Российская Академия Наук

РОССИЙСКАЯ АКАДЕМИЯ НАУК

ИНСТИТУТ ПРОБЛЕМ БЕЗОПАСНОГО РАЗВИТИЯ АТОМНОЙ ЭНЕРГЕТИКИ

RUSSIAN ACADEMY OF SCIENCES

NUCLEAR SAFETY INSTITUTE

Препринт ИБРАЭ № IBRAE-2002-13

Preprint IBRAE-2002-13

R. Parkin, M. Kanevski, A. Pozdnukhov, V. Timonin, M. Maignan, B. Yatsalo, S. Canu

ENVIRONMENTAL DATA MINING AND MODELLING BASED ON MACHINE LEARNING ALGORITHMS AND GEOSTATISTICS

Москва
2002

Moscow
2002

УДК 502.3

Паркин Р., Каневский М., Позднухов А., Тимонин В., Майгнан М., Яцало Б., Каню С. РАСКОПКА И МОДЕЛИРОВАНИЕ ДАННЫХ ПО ОКРУЖАЮЩЕЙ СРЕДЕ С ПОМОЩЬЮ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ И ГЕОСТАТИСТИКИ. (На англ. яз.). Препринт № IBRAE-2002-13. Москва: Институт проблем безопасного развития атомной энергетики РАН, 2002. 13 с. — Библиогр.: 9 назв.

Аннотация

В данной работе представлено несколько современных подходов для анализа, обработки и представления пространственных данных по окружающей среде. Набор используемых средств содержит алгоритмы машинного обучения (Многослойный Персептрон и Регрессия на поддерживающих векторах) и недавно разработанные геостатистические предсказатели и симуляционные модели. Новаторская часть данной работы представлена смешанными/гибридными моделями, включая как кригинговые оценки последовательные гауссовы симуляции на невязках алгоритмов машинного обучения. Цель смешанных моделей двояка: с одной стороны, алгоритмы машинного обучения эффективно решают проблемы пространственной нестационарности, которые сложны для геостатистического подхода; с другой стороны, геостатистические средства широко и успешно применяются, чтобы охарактеризовать возможности методов машинного обучения, анализируя качество и количество пространственной структурированной информации, выделенной из данных с помощью методов машинного обучения. Кроме того, подходы, основанные на композиции алгоритмов машинного обучения и геостатистической модели, являются привлекательными для различных процессов принятия решений.

Parkin R., Kanevski M., Pozdnukhov A., Timonin V., Maignan M., Yatsalo B., Canu S. ENVIRONMENTAL DATA MINING AND MODELLING BASED ON MACHINE LEARNING ALGORITHMS AND GEOSTATISTICS. Preprint IBRAE-2002-13. Moscow: Nuclear Safety Institute RAS, 2002. 13p. — Refs.: 9 items.

Abstract

The paper presents some contemporary approaches to the spatial environmental data analysis, processing and presentation. The main topics are concentrated on the decision–oriented problems of environmental and pollution spatial data mining and modelling: valorisation and representativity of data with the help of exploratory data analysis, topological, statistical and fractal measures of monitoring networks, spatial predictions and classifications, probabilistic and risk mapping, development and application of conditional stochastic simulation models. The set of tools used consists of machine learning algorithms (MLA) – Multilayer Perceptron and Support Vector Regression, and recently developed geostatistical predictive and simulation models. The innovative part of the report deals with integrated/hybrid models, including ML Residuals Kriging predictions and ML Residuals Sequential Gaussian simulations. The objective of the integrated models is twofold: from one side ML algorithms efficiently solve problems of spatial non-stationarity, which are difficult for geostatistical approach; from another side geostatistical tools are widely and successfully applied to characterise the performance of the ML algorithms, analysing the quality and quantity of the spatially structured information extracted from data by ML. Moreover, mixture of ML data driven and geostatistical model based approaches are attractive for decision-making process.

Environmental Data Mining and Modelling Based on Machine Learning Algorithms and Geostatistics

R. Parkin, M. Kanevski (1), A. Pozdnukhov, V. Timonin, M. Maignan (2), B. Yatsalo (3), S. Canu (4)

Institute of Nuclear Safety (IBRAE)
B. Tulskaya 52, 113191 Moscow, Russia
phone: (095) 955-22-31, fax: (095) 958-11-51, e-mail: park@ibrae.ac.ru

(1) IDIAP Dalle Molle Institute for Perceptual Artificial Intelligence, Simplon 4, 1920 Martigny, Switzerland

(2) Lausanne University, Lausanne, Switzerland

(3) OINPE Institute, Obninsk, Russia

(4) INSA, Rouen, France

Contents

Contents....3
1 Introduction....3
2 Machine Learning Residual Gaussian Simulations....4
2.1 The methodology of ML Residual Gaussian Simulations....4
2.1 Description of Multilayer Perceptron model....5
2.3 Description of Support Vector Regression model....6
2.3.1 Prediction problem....6
2.3.2 Empirical and Structural Risk Minimization....7
3 Case Study....8
4 Conclusions....13
5 Acknowledgements....13
6 References....13

1 Introduction

Most environmental data represent a combination of several spatial phenomena of different origin and appear as the complex spatial patterns at different scales. In some cases the original observations are taken with significant measurement errors and may contain a number of outliers. Spatial trends reproducing large-scale processes complicate variography – a basic geostatistical tool, describing spatial correlations, and sometimes make difficult or impossible developing of a valid variogram model. These and other facts complicate analysis, processing and interpretation of the results. Usually it is supposed that data can be decomposed into two parts: $Z(x)=M(x)+e(x)$, where $M(x)$ represents large scale deterministic variations (trends), and $e(x)$ represents small scale stochastic variations. Geostatistical approach offers several possible models in case of spatial trends (spatial non-stationarity): universal kriging, residual kriging, moving window regression residual kriging, science-based approaches, etc. These approaches have been considered by Cressie [1991], Deutsch et al. [1992], Dowd [1994], Neuman et al. [1984], Gambolati et al. [1987] and Haas [1996]. Each of these methods has its own advantages and drawbacks.

