Geophysical Research Abstracts, Vol. 9, 03031, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-03031

Machine learning and geostatistics for multivariate soil contamination mapping

M. Kanevski (1), A. Pozdnoukhov (1), V. Timonin (1), M. Maignan (2)

(1) Institute of Geomatics and Analysis of Risk, University of Lausanne
(Mikhail.Kanevski@unil.ch/+41 21 692 3535)

The research presents overview of approaches and methods recently developed for the analysis and modelling of environmental and pollution data. The main attention is paid to the application of machine learning algorithms (artificial neural networks of different types, support vector machines) and geostatistical simulations for decision-oriented and prediction mapping of soil contamination by heavy metals and radionuclides. Both approaches have some advantages and drawbacks. In particular, geostatistical modelling is a model-based approach and demands analysis and modelling of spatial correlation structures (variography) and is based on statistical hypotheses of spatial stationarity. Machine learning algorithms, being nonlinear data-driven tools highly depends on the quality and quantity of data and their results are not easy for the interpretation. But they are excellent exploratory data analysis tools capable to reveal hidden patterns and structures, detect outliers, etc. In the present paper coherent and consistent analysis based on real case studies on soil pollution by heavy metals and radionuclides using machine learning algorithms and geostatistics is presented. The considered problems include regional classification, multivariate and multiscale spatial regression and probabilistic/risk mapping. The efficiency and quality of the analysis, mapping and spatial predictions are controlled by validation data sets. Hybrid models using both machine learning and geostatistics are considered as a promising way of spatial data analysis. Geographical Information Systems are used to produce decision-oriented maps.