Decision-Oriented Environmental Mapping with Radial Basis Function Neural Networks

V. Demyanov (1), N. Gilardi (2), M. Kanevski (2), M. Maignan (3), V. Polishchuk (1)

(1) Institute of Nuclear Safety (IBRAE), B. Tulskaya 52, 113191 Moscow, Russia vasia@ibrae.ac.ru
(2) IDIAP Research Institute, Martigny, Switzerland; gilardi@idiap.ch
(3) Institute of Mineralogy and Petrology, University of Lausanne, Switzerland; michel.maignan@imp.unil.ch

Abstract. The work deals with development and application of Radial basis functions neural networks (RBFNN) for spatial predictions. Geostatistical tools for spatial correlation analysis (variography) are used to qualify and quantify the estimation results. Geostatistical analysis is performed on the residuals obtained at the training and test sample locations. Variogram of residuals explores spatial correlation remaining after RBF modelling and is an independent criteria of the results quality. Networks with different number of hidden units and different RBF functions were considered. Extended RBF networks allow to obtain not only the estimate itself by also the estimate variance, which qualifies the prediction map. Two real case studies were considered: radioactive soil contamination after the Chernobyl fallout and heavy metal sediments in Lake Geneva. Decision-oriented prediction maps accompanied by maps of estimation errors and "thick contours" are presented as the outputs for decision making.

Introduction

The problem of decision-oriented mapping is faced very often by decision-makers, while dealing with spatially distributed data. Environmental decision-oriented mapping (DOM) includes comprehensive data analysis, monitoring networks design and redesign, spatial predictions/estimations along with mapping of estimates variances, probabilistic/risk mapping, conditional stochastic simulations. In the present work application of the RBFNN for the spatial predictions and mapping of estimate variances as a first step in the DOM process is considered. In contrast to other papers, geostatistics is widely used from the beginning of the analysis (exploratory data analysis, declustering, spatial continuity description – variography) until the interpretation of the results in the present research.

Traditionally there exists a comprehensive spatial data analysis methodology including a wide range of geostatistical models and tools to assess this task and provide the full range description of the spatial phenomenon [1,2]. The most commonly occurred problems are multivariate and multiscale structure of the spatial data, high variability of the data, presence of noise and outliers. Presence of spatial trends limits application of stationary geostatistical models like ordinary kriging. To overcome this problem geostatistics offers several type of kriging for non-stationary cases: kriging with a trend (universal kriging); moving window kriging, when the region is split into smaller areas, within which local stationarity can be assumed; intrinsic random function of the order *k* kriging, which employs correlation function of the order higher than two.

Artificial neural networks (ANN) offers an alternative way of dealing with spatial information when no stationary model is assumed. ANN are adaptive, non-linear, model free estimators, which are robust in case of noisy data and extreme values. ANN have got a remarkable feature of learning from data and ability to reproduce data pattern at unsampled locations according to the learned experience. ANN do not depend on any built in or assumed model, but only on their inner structure (architecture, learning algorithm) and the training data itself, which is used for learning. However, it should be noted that ANN have a solid mathematical foundation, by means of which ANN can be interpreted as a universal approximators [3,4].

Feedforward neural network (FFNN) application to the problem of spatial prediction has provided very promising results when the model was supplemented with geostatistical kriging estimation of the ANN residuals. The work [5] showed that multi-layer perceptron (MLP) is able to catch non-linear large scale spatial structures and the local correlation remaining in the residuals is fairly well modelled by stationary ordinary kriging model if it is needed. Such combined model has demonstrated its advantage upon pure ANN or pure geostatistical predictors. Another spatial estimation application of ANN dealt with General Regression Neural Networks (GRNN) [6,7]. Unlike MLP GRNN was able to model both large and small scale correlation. An extension of the GRNN for modelling prediction variance is given in [7].

The subject of the present work is to adapt a radial basis function (RBF) neural network for spatial predictions/estimations and use geostatistics to qualify the results and to improve the prediction if possible. RBF networks possess the property of the best universal approximation. The characteristic feature of RBF network is clear distinction between the contribution of the first layer and second layer parameters to network output. Thus, training procedure can be divided into two distinct stages. During the first stage relatively fast non-parametric methods for finding optimal centres positions can be used given the number of hidden units in the model. This stage makes use only of the monitoring network. When the number, type and centres positions of basis functions are fixed, the problem of finding optimal weights becomes linear and allows implementing simple linear algorithms for solving the equation for optimal weight vector.

One of the present case studies deals with radioactive soil contamination with caesium 137 (CS137) originated after the accident at Chernobyl Nuclear Power Plant [5]. Training set consists of 300 and test set of 365 samples of contamination level as a function of two location co-ordinates. The other case study deals with Ni contamination of Geneva Lake. 200-data set was randomly split into 150-points training set and 50-point test set.

Radial Basis Function Neural Networks

Radial basis function neural networks express the regression function in the form

$$Z=\Sigma_{j=1,m} \lambda_j h_j+\lambda_0+\varepsilon$$

where is a zero-mean noise, and h – radial basis functions.

Radial basis functions are a class of functions which key feature is that the distance from a centre determines their response.

$$h_i(x,c_i,r) = h(\| x - c \|/r)$$

where r is a smoothing parameter (bandwidth).