The present work is an extension (development of Neural Network Residuals Sequential Gaussian simulations (NNRSGS)) of the ideas presented by Kanevsky et al. [1996], where hybrid model – Neural Network Residuals Kriging (NNRK) – has been presented for the first time. The basic idea is to use feedforward neural network

(FFNN), which is a well-known global universal approximator to model large-scale nonlinear trends, and then to use geostatistical interpolators/simulators for the residuals.

One of the principal advantages of ML algorithms is their ability to discover patterns in data, which are so obscure as to be imperceptible to human researches and standard statistical methods; the data exhibit significant unpredictable non-linearity. Containing no data behaviour model, MLA depends only on the input data and the inner structure of the model, e.g. number of neurones, hidden layers, types of connections, information flow direction. MLA, depending on its architecture, can capture spatial peculiarities of the pattern at different scales describing both linear and non-linear effects. The performances of MLA are based on solid theoretical foundations.

The objective of the integrated models developed in the current paper is twofold: from one side MLA efficiently solve problems of spatial non-stationarity, which are difficult for geostatistical approach; from another side geostatistical tools are widely and successfully applied to characterise the performance of the MLA by analysing the quality and quantity of the spatially structured information extracted from data. Moreover, mixture of ML data driven and geostatistical model based approaches are attractive for decision-making process because of their interpretability. The real case study on soil pollution is considered in detail: Chernobyl fallout – large-scale contamination of environment by radiologically important radionuclides. Details of the data can be found in Kanevski et al [1996].

2 Machine Learning Residual Gaussian Simulations

2.1 The methodology of ML Residual Gaussian Simulations

The present work deals with an important development of hybrid MLA+geostatistical models firstly presented by Kanevsky et al. [1996] towards probabilistic/risk mapping. In short, the basic idea is to use MLA to develop a nonparametric, robust model for the large-scale non-linear structures (detrending) and then to use geostatistical models for the analysis of residuals - modelling of small scale structured variations. Lets look more closely into the original ML Residual Simulations.

1. The data is prepared for the analysis: split into training and validation set, checked for outliers, analysed with variography tools. If ANN is used at the first step, the data is scaled on the interval [0.1, 0.9] to facilitate the training procedure.

2. Further ML algorithm is applied. Without loss of generality, in the present study Multilayer Perceptron (MLP) and Support Vector Regression (SVR) are used. They are well known function approximators. The consideration of these algorithms for application in ML Residual Simulations is presented below. Accuracy test is MLA estimation at training points. It shows how well the MLA has been trained. Validation procedure – when MLA estimates values at validation points, which have not been used for training, – is a test of overall MLA performance, its ability to generalise and is especially used to avoid overtraining.

3. Accuracy test provides MLA residuals (*estimated* - *measured*) which are the base of the further analysis. Two cases are possible:

   - residuals are not correlated with the measurements, which means, that ANN has modelled all spatial structures represented in the raw data;

   - residuals show some correlation with the samples, than further analysis must be performed on the residuals to model this correlation.

The remaining spatial correlation represents short-range correlation structures. Long-range correlation (trend) in the whole area beyond the hot spots is very well modelled by MLA.

4. MLA residuals are explored using variography tools. Normal score transformation is performed to prepare data for further Gaussian simulations.

5. Sequential Gaussian simulation is applied to the MLA residuals.

The idea of stochastic simulation is to develop a spatial Monte Carlo model/generator that will be able to generate many, in some sense equally probable, realisations of the random function (in general, described by joint probability density function). Any realisation of the random function is called a nonconditional simulation.

Realisations that honour the data are called conditional simulations. Basically, the simulations are trying to reproduce first (univariate distributions) and second moment (variograms). The realisations are determined by the conditional data, simulation model and random seed. The similarities and dissimilarities between realisations describe spatial variability and uncertainty. Simulations bring valuable information for the decision-oriented mapping of pollution. Postprocessing of simulations gives rise to probabilistic maps: maps of probabilities to be above/below some predefined decision levels. Gaussian random function models are widely used in statistics and simulations due to their analytical simplicity, they are well understood, and they are limit distributions of many theoretical results. They were successfully applied in many cases. In this work we shall use algorithm known as a Sequential Gaussian Simulations. Details on the models description and on the implemented algorithm can be found in Deutsch and Journel [1992].

6. Simulation value of the residuals appears after back normal transformation. Final ML Residual Simulations value is a sum of MLA estimate and sequential Gaussian simulation value of the residuals.

2.1 Description of Multilayer Perceptron model

Let $Z(\mathbf{s})$ be the Spatial Random Field (SRF) representing, for example, the ^{137}Cs concentration at the location $\mathbf{s}=(s_1,s_2)$. In this work we assume that the non-stationary SRF $Z(\mathbf{s})$ may be written as $Z(\mathbf{s})= m_z(\mathbf{s})+ X(\mathbf{s})$, where $m_z(\mathbf{s})$ is the large-scale spatial trend and $X(\mathbf{s})$ is a stationary SRF. We use the MLP method to estimate the spatial trend $m_z(\mathbf{s})$.

