Geophysical Research Abstracts
Vol. 16, EGU2014-2934, 2014
EGU General Assembly 2014

Multivariate, non-linear trend analysis of heterogeneous water quality monitoring data

Gunnar Lischeid (1,2), Thomas Kalettka (1), Jörg Steidl (1), Christoph Merz (1), Christian Lehr (1,2)

(1) Leibniz Centre for Agricultural Landscape Research, Institute of Landscape Hydrology, Müncheberg, Germany, (2) University of Potsdam, Institute for Earth and Environmental Sciences, Potsdam, Germany

Comprehensive water quality monitoring is considered a necessary prerequisite for sound water resources management and a valuable source for science. In practice, however, use of large monitoring data sets is often limited due to heterogeneous data sources, spatially and temporally variable monitoring schemes, non-equidistant sampling, large natural variability, and, last but not least, by the sheer size of the data sets that makes identification of unexpected peculiarities a tedious task. As a consequence, any initiation of gradual long-term system shifts can hardly be detected, especially as long as it is restricted to a small fraction of sampling sites. In addition, trends might be limited to a rather small subset of sampling sites or to certain periods of time and might thus escape attention.

Usually, numerous solutes are monitored in parallel, but trend analyses are performed for each solute separately. However, in water quality samples trends are hardly restricted to single solutes, but affect various solutes synchronously in a characteristic way. Thus performing joint multivariate trend analyses would not only save effort and time, but would yield more robust assessments of system shifts.

We present a non-linear multivariate data visualization approach that allows a rapid assessment of non-linear, possibly local trends and unexpected behaviour in large water quality monitoring data sets. It consists of a combination of Self-Organizing Maps and Sammon's Mapping (SOM-SM). The approach was applied to a data set of 2900 water samples, each comprising 13 solutes, compiled from various monitoring programs in the Federal State of Brandenburg (Germany). In total, 128 stream water, groundwater and small pond sites had been sampled between 1994 and 2012 at different and irregular time intervals.

The SOM-SM product is a graph where every sample is represented by a symbol. Location of the symbols in the graph is optimized such that the distance between any two symbols in the graph is proportional to the dissimilarity of the two respective water samples with respect to all 13 solutes. In our study, the non-linear 2D projection of the SOM-SM reflected 75% of the variance of the 13D data set. For further analyses the same graph was used again and again, where different colouring revealed different information. Thus the user rapidly became acquainted with the large, high dimensional data set.

At a first glance outliers easily could be identified as well as clusters of samples with similar solute concentration. Different groups of samples were analysed for the degree of overlap. Multivariate trend analysis was performed that did not only account for increasing or decreasing concentration of single solutes but for systemic shifts of characteristic solute concentration patterns as well. Partly converging trends were found, that is, sampling sites becoming more similar to each other. In addition, long-term decreasing variance was found at some sites. For checking for significant differences between different time periods confidence intervals were included in the graph. We conclude that the SOM-SM proved to be a powerful and extremely helpful tool for analysis of this large, heterogeneous water quality data set.