Geophysical Research Abstracts, Vol. 9, 01307, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-01307

Automatic Regional Classification of Environmental Data

R. Tapia (1), V. Timonin (1), A. Pozdnukhov (1), M. Kanevski (1), M. Gruson (2)
(1) IGAR, UNI Lausanne (2) OFSP, Bern

Predictive mapping of continuous variables is a well-established field in spatial data modeling. A variety of methods exists for the task: deterministic interpolators, geostatistical approaches, non-parametric neural network and machine learning methods. However, considerably less attention was paid to the mapping of categorical variables, or classification; and particularly to the application of spatial classification methods for risk mapping. Given a continuous variable and a set of decision thresholds, the problem of decision-oriented risk mapping can be considered as a regional classification one. Then, spatial areas where the concentration\activity of some contaminant exceeds the predefined decision threshold have to be determined.

A common geostatistical method for predictions over ranked data is indicator kriging. It is based on variogram modeling and may fail if the number of instances of some class is not large enough; this is a common situation for high value thresholds and hot spots. Machine learning classification methods are more robust in this sense. Moreover, in the context of automatic mapping; they have the advantage of not requiring expert guided variogram modeling and they could provide an easy way for tuning their intrinsic parameters, using cross-validation error minimization.

In this paper, Support Vector Machine and Probabilistic Neural Network are considered. The well-known deterministic classifier, K-Nearest Neighbor is considered as a benchmark model, and also indicator kriging is applied along for comparison. These methods of predictions were tested and validated using very complex (high local and regional variability, important spatial clustering) Swiss indoor radon data. Raw data set was split into training and testing in order to verify and validate models applied. The problem was considered as a problem of regional decision-oriented classification.