Multi-layer perceptron (MLP) is a type of artificial neural network with a specific structure and training procedure that are briefly described below (Haykin, S., 1999; Bishop, C.M., 1995.).

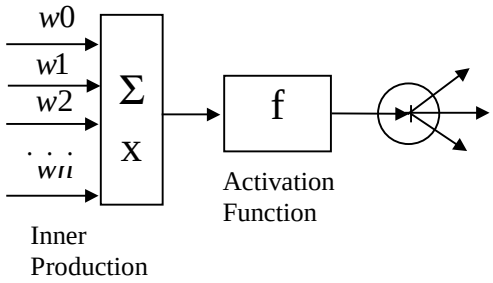

Figure 1. Formal neuron

The main component of MLP is the *formal neuron* (Figure 1), which sums the inputs, and performs a transform via the *activation function.* The activation function (or non-linear transformer) may be given by any continuous, bounded and non-decreasing function, such as the commonly used exponential sigmoid or hyper tangent. Activation function was employed only in hidden unites – output neurons have linear function. MLP of such structure is a universal approximator. The weighs $\boldsymbol{W}$ are adaptive parameters for the optimisation procedure, which uses the following quadratic MSE cost function,

$$E = \frac{1}{N}\sum_{n=1}^{N}(t_n - o_n)^2 \tag{1}$$

where E is the mean square error (MSE), N is the number of samples, o is the net output (prediction), and t is the real function value.

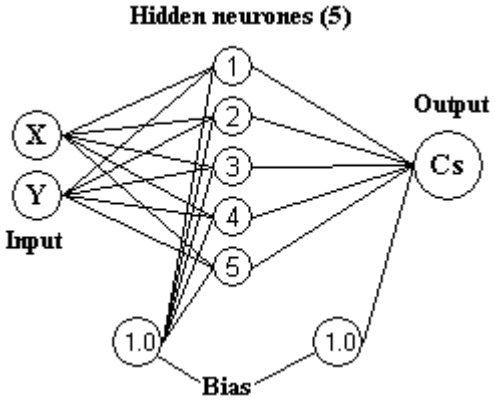

Figure 2. Multilayer Perceptron

In a standard MLP neurons are arranged in input, hidden and output layers. The sampled function parameters (X and Y co-ordinates) are exposed to the input layer, the output layer produces and compares the target functions value (value of ^{137}Cs concentration), and (one or two) hidden layers allow(s) to deal with non-linearity (see Figure 2). The number of neurons in the hidden layers can vary and it is the user's task to choose the optimum configuration. As long as the aim of MLP in the present work is to extract large-scale trend, the number of hidden neurons was chosen as small as possible to be able extract the non-linearity. Further increase of the number of hidden neurons leads to extracting more detailed local peculiarities of the pattern, then in such case not only general trend will be modeled by MLP that contradicts the idea of the approach. Choosing too many hidden neurons will then lead to over-fitting (or over-learning) when MLP looses its ability to generalize the information from the samples. On the other using too few hidden neurons does not provide explicit extraction of the trend, hence some large-scale correlation will remain in the residuals restricting further procedure. Also the complexity of the MLP must be consistent with the amount of information for training – there should be enough data to match every connection. The conclusion on how many hidden neurons to choose was made basing on the variogram analysis of MLP estimates and remaining residuals at training points.

Back-propagation error algorithm is applied to calculate gradient of MSE on adaptive weight, $\partial E / \partial W$. Various optimisation algorithms, which employ back propagation, can be used such as the conjugate gradient deseand method, second-order pseudo-Newton Levenberg-Marquardt method, or the resilient propagation method. The selection of the training algorithm was governed by the same principle as the number of hidden neurones – optimal for extracting the large-scale trend where variogram was used as a selection criterion.

2.3 Description of Support Vector Regression model

Support Vector Regression (SVR) is a recent development of the Statistical Learning Theory (Vapnik-Chervonenkis theory). It is based on Structural Risk Minimisation and seems to be promising approach for the spatial data analysis and processing Kanevski et al [2001]. There are several attractive properties of the SVR: robustness of the solution which is important in many applications, sparseness of the regression, automatic control of the solutions complexity, good generalisation Vapnik [1998]. In general, by tuning SVR hyper-parameters it was possible to cover the range of spatial function regression from overfitting to oversmoothing.

2.3.1 Prediction problem

Assuming **f** is a prediction function (i.e. a function used to predict the value of Z knowing the geographical co-ordinates), we define the cost of choosing this particular function for a given decision process. First, for a given observation (x,y,Z) we define the ε-insensitive cost function:

$$C\{(x,y),Z,\varepsilon,f\} = \begin{cases} |f(x,y)-Z|-\varepsilon & \text{if } |f(x,y)-Z| \rhd \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where ε characterizes some acceptable error.

Now, for all possible observations we define the global or generalization error also known as the integrated prediction error IPE:

$$IPE(f) = \int E_Z(C((x,y,)Z,\varepsilon,f))\omega(x,y)dxdy \quad (3)$$

where ω(x,y) is some economical measure, indicating the relative importance of a mistake at point (x,y). In case of non-homogeneous monitoring networks this function can take into account spatial clustering. Usually ω(x,y)=1, so that all positions are assumed to be equally important.