Centres positions, type of the radial functions and their number are parameters of the model. We must keep them unchanged if we want the problem of weight optimisation to be linear.

Most commonly used RBF is a Gaussian radial basis function, which in the case of scalar input, is

$$h(x,c,r)= exp(-(x-c)^2/r^2)$$

Gaussian provides an example of local RBF that gives significant response only in a neighbourhood of its centre. In contrast, global RBF response increases monotonically with distance from the centre. Thin Plate Spline (tps) RBF which, in the case if scalar input, is

$$h(x,r)=(x/r)^2 \, Ln(x/r)^2$$

RBF neural network has a well known architecture presented in Fig. 1. Extension of the vanilla RBFNN deals with including one more output neuron for the error bars estimates (see details below). Gaussian and thin plate spline RBFs are shown as well.

Each of the radial functions are motivated from different points of view and each may be shown to be optimal under some conditions [9]. These functions are very different: Gaussian is compact and positive, the second one is global and diverges at infinity.

The weight vector minimising the MSE of the network outputs is given by

$$\lambda \quad = (H^T H)^{-1} H^T t$$

where $\boldsymbol{H}$ is a design matrix.

Details of the numerical problems and implementation, including generalisation of the RBFNN with regularisation terms are discussed in detail in [3,4].

Predictive Eror Bars

In applying RBFNN for data mapping it is rare to find estimates of confidence placed on the neural network predictions (in our case – “thick” isolines, see below). We have used so-called perdictive error bars (PEB) approach to error bar estimation by neural networks developed in [8]. This approach is similar to predictions of estimate variance based on

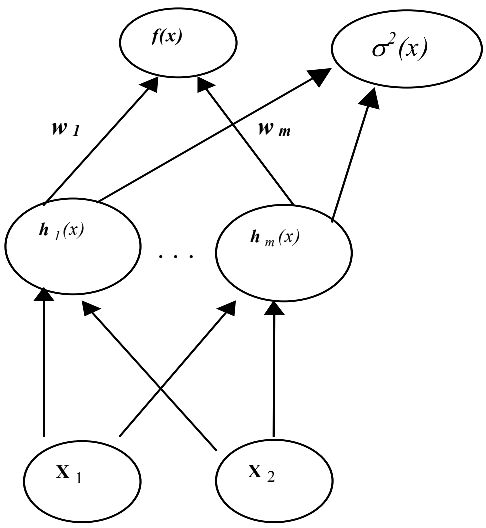
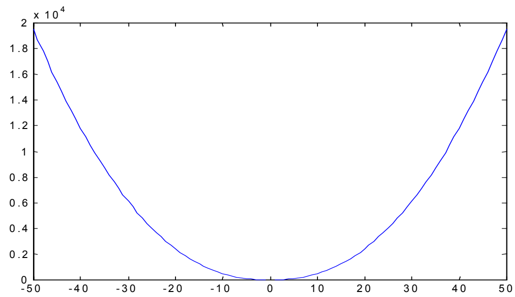
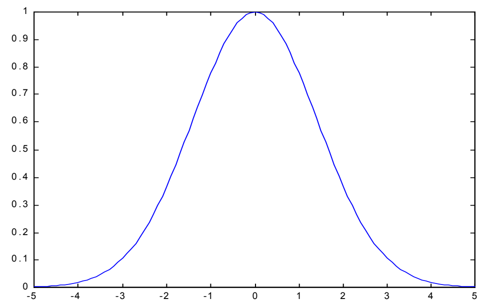


Figure 1. Single-layer RBFNN structure.



