Geophysical Research Abstracts
Vol. 15, EGU2013-10786, 2013
EGU General Assembly 2013

Multivariate Analysis and Modeling of Sediment Pollution Using Neural Network Models and Geostatistics

Jean Golay and Mikhaïl Kanevski
Center for Research on Terrestrial Environment, University of Lausanne, Switzerland (jean.golay@unil.ch)

The present research deals with the exploration and modeling of a complex dataset of 200 measurement points of sediment pollution by heavy metals in Lake Geneva. The fundamental idea was to use multivariate Artificial Neural Networks (ANN) along with geostatistical models and tools in order to improve the accuracy and the interpretability of data modeling. The results obtained with ANN were compared to those of traditional geostatistical algorithms like ordinary (co)kriging and (co)kriging with an external drift.

Exploratory data analysis highlighted a great variety of relationships (i.e. linear, non-linear, independence) between the 11 variables of the dataset (i.e. Cadmium, Mercury, Zinc, Copper, Titanium, Chromium, Vanadium and Nickel as well as the spatial coordinates of the measurement points and their depth). Then, exploratory spatial data analysis (i.e. anisotropic variography, local spatial correlations and moving window statistics) was carried out. It was shown that the different phenomena to be modeled were characterized by high spatial anisotropies, complex spatial correlation structures and heteroscedasticity. A feature selection procedure based on General Regression Neural Networks (GRNN) was also applied to create subsets of variables enabling to improve the predictions during the modeling phase.

The basic modeling was conducted using a Multilayer Perceptron (MLP) which is a workhorse of ANN. MLP models are robust and highly flexible tools which can incorporate in a nonlinear manner different kind of high-dimensional information. In the present research, the input layer was made of either two (spatial coordinates) or three neurons (when depth as auxiliary information could possibly capture an underlying trend) and the output layer was composed of one (univariate MLP) to eight neurons corresponding to the heavy metals of the dataset (multivariate MLP). MLP models with three input neurons can be referred to as Artificial Neural Networks with EXternal drift (ANNEX). Moreover, the exact number of output neurons and the selection of the corresponding variables were based on the subsets created during the exploratory phase. Concerning hidden layers, no restriction were made and multiple architectures were tested. For each MLP model, the quality of the modeling procedure was assessed by variograms: if the variogram of the residuals demonstrates pure nugget effect and if the level of the nugget exactly corresponds to the nugget value of the theoretical variogram of the corresponding variable, all the structured information has been correctly extracted without overfitting. Finally, it is worth mentioning that simple MLP models are not always able to remove all the spatial correlation structure from the data. In that case, Neural Network Residual Kriging (NNRK) can be carried out and risk assessment can be conducted with Neural Network Residual Simulations (NNRS).

Finally, the results of the ANNEX models were compared to those of ordinary (co)kriging and (co)kriging with an external drift. It was shown that the ANNEX models performed better than traditional geostatistical algorithms when the relationship between the variable of interest and the auxiliary predictor was not linear.

References

Kanevski, M. and Maignan, M. (2004). Analysis and Modelling of Spatial Environmental Data. Lausanne: EPFL Press.