Geophysical Research Abstracts, Vol. 7, 04654, 2005
SRef-ID: 1607-7962/gra/EGU05-A-04654

Non-linear visualization and trend analysis of multivariate data sets using Self-Organizing Maps

G. Lischeid

University of Bayreuth, Germany, email: gunnar.lischeid@bitoek.uni-bayreuth.de

Hydrological monitoring programs yield large data sets that are likely to reflect a variety of hydrological and biogeochemical processes. However, rarely methods are applied that allow to identify unexpected structures in the data sets, e.g., for early warning or for formulating new hypotheses. These methods should be able to detect non-linear relationships and local patterns, e.g., correlations that hold only for a certain subset of the data. Visualization of the data could be an efficient first step for analyzing large data sets and in order to organize further analyses as it makes efficient use of the human brain's capability of image analysis. Here, a data set was analyzed that consisted of time series of concentration of 13 solutes from 40 stream and groundwater sampling sites with about 4000 samples in total. Self-Organizing Maps were applied to visualize the data, and to investigate spatial patterns and temporal trends. This type of artificial neural networks projects high-dimensional data into a low-dimensional space in a non-linear way, aiming at preserving the similarity between the data vectors. Compared to linear methods of dimensionality reduction, the Self-Organizing Maps reproduced a considerably higher fraction of variance. Clusters and outliers could be easily identified. Even subtle long-term trends at some sampling sites could be determined and differentiated from short-term fluctuations. This provided substantial insight into long-term changes of the dynamics of stream water and groundwater quality. In addition, contamination of one of the groundwater wells could easily be traced through space and time. Last but not least, the Self-Organizing Map analysis helped to optimize the sampling design.