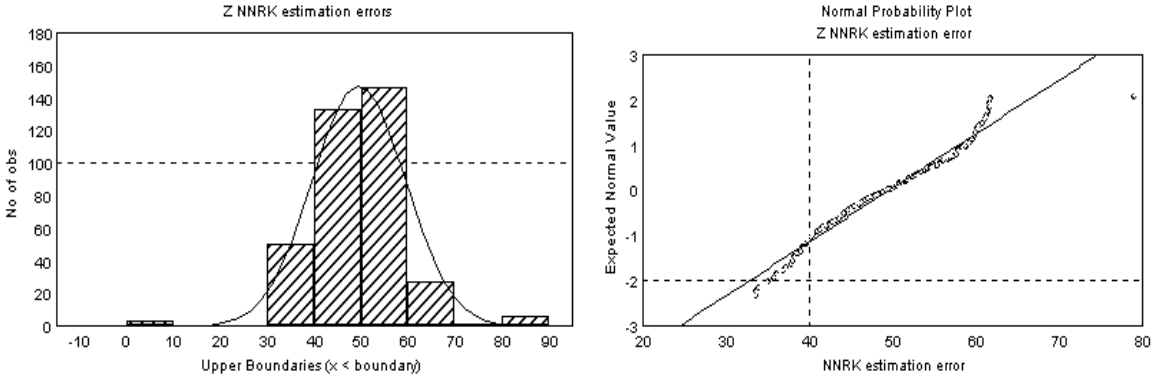

Figure 23 NNRK estimation errors: histogram (left) and normal probability plot (right)

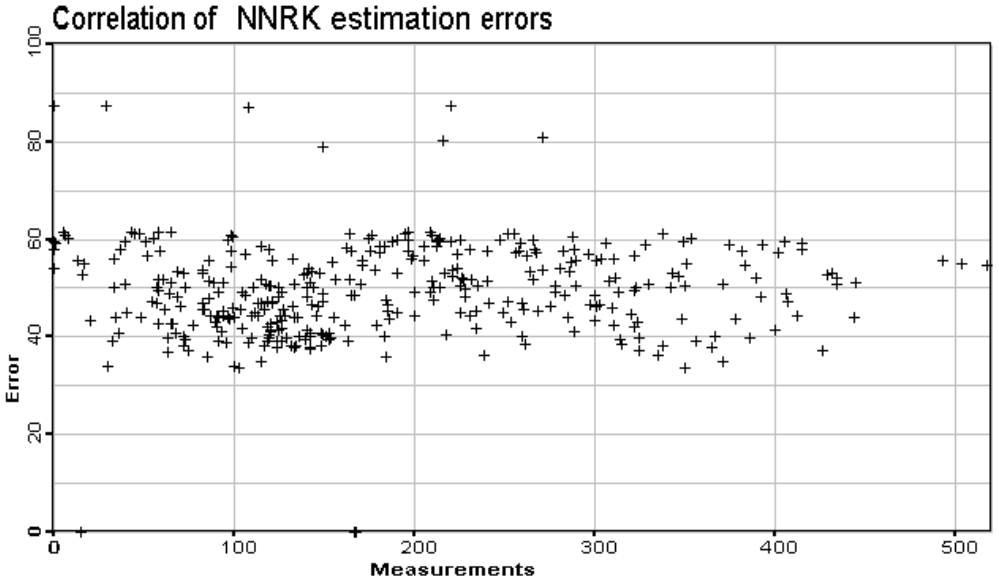

Figure 24 Validation: NNRK estimation error / measurements scatter plot

7. Software Tools

A wide collection of software tools (a part of them is known as Geostat Office - http://www.ibrae.ac.ru/~mkanev) were used to produce and to present the results. Data handling, preparation and visualisation was performed by using GeoPlot/3Plot software (Kanevsky *et al.* 1997b) and a data base management system. ANN were trained and applied using programs from (Masters 1993). Variography were performed with the help of VarRose and VARIOWIN (Pannatier 1996). Ordinary kriging predictions were carried out with the help of WinGSLIB software, which is a Windows extension of GSLIB (Deutsch and Jounel 1996) and allows calculation of estimates on arbitrary grids. Statistical description was made with the help of Statistica package and 3Plot. Black and

white contour maps were produced by SURFER software, while 3Plot produced the same maps in colour and Voronoi polygon maps. For more details about GeoPlot/3Plot, VarRose, NetMan and WinGSLIB (see http://www.ibrae.ac.ru/~mkanev/).

