ARTIFICIAL NEURAL NETWORKS AND SPATIAL ESTIMATION OF CHERNOBYL FALLOUT

Mikhail Kanevsky1) Rafael Arutyunyan1) Leonid Bolshov1) Vasily Demyanov1) Michel Maignan2)
1) Institute of Nuclear Safety, B. Tuiskaya 52, 113191, Moscow, Russia
2) Mathematics/Mineralogy, University of Lausanne, 1015, Lausanne, Switzerland

(Received: 1 November 1995; Accepted 31 January 1996)

Abstract: The present work continues advanced spatial data analysis of surface contamination by radionuclides after severe nuclear accident on Chernobyl NPP. Feedforward neural networks are used for the Cs137 and Sr90 radionuclides prediction mapping and spatial estimations. Neural networks are used to model complex trends over the entire region. Residuals are analyzed with the help of geostatistical approach within the framework of NNRK (neural network residual kriging) model. Another set of data is used to validate obtained results.

Key Words: Radioactively contaminated territories, Artificial neural networks, Residual kriging

INTRODUCTION

On April 26, 1986, a serious accident was happened at the fourth unit of the Chernobyl nuclear power plant in the Ukraine. Radioactive materials were released from the reactor to the atmosphere during ten days following the accident. This resulted in large scale contamination of the environment. Comprehensive spatial data analysis on Chernobyl fallout was started by using different deterministic and statistical approaches, including traditional interpolators, geostatistical estimations, stochastic simulations, fractal interpolations (Kanevsky, 1995; Kanevsky et al., 1995). The present work continues advanced data analysis of Chernobyl fallout with the help of multilayer feedforward neural networks (FFNN) which are a workhorse of neural computation.

In general, artificial neural networks are a collection of a simple computational units (cells, neurons) interlinked by a systems of weighted connections (synaptic connections). The number of units and their connections form the network topology. Recently a few papers have been published on the use of neural networks for spatial data analysis (Dowd, 1994; Rizzo, 1994; Wu, 1993). Different paradigms (neural architectures), learning rules, measures of success or performance were used. It was shown that ANN can be used for the spatial estimations.

It is well known that standard multilayer FFNN with as few as one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function from one finite dimensional space to another with any desired degree of accuracy, provided sufficiently many hidden units are available. In this sense multilayer FFNN are a class of universal approximators (Hornik et al., 1989; White, 1990; Haykin, 1994; Kreinovich, 1991). Unlike statistical estimators, they estimate a function without an explicit mathematical model of how outputs depend on inputs. Neural network are model-free estimators. As a superregression nonlinear model, FFNN can be useful during analysis of data having complex trends over the entire region of interest. It should be noted that unlike most other software systems, the character of a neural network is as much determined by the data in its experience as by the algorithms used to built it.

Between others, there are several important problems and open questions for the present study: development of a methodology for using artificial neural networks within the framework of spatial data analysis; analysis of residuals (ANN mapping is based on some theoretical assumptions); study of possibilities for coestimations of correlated variables, possibility of development of hybrid models (ANN + geostatistics).

The present work deals with several questions among those listed above. One of the most important problem is the analysis of residuals and of their spatial correlation structures. Theory supposes that residuals after ANN estimations have zero mean value and are not correlated. We investigated this suggestion by analyzing spatial correlation structures.

In the present study, backpropagation training algorithm which is a supervised learning algorithm was applied. This algorithm is an iterative gradient algorithm designed to minimize the error measure between the actual output of the neural network and the desired output. We have to optimize nonlinear system consisting of a large number of highly correlated variables. After training with training data set and validation with an independent data set, network can be used for the interpolations.
It should be noted that environmental and ecological data usually have complex trends and are highly variable at different spatial scales. These facts complicate both analysis and interpretation of the results. It is supposed that data can be split on two parts: 
\[ Z(x) = M(x) + e(x), \]
where \( M(x) \) represents large scale variations (trends), and \( e(x) \) represents small scale variations. \( M(x) \) and \( e(x) \) can be treated also as deterministic and stochastic parts, respectively.