2.3.2 Empirical and Structural Risk Minimization

Let us assume that solution is a function that can be decomposed into two different components: a trend plus a remaining random process.

$$\hat{f}(x,y) = \sum_{k=1}^{m} w_k \varphi_k(x,y) + \sum_{j=1}^{J} \beta_j K_j(x,y) \quad (4)$$

where $K_j(x,y)$ is a basis of the trend component and φ_k, k=1,..m is an orthonormal basis of the remaining part (note that m can be infinity).

The complexity of the solution can be tuned through $||w||^2=\Sigma_{k=1..m}w_k^2$. Thus, a relevant strategy to minimize IPE is to minimize the empirical error together with maintaining $||w||^2$ small. This can be obtained by minimizing the following cost function:

$$\begin{cases} \text{minimize} \quad \frac{1}{2}\|w\|^2 \\ \text{subject to } |f(x_i,y_i) - Z_i| \le \varepsilon_i \text{ , for i = 1,...n} \end{cases} \quad (5)$$

When data lie outside of this epsilon tube due to noise or outliers making these constraints too strong and impossible to fulfill, Vapnik suggested to introduce slack variables ξ_i , ξ_i^*. These variables measure the distance between the observation and the ε tube.

Note that by introducing the couple (ξ_i , ξ_i^*) the problem has now 2n unknown variables. But these variables are linked since one of the two values is necessary equals to zero. Either the slack is positive ($\xi_i^* = 0$) or negative ($\xi_i = 0$). Thus, $Z_i \in [f(x,y)- \varepsilon -\xi_i, f(x,y)+ \varepsilon +\xi_i^*]$.

Following the ideas as in the case of SVM classification we arrive at the following optimization problem:

$$\text{minimise } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (6)$$

$$\text{subject to } \begin{cases} f(x_i,y_i) - Z_i - \varepsilon_i \le \xi_i \\ -f(x_i,y_i) + Z_i - \varepsilon_i \le \xi_i^* \\ \xi, \xi_i^* \ge 0 \quad \text{for i = 1,...n} \end{cases} \quad (7)$$

The dual formulation of the problem (21) is formulated as follows:

$$\begin{aligned} \text{maximise } & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)\left(\sum_{k=1}^{m}\varphi_k(x_i,y_i)\varphi_k(x_j,y_j)\right)(\alpha_j^* - \alpha_j) \\ & - \sum_{i=1}^{n}\varepsilon_i(\alpha_i^* + \alpha_i) + \sum_{i=1}^{n} Z_i(\alpha_i^* - \alpha_i) \\ \text{subject to } & \begin{cases} \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)K_j(x_i,y_i) = 0 \text{ for } K_j = 1,...m \\ 0 \le \alpha_i^*, \alpha_i \le C \quad \text{for i,...n} \end{cases} \end{aligned} \quad (8)$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ are the Lagrangian multipliers associated with the constraints. They can be roughly interpreted as a measure of the influence of the constraints in the solution. A solution with $\alpha_i = \alpha_i^* = 0$ can be interpreted as “the corresponding data point has no influence on this solution”.

By using kernel trick this problem can be solved without direct modeling in a feature space (the same as in non-linear classification). To do so it is necessary to choose φ_k such that:

$$\sum_{k=1}^{m} \varphi_k(x_i, y_i)\varphi_k(x_j, y_j) = G\big((x_i, y_i),(x_j, y_j)\big) \tag{9}$$

This is the case in reproducing kernel Hilbert space, where G is the reproducing kernel. Functions φ_k are the eigen functions of G. In this case the solution can be formulated in the following form:

$$\hat{f}(x,y) = \sum_{i=1}^{n} v_i G((x,y),(x_i,y_i)) + \sum_{j=1}^{m} \beta_j K_j(x,y) \tag{10}$$

with $v_i = (\alpha_i^* - \alpha_i)$. This solution only depends on the kernel function *G*.

3 Case Study

Radioactive soil contamination caused by the Chernobyl fallout feature anisotropic highly variable and spotty spatial pattern. The multi-scale character of the pattern is due to numerous influencing factors. Structural analysis of sample data discovers limitation for use of stationary estimation/simulation models, like kriging or stochastic simulation.

Exploratory spatial data analysis deals with the following steps: statistical analysis, spatial moving window statistics and trend analysis. This is an important phase of the study both for the MLA and geostatistical analyses. The basic statistical parameters of the Chernobyl data are following: minimum value Cs137=5.9, mean value ^{137}Cs =571.8, maximum value ^{137}Cs =4333.9, variance ^{137}Cs =315372, skewness ^{137}Cs =2.7 and kurtosis ^{137}Cs =16.9. As usually environmental data are positively skewed and their distributions are far from normal. Concentrations are measured in kBq/m^2 (see Figure 3).

