Geophysical Research Abstracts
Vol. 13, EGU2011-9359, 2011
EGU General Assembly 2011

Spatio-Temporal Data Analysis Using Geostatistics and Machine Learning

Mikhail Kanevski
Institute of Geomatics and Analysis of Risk, Lausanne, Switzerland (Mikhail.Kanevski@unil.ch)

The research presents a review of general problems concerning spatio-temporal data analysis (STDA), modelling and visualisation using geostatistics (spatio-temporal predictions and conditional simulations) and machine learning algorithms (artificial neural networks of different architectures, support vector machines). There are several important factors that complicates the STDA: the environmental spatio-temporal data usually are non-stationary in space and/or in time; have high variability at several spatial and temporal scales; are collected on clustered monitoring networks; in general, should be considered in high dimensional geo-feature spaces, constructed from geographical coordinates and additional features, e.g. generated from digital elevation models [1, 2]. The analysis and treatment of an increased number of environmental data produced by numerous monitoring networks (including automatic ones), remote sensing/satellite data, etc. is a challenging task of data mining. Moreover, in many cases, there are science-based models (e.g., hydrological, geophysical, meteorological, pollution diffusion) which have to be integrated during the analysis.

The STDA methodology presented in this study includes several important phases: construction of geo-feature input (independent variables) and multivariate output (dependent variables) spaces; comprehensive exploratory data analysis; analysis and optimization of monitoring networks; features selection/feature extraction; detection and modelling of spatial structures/patterns; spatio-temporal predictions/simulations; testing and validation of the results; quantification of the uncertainties; visualisation of high-dimensional data and decision-oriented mapping. The advantages and drawbacks of each approach at different steps of the analysis are considered. The emphasis is put on a complementary nature and use of geostatistics (model-based) and machine learning (data-driven) algorithms to solve a wide range of environmental tasks. Finally, the hybrid models (geostatistics + machine learning), which can help to improve the interpretability of the results and to overcome some problems with stationaritires, are revised.

The main concepts and models are illustrated by using both simulated and real life data from the pollution of the environment, topo-climatic modelling, and automatic mapping.

Acknowledgements. The research was partly supported by Swiss NSF grants N 200021-126505 and N 200020-121835

References

1.Kanevski M. (Editor). Advanced Mapping of Environmental Data. Geostatistics, Machine Learning and Bayesian Maximum Entropy. iSTE and Wiley, 2008.
2. Kanevski M., Pozdnoukhov A. and Timonin V. Machine Learning for Spatial Environmental Data. Theory, Applications and Software. EPFL Press, Lausanne, 2009.