There are several possible approaches in case of trends (nonstationarity): universal kriging, residual kriging (Neuman and Jacobson, 1984; Gambolati and Galeati, 1987), moving window regression residual kriging (Haas, 1995), trend surface analysis, science-based approaches (Venkatram, 1988), etc. Each of these methods has its own advantages and disadvantages. In our case science-based estimates of \( M(x) \) have to rely on atmospheric dispersion models. The problems are that there is still uncertainty about accident scenario and details on physical and chemical composition of time dependent source term, wind and rain fields at different scales, etc. Moreover, atmospheric dispersion model's nonlinearity depends on many parameters (wet and dry deposition velocities, boundary layer parametrization, orography, etc.) and measurements are used to estimate/reestimate them. It is not evident that the use of atmospheric dispersion model should lead to the stationary residuals.

The present work is based on a simple idea: if data represent large scale trends over entire region and small scale (possibly correlated) variability, try to estimate nonlinear trends with the help of simple feedforward neural network and then analyze residuals. The approach is similar to the moving window regression residual kriging approach recently developed in Haas (1995), and earlier works (Neuman and Jacobson, 1984; Gambolati and Galeati, 1987). The main difference is that we are modeling nonlinear trends with the help of ANN in one window (entire region). Another important question is how to analyze correlated residuals. In Neuman and Jacobson (1984) and Haas (1995) the stepwise procedure by using generalized least squares regression have been applied. It was shown in Haas (1995) by using cross-validation that bias in this case can be negligible. In Gambolati and Volpi (1979) and Gambolati and Galeati (1987), only one step procedure (modelling of nonlinear trends with the help of ordinary least squares regression and then geostatistical analysis of residuals) leading to non self-consistent model was used. It was shown that although stepwise regression is superior from a strictly theoretical point of view, the results are not more reliable then in one step procedure.

The present study is similar to one step procedure. We used feedforward artificial neural networks which are capable to extract deterministic features hidden in the data and are robust to noise data to model nonlinear trends. Then residuals were analyzed and in case of correlated residuals geostatistics was used.

**MAIN PROCEDURE**

The main stages of the present study are as follows:

1. Preparing Input data and exploratory data analysis that includes looking at the outliers, paying attention to data magnitude and variability, minimizing nonlinearities (the great strength of neural networks is that they work well on nonlinear problems but minimizing nonlinearities leads to faster training, less complicated network, and better performance), etc. The summary of univariate statistics along with histograms both for Cs137 and Sr90 are presented in Figures 1-4.

1a. Preparing training and validation data sets. Data on cumulative surface deposition of Cs137 and Sr90 in the most contaminated part of Briansk region of Russia were prepared and used. It is essential that radionuclides are correlated with correlation coefficient greater than 0.7. Original data were splitted into two data sets: training and validation sets. The last one was used only as an independent ("additional sampling") data set for the validation. Training data set have been used for training. Validation data set is about 14% of the training data set. For the selection of validation data, spatial cell declustering (random selection) was used. As such, validation data more or less represent the investigated region.

1b. Univariate and multivariate statistical analysis. Exploratory variography, analysis of spatial correlation structures. Trend analysis.

2. Designing network architectures. As usually, the multilayered FFNN consists of input layer, output layer and few hidden layers (Figure 5). Two spatial coordinates are inputs of the network and are described by two input neurons. Surface contamination by Cs137 and/or Sr90 are outputs. So, we had both one output neuron describing one radionuclide and two output neurons during "coestimations". Number of hidden layers was never more than two. Multilayer FFNN are powerful tools because they can generate their own internal representation in the hidden units. Number of hidden units can be varied. Choosing an appropriate number of hidden neurons is very important. Using too few will starve the network of the resources it needs 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, that are irrelevant to the general population). In this case it was important to investigate the behaviour of the predictions and residuals with variable number of hidden layers and neurons. We used as
Artificial Neural Networks and Spatial Estimation of Chernobyl Fallout

Fig. 1 South-West Part of Briansk region Cs137. Univariate statistics of training data.

Fig. 2 South-West Part of Briansk region Sr90. Univariate statistics of training data.

Fig. 3 Cs137 - Sr90 correlations between training data.
few hidden neurons as possible to catch large scale structure. FFNN notions: [2-5-0-2] means 2 input neurons, 5 neurons in the first layer, 0 neurons in the second layer, and 2 output neurons.