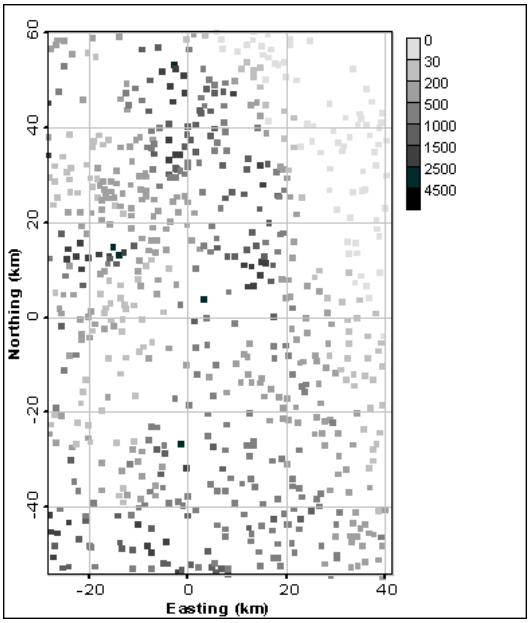

Figure 3. Raw data on ^{137}Cs concentration in the Briansk region

Variogram analysis of normal score data discovered long-range structures (50 km) and local correlation (10-15 km) (see Figure 4 and 5). This conclusion leads to MLA use for trend modelling. Another problem with not de-trended data is that normal score variogram does not reach the *sill=1*, which is required for normally distributed variable.

Then the phase of study deals with preparing of training and testing (validation) datasets, scaling of the data, non-linear transformations, if necessary. Distributions of training and validation locations are presented in Figure 6. The following step is to select of MLA algorithms (architecture and learning model): Multilayer Perceptron and Support Vector Regression.

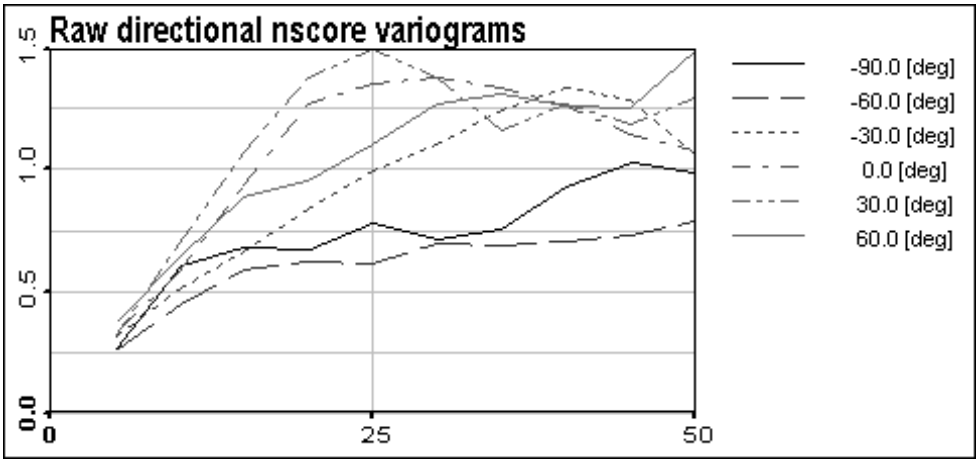

Figure 4. Raw directional variograms for normal score ^{137}Cs samples

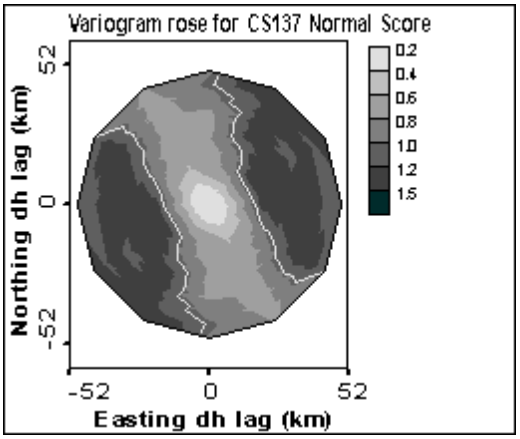

Figure 5. Raw variogram rose for normal score ^{137}Cs samples

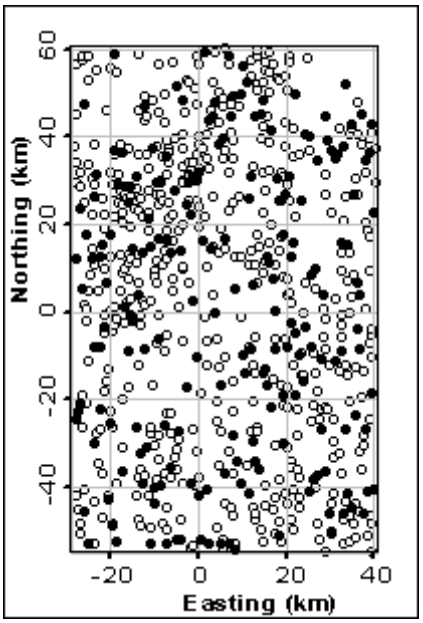

Figure 6. Location of training (circles) and validation (filled circles) points of the dataset

In the present study the MLP models with the following was used: 2 input neurons, describing spatial co-ordinates (X, Y); one or two hidden layers; output neuron describing ^{137}Cs contamination. An important step deals with training and testing of the network. Backpropagation training with conjugate gradient, steepest descent, Levenberg-Marquardt, simulated annealing and genetic optimisation algorithms in order to avoid local minima were used. The trained network has been evaluated by using cross-validation, and accuracy tests - prediction of the training data set with trained ANN. Accuracy test is used as a simple test describing how ANN captured the correlation between locations and contamination. The network has been validated by using independent data set. Then ANN is used for Cs137 spatial predictions/generalisations - mapping. Result for the ^{137}Cs ANN large-scale mapping is presented in Figure 3a.

This result was obtained by using 5 hidden neurones respectively. It is evident that ANN has learned non-linear trends and that small-scale variations have been ignored. By using more hidden neurones it was practically impossible to detect all structured small-scale variations.