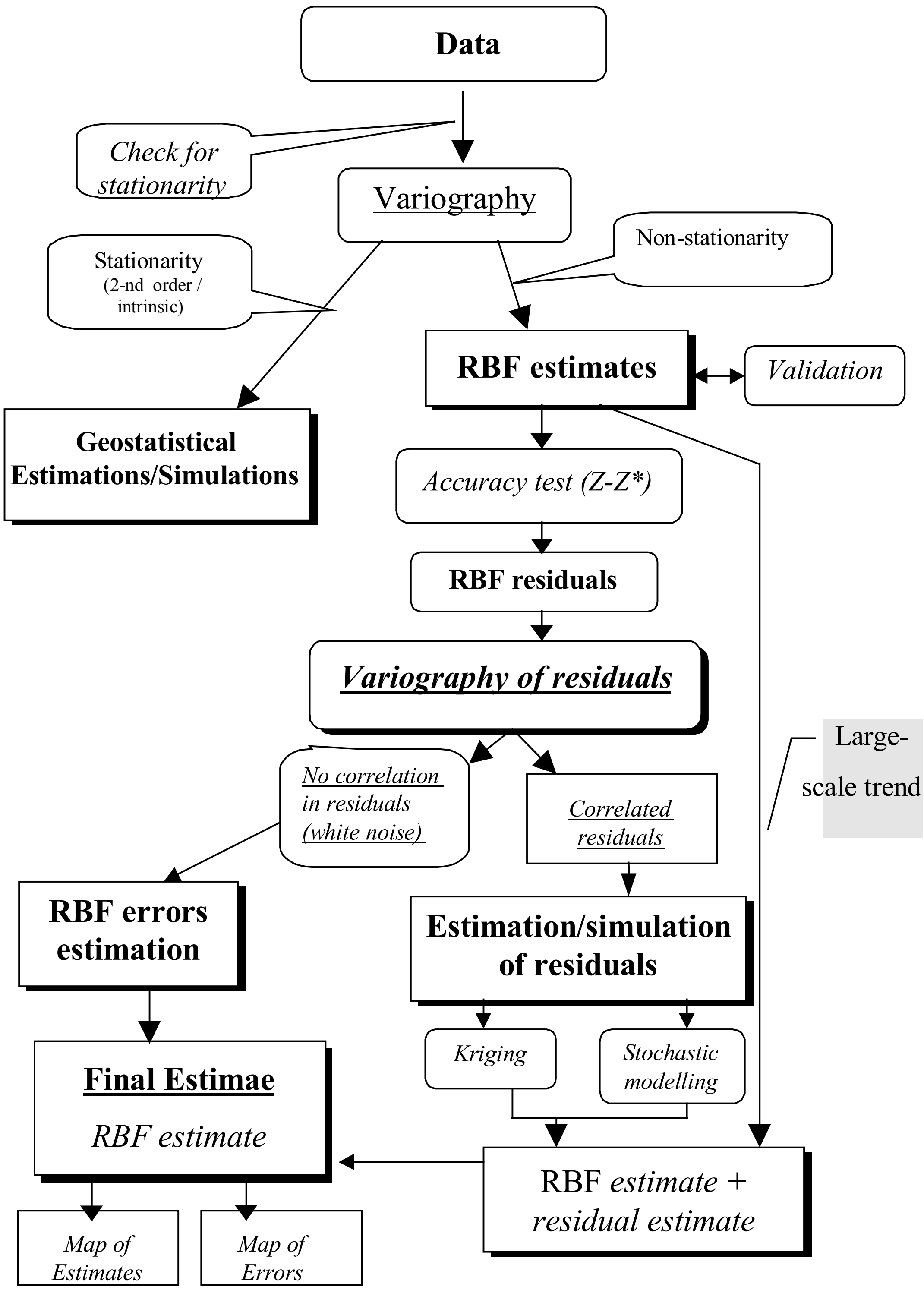


residuals and applied to General Regression Neural Networks in [7]. In the implementation of PEB there are two interlocked neural networks with the same input and hidden nodes but different output layer links to separate out[put nodes. One set produces the required regression to the desired target values whereas the second one (the error network) estimates the residual error on the training data set. Thus, the second network predicts the noise variance of the main neural model. [8].

Optimisation of the weights is a two-stage procedure:

1. The first stage determines the weights conditioning the regression on the mapping surface. Once these weights have been determined, the network approximations to the target values are known, and hence so are the conditional error values on the training samples.
2. In the second stage, the inputs to the network remain exactly as before, but now the target outputs of the network are the conditional error values. This second pass determines weights which condition the second set of output nodes to the variance (squared error values) $\sigma^2 (x,y)$.

Following [8], we employ the same hidden units for both the regression and the error networks, since the first layer determines the unconditional distribution of the data, or extracts the relevant feature space of the problem if the first layer weights are optimised.


Data
Check for stationarity
Variography
Stationarity (2-nd order / intrinsic)
Non-stationarity
RBF estimates
Validation
Geostatistical Estimations/Simulations
Accuracy test (Z-Z*)
RBF residuals
Variography of residuals
Large-scale trend
No correlation in residuals (white noise)
Correlated residuals
RBF errors estimation
Estimation/simulation of residuals
Kriging
Stochastic modelling
Final Estimae
RBF estimate
RBF estimate + residual estimate
Map of Estimates
Map of Errors

Decision-oriented mapping. Case studies

Training of the RBFNN

The flow-chart of the overall approach to the decision-oriented mapping with RBFNN and geostatistics is presented above.

RBF networks with different number *m* of hidden units, different types of basis function and different covariance matrices were applied. Based on Netlab software [10] the following training algorithm was updated implemented:

First stage:

1. *m* random input data points were chosen as centres for RBFs.
2. A few (3-8) iterations of K-means algorithm were performed.
3. Centres positions were adjusted by means of Expectation-Maximisation (EM) algorithm. Not more than 30 iterations of EM algorithm were enough for error to stabilise.

Second stage:

Weight vector was found using pseudoinverse activation matrix.

Root-Mean-Square-Error (RMSE) of network first decreases with the increase of the number of hidden units for both training and test sets. However, with further enlargement of hidden layer the networks with local and global RBFs showed different behaviours. Global RBF networks' RMSE on test set increased rapidly, revealing over-fitting to the data. Networks with Gaussian approximated the mapping with almost constant function. Therefore, their RMSE remained unchanged regardless the number of RBFs.