3. Training of the network. Training data set was used for the supervised learning. In accordance with Masters (1993) a few essential modifications to classical vanilla backpropagation algorithm were implemented: initial weights were selected with the help of genetic optimization 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.

4. Evaluating performance of the network. For the evaluation phase different tools can be used, like jackknife, cross-validation, accuracy test - after learning training data set is used for the predictions. Scatterplots "estimated versus real data" describe how the FFNN captured the correlation between locations and contamination. Results of the accuracy tests for the two networks: 1. one hidden layer with five neurons and two (coestimations) output neurons; 2. one hidden layer with 15 neurons and one Cs137 (estimations) output neuron are presented in Figure 6. It is clear that the networks learned training data sets.

Fig. 4 South-West part of Briansk region, Cs137. Univariate statistics of validation data set.

Fig. 5 Examples of the feedforward neural networks architectures used for spatial estimations and coestimations.

Fig. 6 Accuracy test: FFNN estimates for training data set (after learning).
5. Validation is a process of estimating the FFNN ability to generalize that delivers a correct response to inputs it has never been exposed before. At this phase validation data set was used. Results of validation are presented in Figure 8.

6. Operation phase: prediction mapping, interpolations. Coordinates on a regular grid are presented to the input of network, and mosaic maps of surface contamination are the outputs. We used regular grid with 80x125 nodes in X (Easting) and Y (Northing) directions which means 1x1 squared km. Mapping with neural network consisting of 5 hidden neurons is presented in Figure 9. Both coordinates are in node index.

7. Analysis of the residuals, structural analysis and modelling, kriging. Residuals obtained after learning phase were analyzed with the help of exploratory variography. There are two possibilities: 1) network
was able to learn data and residuals are not correlated (neural network regression model), 2) network was able to catch only large scale structure and residuals are spatially correlated. In the present case study spatial behaviour Sr90 was described with one hidden layer consisting of 5 neurons, residuals are uncorrelated. Spatial behaviour of Cs137 reflects both large scale and small scale variations. General results from variography for the original data and residuals generated by different networks are presented in Figure 7. It was found that unlike original data residuals have shown stationarity and well behaved semivariograms (Figure 7). The semivariogram model used for the kriging of residuals have been modeled with the help of anisotropic structure. An interesting fact is that varigrams of residuals show more or less the same structures. It can be explained by two things: first - robust behaviour of neural network when finding large scale structures and second, in order to learn small scale variations much more powerful (number of hidden neurons) network should be designed, trained, and validated.

8. Prediction mapping. Developed anisotropic variograms were used for the kriging prediction mapping. After kriging predicted residuals were added to the results of predictions with neural network (NNRK - neural network residual kriging). The result of prediction mapping is presented in Figure 10.

9. Validation. NNRK model was also validated by using the independent data set (consisting of N data). Results are presented in Figure 8. Mean squared error: \[ \text{MSE} = \frac{1}{N} \sum (\text{estimates} - \text{measurements})^2 \] for the FFNN was about two times greater then for the NNRK model.

CONCLUSIONS

The work presents advanced spatial data analysis of Chernobyl fallouts with the help of feedforward neural networks. Methodology of using artificial neural networks for the spatial analysis of environmental data along with detailed case study is presented. Special attention was paid to geostatistical analysis of the residuals. It was shown that neural network residual kriging model can be used in case of complex nonlinear trends over the entire region and small scale spatial variability. Validation test based on independent data ("additional sampling") has shown that in terms of MSE NNRK model in the present case study is about two times better then FFNN. Unlike simple neural network interpolations NNRK estimates are conditional. Preliminary results with "coestimations" of correlated variables have shown improved and more stable behaviour at the learning phase and better predictions.

ACKNOWLEDGEMENTS

The work was supported in part (access to data) by the grant INTAS 94 - 2361. The work is based on the report presented at The IAMG 1995 Annual Conference. One of the author (M. Kanevsky) thanks the IAMG for the partial financial support in order to present report at this conference.

REFERENCES


Artificial Neural Networks and Spatial Estimation of Chernobyl Fallout