Support Vector Regression (SVR) is a recent development of the Statistical Learning Theory (Vapnik-Chervonenkis theory). It is based on Structural Risk Minimisation and seems to be promising approach for the spatial data analysis and processing Kanevski et al [2001]. There are several attractive properties of the SVR: robustness of the solution which is important in many applications, sparseness of the regression, automatic control of the solutions complexity, good generalisation Vapnik [1998]. In general, by tuning SVR hyper-parameters it was possible to cover the range of spatial function regression from overfitting to oversmoothing.

Let us present the results of large scale modelling using Support Vector Regression approach. High flexibility of SVR makes different combinations of the parameters suitable for trend modelling. The following parameters were selected: isotropic RBF kernel, kernel bandwidth – 20 km. This choice is based both on the analysis of training and testing errors and the analysis of the variogram of the resulting trend model. The results are presented in Figure 7b. X and Y co-ordinates are in cell numbers, cell size = dXdY=[1x1] sq.km.

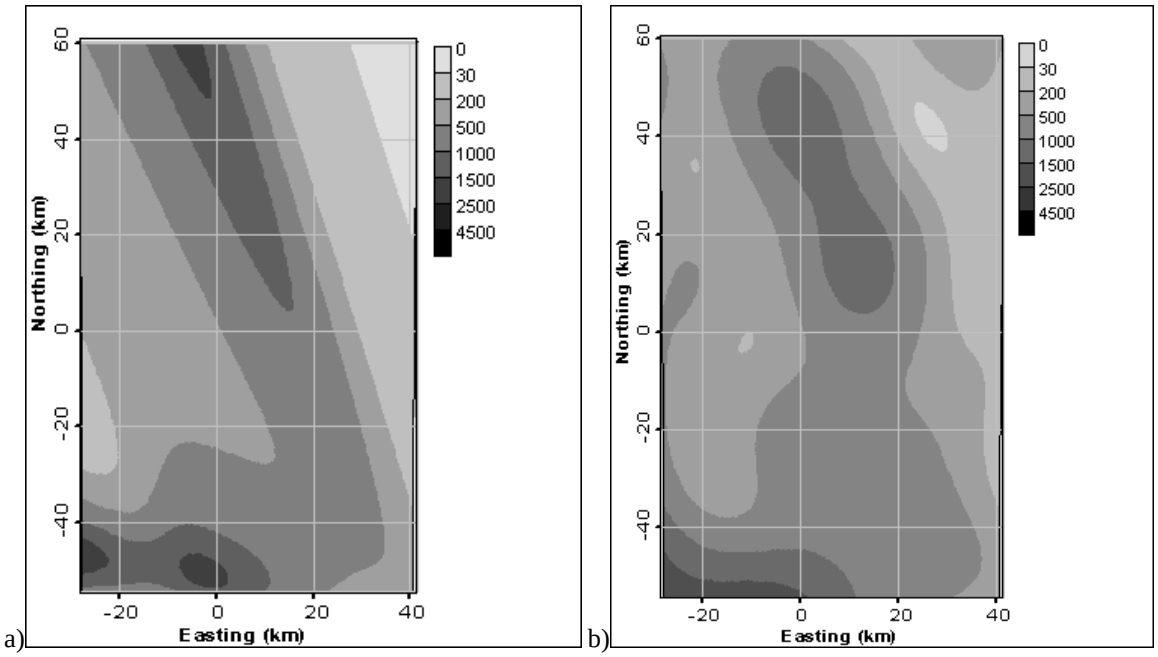

Figure 7. ^{137}Cs concentrations, a) Artificial Neural Network (one hidden layer with 5 neurones) spatial predictions and b) Support Vector Regression trend modelling

Trained neural network and Support Vector Regression are able to extract some information described by spatial correlations from the data. The rest information – small scale spatially structured residuals - was analysed and modelled with the help of geostatistical approach, using conditional stochastic simulations model. Obtained residuals are spatially correlated with the original data and are not correlated with MLA estimates. In the present research so called sequential Gaussian simulations were applied to the MLA residuals.

Exploratory variography of spatial correlation structures (variogram) of the Normal score transformed residuals are presented in Figures 8 and 9. Variograms of the Normal score transformed residuals can be easily modelled (fitting to theoretical model) and Sequential Gaussian simulations can be applied (variogram reaches a

sill and stabilises). Range (distance at which variogram reaches a sill - a priori variance of data, dashed line) of the variogram has been changed to shorter distances in comparison with Figure 4.