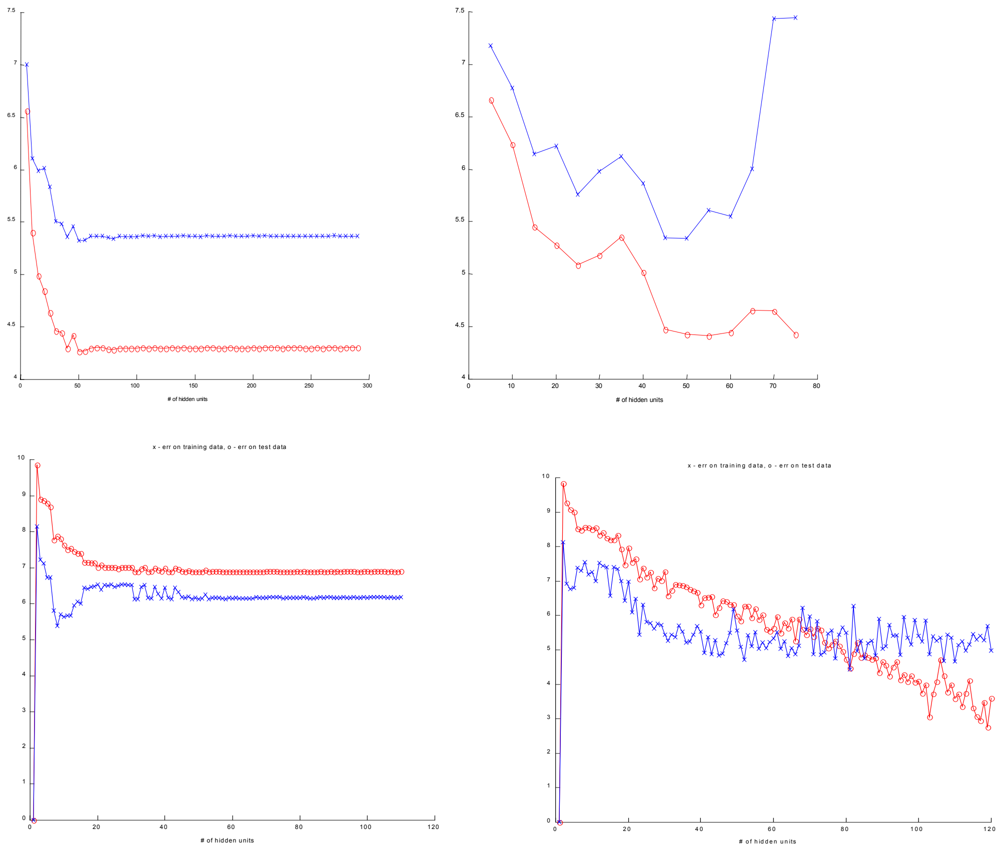


Figure 2. RMSE of network with different number of hidden units. Training (o) and testing (x) data sets: a – gaussian , b – thin plate spline RBF. I – 1st case study II – 2nd case study

Results of the Case Studies

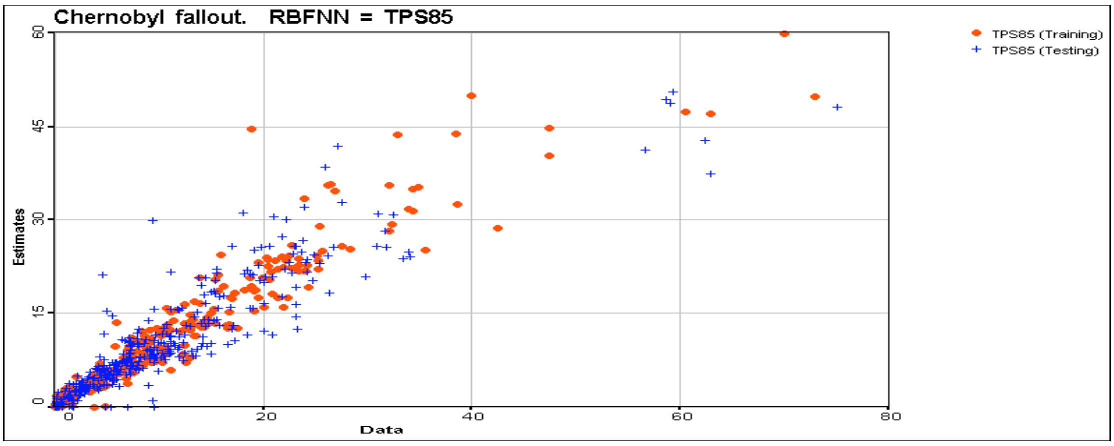


Fig. 3. Estimates (training and testing data sets) versus measurements (data). RBFNN: thin plate spline with 85 neurons.

