Geophysical Research Abstracts, Vol. 7, 03441, 2005
SRef-ID: 1607-7962/gra/EGU05-A-03441

First results in the attempt to adept a Data Mining Strategy for distinguishing global, regional and local factors reflecting variation patterns in river water chemistry

J. Hartmann (1), T. Fernandez-Steeger (2), S. Kempe (1)

(1) Institute of Applied Geosciences, Darmstadt University of Technology, Darmstadt, Germany, (2) Chair of Engineering Geology and Hydrogeology (LIH), Aachen University of Technology (RWTH), Aachen, Germany (hartmann@geo.tu-darmstadt.de / Fax: +49 6151 16 6539)

This study tests the feasibility of a data mining approach to reveal statistical significant global, regional and local factors, representing changes in the composition of river water in an anthropogenic altered river basin. For this purpose, the standard technique of Principal Component Analysis (PCA) was applied. More than one decade of biweekly data from the environmental monitoring programme of the Neckar river system, a major tributary of the river Rhine, were used for this study. The environmental monitoring data included biogeochemical relevant parameters such as dissolved major ions, nitrate, silica, as well as partial pressure of CO_2 (PCO_2), pH, O_2 and temperature.

PCA was run for the entire data set and for subsets of mainstream stations and headwater stations as well as for all of the individual sites. When running the entire data set, artificial noise is produced. Comparing the outcome of different groups of subsets and the global data set, stable factors for each set can be distinguished. Nevertheless, three significant global factors can be extracted, representing: (i) a major ion component, including nitrate; (ii) a silica-magnesium-carbon-dioxide component, reflecting weathering; and (iii) a seasonal component, representing temperature and dissolved oxygen. In the main stem subset four factors can be extracted: (i) representing the main ions; (ii) the seasonal changes in O_2 and temperature but also in PCO_2; (iii) an anaerobic factor including ammonia and nitrite (negative correlated with tempera-

ture, but not correlating with oxygen or PCO_2), and (iv) a silica and iron factor. The headwater stations yielded also four factors, however slightly different: (i) the main ion factor, (ii) the seasonal O_2 and temperature factor (not including PCO_2), (iii) a nitrate-PCO_2-factor (anticorrelating); and (iv) a Na-Cl factor, which probably represents atmospheric precipitation.

It is not surprising that the main ions dominate the factor scores, nevertheless we find some unexpected results:

(i) the missing correlations between temperature and oxygen with PCO_2 in the headwaters. Instead a negative correlation of PCO_2 with silica and nitrate was identified.

(ii) the high correlation of silica and iron in the main stem (not enough data for the tributaries were available to confirm this for the whole data set).

However the presented analysis approach is suitable to conduct fast pre-analysis of large of data sets, helping to distinguish significant differences in the factors representing the variation of the hydrochemical matrix.