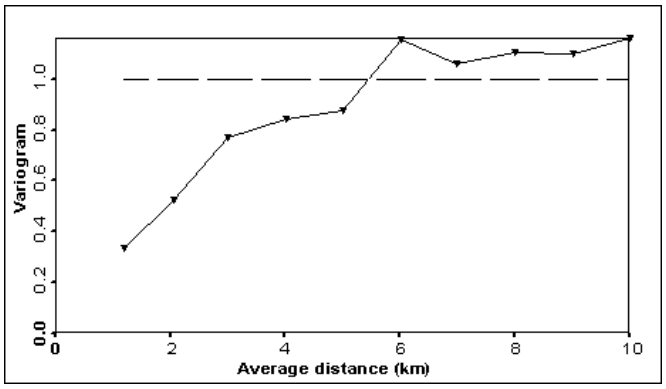

Figure 8. Omni-directional variogram of the ANN residuals

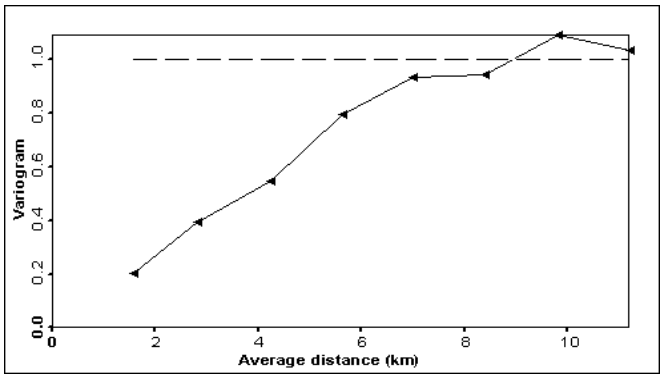

Figure 9. Omni-directional variogram of the SVR residuals

Final ML Residual Sequential Gaussian Simulation results are presented as equiprobable realisations in Figures 10a and 10b.

The final stage is a validation of the ML Residual Sequential Gaussian Simulation results. There is much more variablility on the maps in Figures 10a and 10b than on the maps in Figures 7a and 7b respectively, which describes only large-scale trends. ML Residual Sequential Gaussian Simulations model is exact model - it honours the measured data: when measurements errors are negligible at sampling points ML Residual Sequential Gaussian Simulations estimates equals measurements. Comparisons with geostatistical prediction models were carried out. Proposed models give comparable or better results on different data sets. Comprehensive comparisons with other ML methods are a topic of current research.

Several important points should be mentioned.

1) Analysis of residuals is an important also in case when only MLA mapping is applied. This helps to understand the quality of the results. If there is no spatial correlations between residuals it means that all spatial information from data have been extracted and MLA can be used for prediction mapping as well.

2) Robustness of the approach: how is it sensible to the selection of the MLA architecture and learning algorithm. Kanevsky et al. showed that summary statistics of residuals described by variograms is robust versus ANN architecture – number of hidden layers and neurones. The same robust behavior in the case presented in this study has been obtained both for ANN and SVR (varying model parameters). So, we can choose the simplest models from MLA capable to learn and catch non-linear trends. Usually accuracy test have been used for the analysis and description of what have been learned by MLA. Accuracy test measures correlations between training data set and MLA predictions at the same points.

3) Data clustering is a well-known problem in a spatial data analysis [Deutsch and Journel, 1992]. This problem is related to the spatial representativity of data. We have used spatial declustering procedures for preparing three data sets: training, testing and validation.

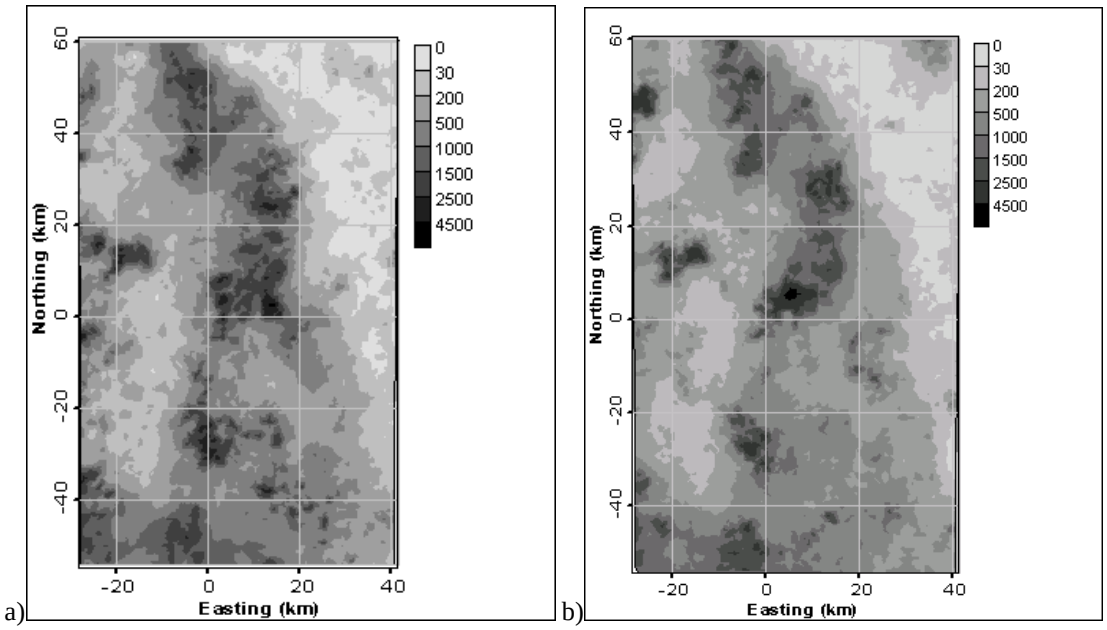

Figure 10. Mapping of Cs137: a) with Neural Network Residual Sequential Gaussian Simulations model (NNRSGS) and b) with Support Vector Regression Residual Sequential Gaussian Simulations model (SVRRSGS)