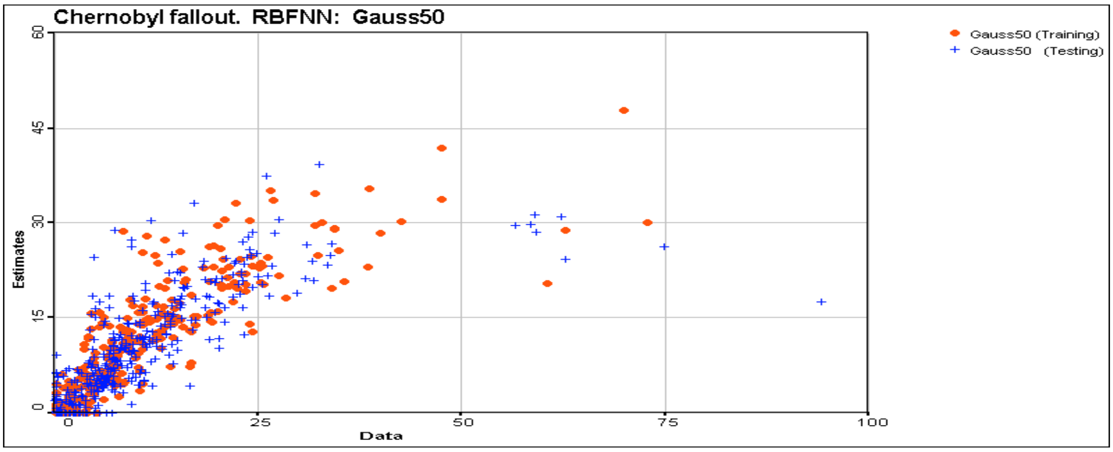


Fig. 4. Estimates (training and testing data sets) versus measurements (data). RBFNN: Gaussian kernels with 50 neurons.

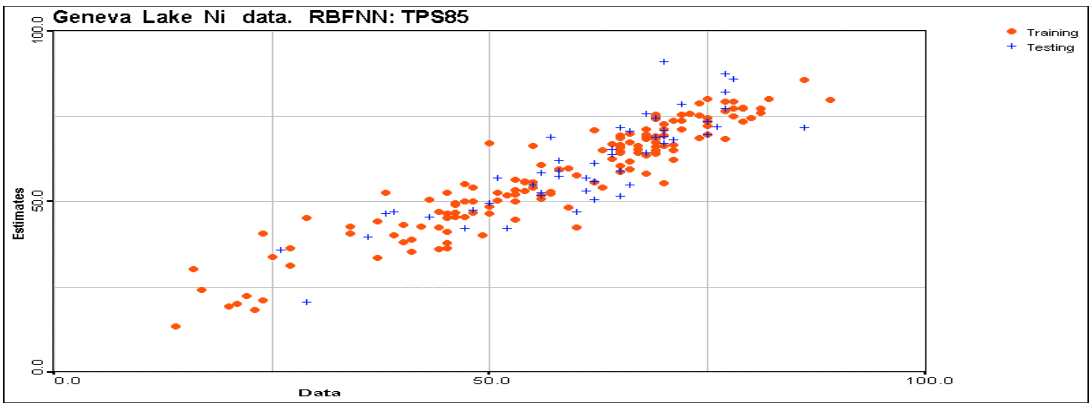


Fig.5 Estimates (training and testing data) versus measurements (data). Geneva Lake, Ni.

The best results were achieved for the network with 85 tps basis functions. Modified MATLAB code developed by Bishop and Nabney (available from http://www.ncrg.aston.ac.uk/netlab/) was used for network training.

In Table 1 RMSE of perceptron outputs for the first case study data are shown for comparison.

Table 1.

MLP structure	2-15-1	2-15-12-1	2-30-1	2-30-25-1	2-7-1	2-7-7-1	RBF
RMSE	7.9	7.5	7.5	7.8	7.9	7.8	4.0 - 6.0

Network predictions versus target data are plotted in Figs. 3-5 for training and testing data sets for the two case studies: Chernobyl fallout and Geneva Lake sediments contamination.

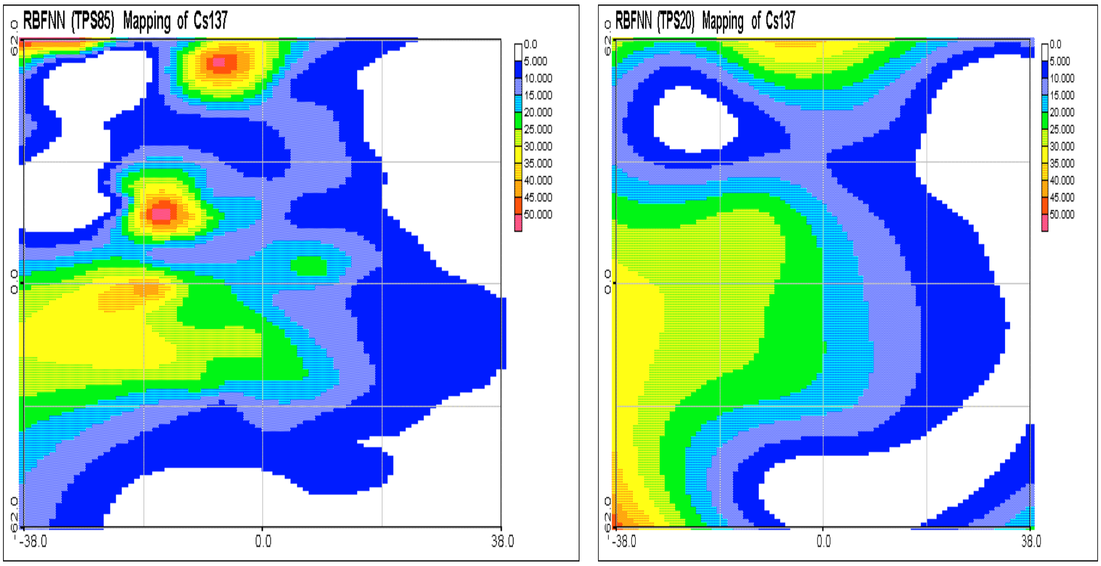


Figure 6. RBFNN mapping of Chernobyl fallout with thin plate spline RBF. TPS85 – 85 RBF centres, TPS20 – 20 RBF centres.