8. Conclusions

NNRK method, applied to climatic data, showed good quality performance in the following aspects:

- Univariate distribution of validation set (minimum, maximum, median, mean and standard deviation);
- Spatial distribution of the validation set including western hot spot;
- Modelling and spatial estimation of large scale structure, including periodicity;
- Modelling of small scale effects given by ANN residuals;
- Acceptable quality of absolute validation errors – minimum bias of mean, no bias of median, almost symmetric distribution.

There appear to be some problems with underestimating the eastern hot spot and positive bias of relative error, which is caused by three points at the boundary. NNRK technology, being very flexible, can be applied to different kinds of data. The method is exact (i.e., if there are no measurement errors - predictions at the sampled points are equal to measurements). Even simple ANN can detect and model highly non-linear trends in the region under study in the case of noisy data. Instead of using geostatistical analysis and modelling of the residuals, incremental methodology, based on ANN only, can be used (at each iteration hidden neurons concentrated on the residual information can be added) (Chentouf *et al.* 1997). The method can be adapted both for emergency situations and long term management.

References

Chentouf R., C. Jutten, M. Maignan and M. Kanevsky (1997). Incremental neural networks for function approximation. *Nuclear Instruments and Methods in Physics Research*. V. A 389, pp. 268-270.

Deutsch C. and A. Jounel (1997) *GSLIB. Geostatistical Software Library*. Oxford: Oxford University Press.

Isaaks H. and R. M. Srivastava (1989) *Introduction to Applied Geostatistics.* Oxford University Press.

Kanevsky M., R. Arutyunyan, L. Bolshov, S. Chernov, V. Demyanov, I. Linge, N. Koptelova, E. Savelieva, T. Haas, and M. Maignan (1997a) Chernobyl Fallouts: Review of Advanced Spatial Data Analysis, In: *geoENV I — Geostatistics for Environmental Applications*, edited by A. Soares, J. Gomez-Hernandes, R. Froidvaux, Kluwer Academic Publishers, pp. 389-400.

Kanevsky M., R. Arutyunyan, L. Bolshov, V. Demyanov and M. Maignan (1995) Artificial neural networks and spatial estimations of Chernobyl fallout. *Geoinformatics*, 7(1-2); 5-11.

Kanevsky M., S. Chernov, and V. Demyanov (1997b) 3Plot Software – Advanced Spatial Data Plot. *Preprint IBRAE* 97-02.

Kanevsky M., V. Demyanov and M. Maignan (1997c) Spatial estimations and simulations of environmental data by using geostatistics and artificial neural networks. In: *IAMG'97 Proceedings of The Third Annual Conference of the International Association for Mathematical Geology*, edited by V. Pawlowsky Glan. Barcelona, Spain, CIMNE, ISBN 84-87867-97-9, vol. 2, p.527.
Kanevsky M., M. Maignan, V. Demyanov and M. F. Maignan (1997d) Environmental decision-oriented mapping with algorithms imitating nature. In: *IAMG'97 Proceedings of The Third Annual Conference of the International Association for Mathematical Geology*, edited by V. Pawlowsky Glan. Barcelona, Spain, CIMNE, vol. 2, p. 520.
Korvin G., D.M. Boyd and R. O'Dowd (1990) Fractal characterisation of the South Australian gravity station network. *Geophys. J. Int.* 100; 535-539.
Masters T. (1993) *Practical Neural Network Recipes in C++*, New York: Academic Press.
Morishita M. (1959) Measuring of Dispersion of Individuals and Analysis of the Distribution Patterns. *Mem. Fac. Sci. Kyushu Univ.*, Ser. E (Biol.), 2, 215-235.
Pannatier Y. (1996) *VARIOWIN Software for Spatial Data Analysis in 2D.* Berlin: Springer-Verlag.

JGIDA vol. 2, no. 2
JGIDA Home