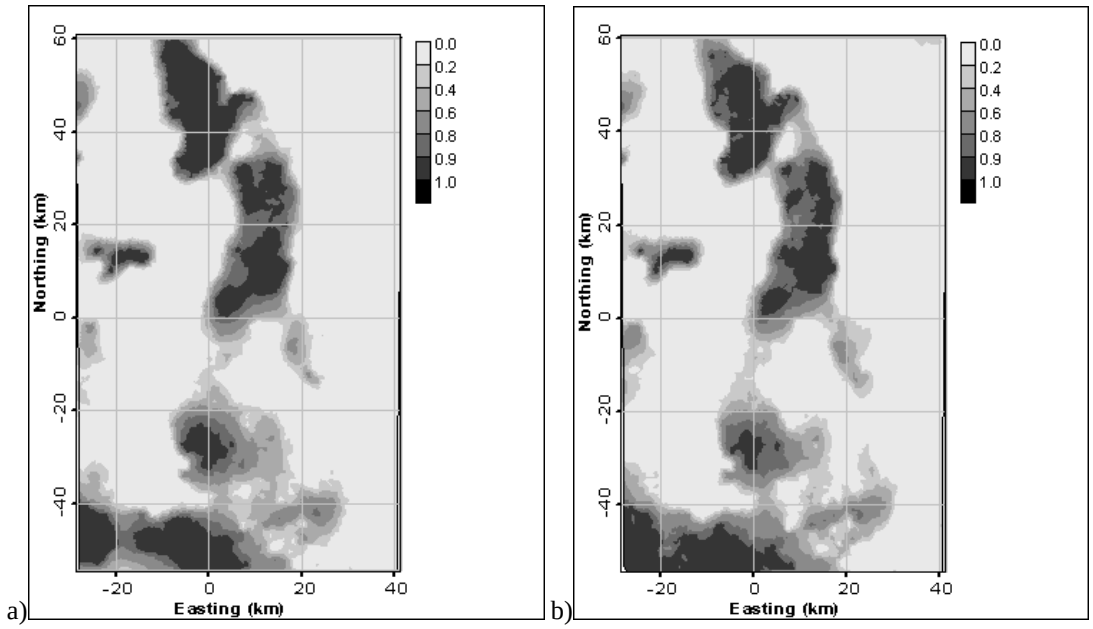

Figure 11. Probability of exceeding level 800 kBq/m^2 for NNRSGS model (a) and for SVRRSGS model (b)

The similarity and dissimilarity between digital models of the reality describes spatial variability and uncertainty. The next step deals with the probabilistic mapping: mapping to be Above some predefined decision level. This is a topic of another research related to decision oriented mapping of contaminated territories. Usually hundreds of simulated models (realizations) are generated. The similarity and dissimilarity between different equiprobable realizations of the reality (using data and available knowledge) describes spatial variability and uncertainty of data. By developing many of equiprobable realizations probabilistic/risk mapping is possible as well: mapping of probability to be above/below some predefined decision/regulation levels (probability of exceeding level 800 kBq/m^2 for Neural Network/Support Vector Regression Residual Sequential Gaussian Simulation models is presented in Figures 11a and 11b respectively). This is important advanced information for real decision-making process.

4 Conclusions

The new non-stationary NNRSGS (neural network residual sequential gaussian simulations model) and SVRRSim models for the analysis and mapping of spatially distributed data have been developed. Non-linear trends in environmental data can be efficiently modelled by the three layer perceptrons. Combinations of MLA and geostatistical models gave rise to decision-oriented risk and probabilistic mapping. The promising results presented are based on an important case study: soil contamination by the most radiologically important Chernobyl radionuclides. Other kinds of ANN models (also local approximators) can be used with possible modifications. The approach seems to be useful in many cases when it is important to model and to remove non-linear trends or large-scale spatial structures. Computational cost of the method is rather cheap for typical geostatistical problems. But application of the method needs deep expert knowledge in geostatistical modelling. Extension of the model to image processing can require improving and adaptation of algorithms, especially from ML side recent developments in ML algorithms implementations, see e.g. www.torch.ch, are promising from the computational point of view

5 Acknowledgements

The work was supported in part by INTAS grant 99-00099 and ISTC 1224

6 References

[1] Cressie, N. *Statistics for Spatial Data*, John Wiley & Sons, New York, 1991.

[2] Deutsch, C.V., and A.G. Journel, GSLIB Geostatistical Software Library and User's Guide, Oxford University Press, New York, Oxford, 1992.

[3] Dowd, P.A., The use of neural networks for spatial simulation, *Geostatistics for the next century*, Ed. R. Dimitrakopoulos, Kluwer Academic Publishers, pp.173-184, 1994.

[4] Gambolati, G., and G. Galeati, Comment on "analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels" by Neuman and Jacobson, *Mathematical Geology*, 19, 249-257, 1987.

[5] Haas, T.C., Multivariate Spatial Prediction in the Presence of Nonlinear Trend and Covariance Nonstationarity. *Environmetrics*, 7, 1996.

[6] Kanevsky, M., R. Arutyunyan, L. Bolshov, V. Demyanov, and M. Maignan, Artificial neural networks and spatial estimations of Chernobyl fallout, *Geoinformatics*, 7, 5-11, 1996.

[7] Kanevski, M., A. Pozdnukhov, S. Canu, M. Maignan, P. Wong, S. Shibli, Support Vector Machines for Classification and Mapping of Reservoir Data, A chapter from "Soft computing for reservoir characterization and modeling", Springer-Verlag, pp. 531-558, 2001.

[8] Neuman, S.P., and E.A. Jacobson, Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels, *Mathematical Geology*, 16, 499-521, 1984.

[9] Vapnik V. Statistical Learning Theory. John Wiley & Sons. 1998.