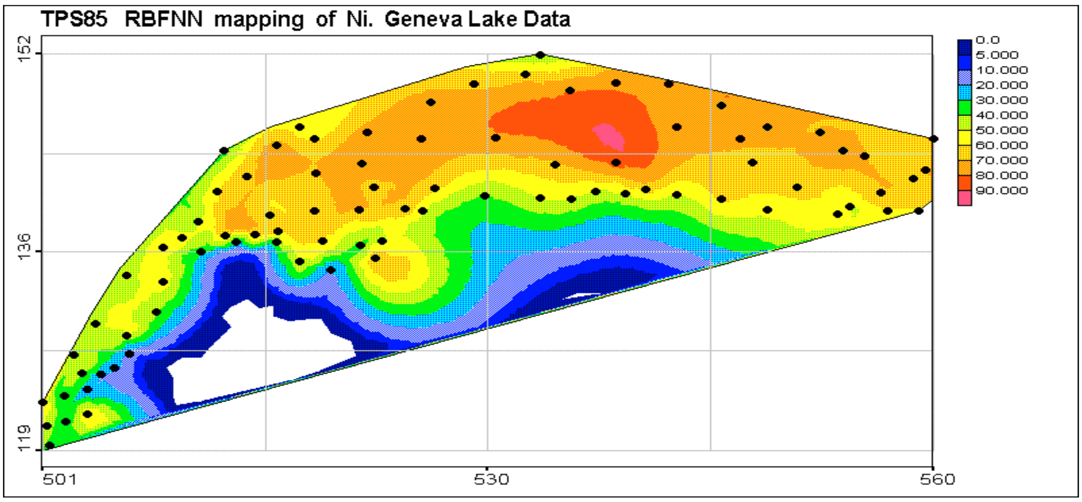


Figure. 7. RBFNN mapping of Geneva Lake Ni data. 85 TPS centres are presented as well.

It can be seen that network gives good approximation even for the test set.

Network predictions on the regular grids are shown in Fig. 6 with optimal (maximal extraction of the useful information from data, pure nugget effect on residuals) –85 centres. Suboptimal mapping with RBFNN is presented in Fig. 6 (oversmoothing, extraction only large scale structures from the data with spatial correlations of the residuals) – 20 thin plate spline RBF centres.

The same training procedure as described above was used for training the RBF net on the net's squares of the residuals (mapping of variance). The mapping of error (square root of the variance) on the regular grid is presented in Fig. 8.

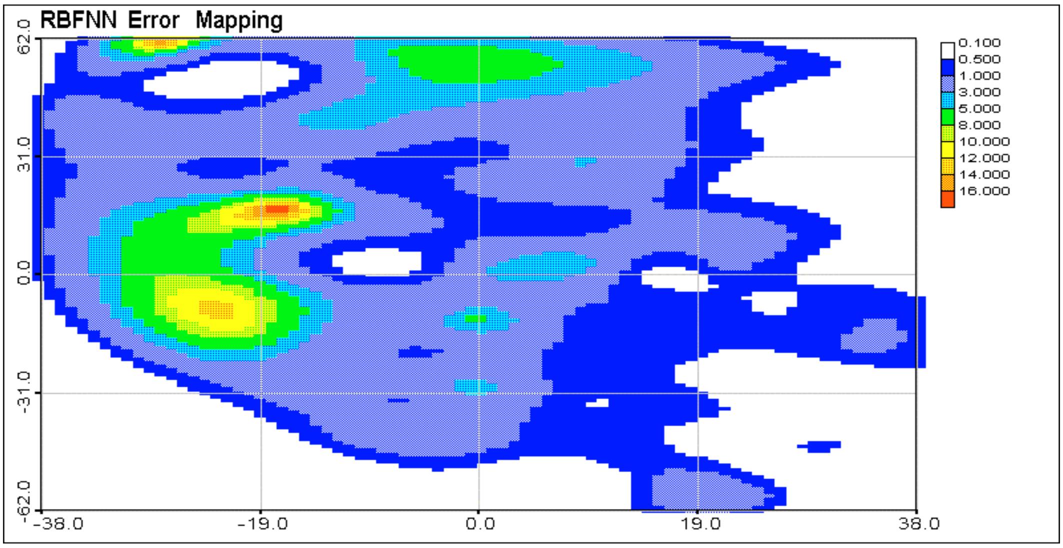


Figure 8. RBF mapping of Cs137 estimates standard deviations (square root of variance) $\sigma(x,y)$.

It is possible to draw "thick" contour isolnes using the results of the function and variance predictions. "Thick" isolines consist of 3 isolines: conditional mean value predicted by RBFNN $Z_m(x,y)$ and two isolines corresponding to $Z_m(x,y) \pm 2\sigma(x,y)$. This kind of visualisation gives an impression of spatial uncertainty as well. It should be noted that in case of normally distributed data "thick" isolines correspond to regression mapping along with confidence intervals.

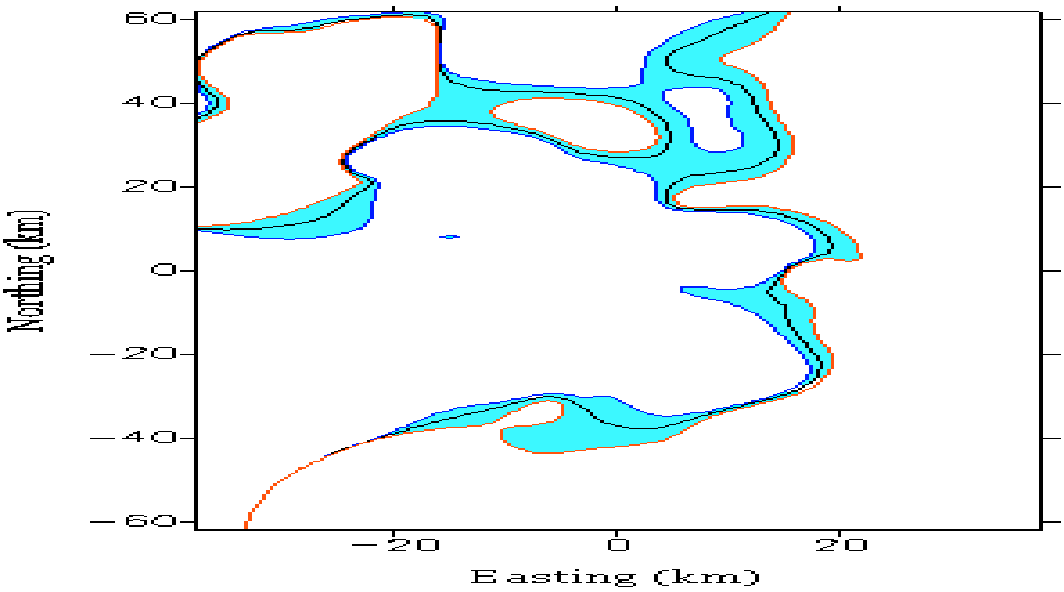


Figure 9. "Thick contour" isolines, Chernobyl fallout case study.

Geostatistical technique applied to the RBF network outputs shows that the variogram for the net residuals is pure nuget type and no correlation between residuals can be extracted further.

In order to control the quality of the spatial model developed by RBFNN geostatistical structural analysis (variography) was used. The main measure for the description of spatial continuity is so-caled variogram $\gamma(\Delta x)= E\{Z(x)-Z(x+\Delta x)\}^2$, where $E\{\ \}$ is an operator of mathematical expectation, Δx - vector of separation distance. Anisotropic variograms may be estimated using raw data [1].

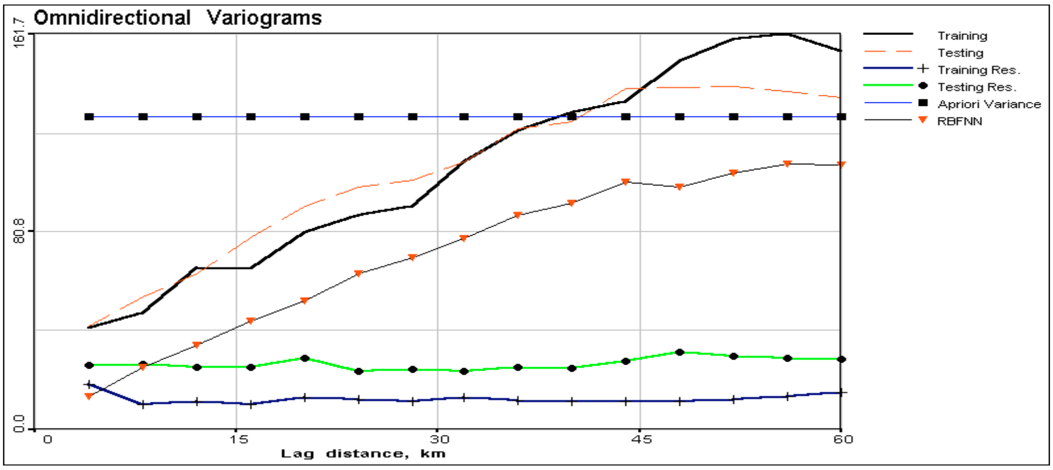


Figure 10. Omnidirectional variograms of training, testing, RBFNN output (training data), and residuals on training and testing data sets.

Omnidirectional variograms for the different data sets, RBFNN predictions and residuals are presented in Figure 10 for the Chernobyl case study. Training and testing data are spatially correlated with correlation distance about 38 km. There is some spatial nonstationarity (variograms go above a priori variance level). Variograms of the residuals both for training data and testing data demonstrate pure nugget effects. From the geostatistical point of view it means that all spatially correlated information has been extracted.

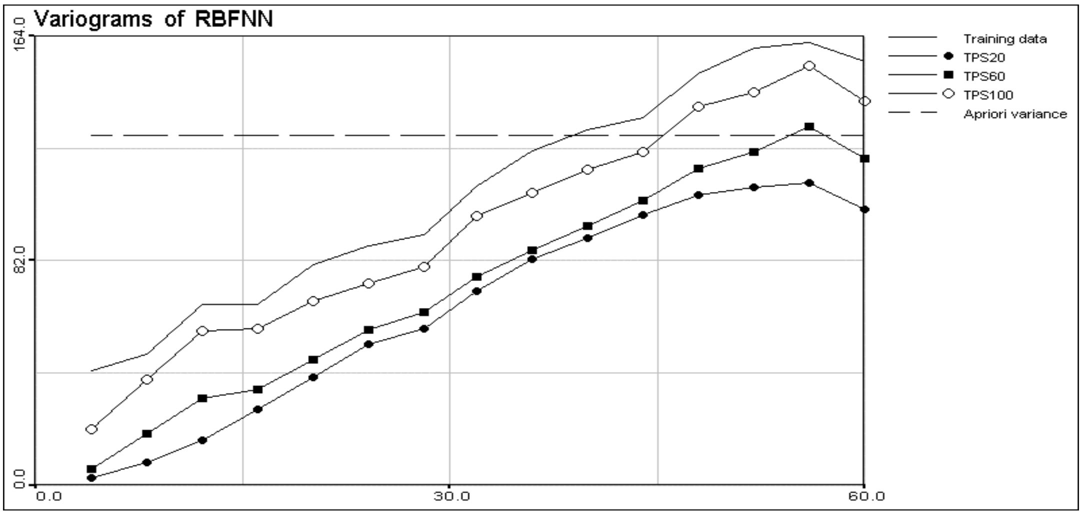


Figure 11. Omnidirectional variograms of the RBFNN models of different complexity.

In the Fig. 11 omnidirectional variograms of the RBFNN outputs are shown. Networks of different complexity (by changing of the number of RBF centres) have been used. The variograms demonstrate smoothing effects of the RBFNN models: from oversmoothing at 20 tps to close to overfitting at 100 RBF centres. Corresponding variograms of the residuals quantify the quality of modelling (extraction of the useful information described by spatial correlations from data).

Conclusion

Radial basis function neural networks were used for the decision-oriented environmental mapping. It was shown that RBFNN give the same results as a standard geostatistical technique: prediction mapping of the regression function and mapping of the predictions variance.

In general, radial basis functions (RBF) networks are special kind of neural networks, which employ both local and global functions to provide spatial estimates. They are easily adaptive to the monitoring networks by means of adjusting kernels positions. Geostatistical tools (variography) is used to control the quality of RBFNN mapping.

There was shown that estimations with global radial basis functions (thin plate splines) are able to model both large-sale and local structures of the spatial data distribution.

An original network was designed to calculate estimation error. "Thick" contours mapping technology was developed to produce decision-oriented maps.

Models developed were applied to the problem of decision-oriented mapping of environmental pollution. Two real case studies were considered: radioactive soil contamination after the Chernobyl fallout and heavy metal sediments in the Geneva Lake. Decision-oriented prediction maps accompanied by maps of estimation errors and "thick contours" are presented as the outputs for decision making.

Geostat Office software was used for the exploratory data analysis, variography and presentation of the results.

Acknowledgements

The work supported in part by Swiss National Science Foundation (Cartann) and European INTAS grant 96-1957.

References

1. P. Goovaerts. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 1997.
2. M. Kanevski, V. Demyanov, S. Chernov, E. Savelieva, A. Serov, V. Timonin, M. Maignan. Geostat Office for Environmental and Pollution Spatial Data Analysis. Mathematische Geologie, N3, April 1999, 1-12 pp.
3. S. Haykin. Neural Networks. A Comprehensive Foundation. Macmilln College Publ. Compny. N.Y. 1994.
4. C. Bishop Neural Networks for Pattern Recognition. Oxford: Oxford University Press, 1995.
5. Kanevsky M., Arutyunyan R., Bolshov L., Demyanov V., Maignan M. Artificial neural networks and spatial estimations of Chernobyl fallout. Geoinformatics. Vol.7, No.1-2, 1996, pp.5-11.
6. M. Kanevski. Lazy Mapping of Environmental and Pollution Data. Proc. Of IAMG'98 (The Fourth Annual Conference of the International Association for Mathematical Geology). A. Buccianti, G. Nardi, R. Potenza (Eds.) Napoli, 1998,, pp. 154-160.
7. M. Kanevski. Spatial Predictions of Soil Contamination Using General Regression Neural Networks. Int. J. Systems Research and Info. Science. 1998, pp. 1-16.
8. D. Lowe and K. Zapart Point-wise Confidence Interval Estimation by Neural Networks: A comparative study based on automotive engine calibration. Neural Computing Research Group, Technical Report NCRG/98/007.www.ncrg.aston.ac.uk
9. A. Webb and S. Shannon. Shape-Adaptive Radial Basis Functions. IEEE Transactions on Neural Networks, vol.9, No.6 pp. 1155-1166.
10. C. Bishop and I. Nabney. Netlab. http://www.ncrg.aston.ac.uk/Netlab/