Tracing biogeochemical processes in stream water and groundwater using non-linear statistics

Gunnar Lischeid [a,*], Jochen Bittersohl [b]

[a] *Chair for Ecological Modelling, BayCEER, University of Bayreuth, Dr.-Hans-Frisch-Strasse 1-3, D-95440 Bayreuth, Germany*
[b] *Bavarian State Agency for the Environment, Hans-Högn-Straße 12, D-95030 Hof, Germany*

Summary   Stream water and groundwater solute concentration are subject to a multitude of biogeochemical processes that act at different scales and are often characterized by non-linear relationships and feedback loops. Different multivariate statistical methods were applied to investigate the interplay of different processes. The data set from the Lehstenbach catchment in South Germany comprised 2641 stream water and groundwater samples from 38 different sites in the catchment, where 13 different solutes had been determined. According to the correlation dimension analysis, the number of dominant processes was four. The first four components determined via principal component analysis comprised 88% of the total variance, whereas the non-linear isometric feature mapping explained 92% with the first four components. These components were ascribed to prevailing biogeochemical processes and were used to investigate spatial and temporal patterns. Redox processes and contamination by road salt explained 35% of the variance each. Another 13% were ascribed to near-surface runoff in the acidified topsoil, and 9% to the impact of contaminated filter gravel in some of the groundwater wells.

The redox component exhibited clear seasonal patterns at most stream water and groundwater sampling sites, with the most reduced conditions in late summer, immediately before the onset of re-wetting. There was clear evidence that redox processes, especially denitrification, play an important role even in the oxic aquifer. During discharge peaks, stream water exhibited higher values of the near-surface runoff component. However, the associated lower values of the redox component pointed to near-surface runoff in the riparian wetlands as the predominating runoff generation process rather than to a contribution of upslope soil water. A series of major rain storms in fall 1998 altered groundwater and stream water solute concentration for months: stream water and groundwater became more oxic and more acidified, and showed a higher impact of road salt contamination at some sites.

* Corresponding author. Present address: Leibniz Centre for Agricultural Landscape Research, Institute of Landscape Hydrology, Eberswalder Str. 84, D-15374 Müncheberg, Germany. Tel.: +49 921 555632; fax: +49 921 555799.
  E-mail address: gunnar.lischeid@zalf.de (G. Lischeid).

Overall, the spatial and temporal patterns of the predominating components gave a consistent picture and helped considerably to better understand the interplay between biogeochemical and hydrological processes. The chosen approach is a promising tool for investigating apparently complex hydrological and biogeochemical systems.

Introduction

A thorough understanding of solute transport and solute turnover in the subsurface is a necessary prerequisite for assessing the impact of anthropogenic pressure as well as of climate change or other natural processes on stream water and groundwater quality. E.g., the EU water framework directive demands wide-ranging assessment of water quality, detection of harmful anthropogenic impacts and to take measures to improve water quality, if necessary. However, solute concentration and solute flux in the subsurface is determined by a multitude of processes that act at different scales and are often interlinked in a highly non-linear way. Thus, anthropogenic impacts often cannot clearly be differentiated from the effects of natural processes due to, e.g., climatic variability. In spite of many process studies a thorough understanding of total system behavior remains a challenge for hydrological and biogeochemical sciences.

One common way to assess system behavior is a bottom-up modeling approach. However, this approach suffers from the fact that the necessary selection of a limited number of processes and the feedback between different processes have to be defined in an ad-hoc manner and can rarely be rigorously tested (Kirchner, 2006; Hooper, 2001; Beven, 2001). Thus, there is urgent need for methods that help to identify prevailing processes directly at the scale of interest.

It has been argued that although many processes interact, often only a very limited number of processes predominate, depending on the scale of observation (Seyfried and Wilcox, 1995; Sivakumar, 2004a). As a consequence, only a limited number of processes need to be considered at given sites although many more might have been proven in small scale studies. On the other hand, this implies that the potential for model calibration or model testing, respectively, is limited.

According to the chaos and fractal theory, the degrees of freedom of a dynamic system can be assessed for sufficient long time series or large data sets and a sufficiently high signal to noise ratio (Moon, 1992). This number is equal to the number of independent processes that affect the behavior of a system, which is often called the "intrinsic dimensionality" of a given data set. To that end, the different processes do not need to be known. This holds even for cause and effect in strongly non-linear relationships. This information can be used to compare different data sets as well as to confront the complexity of a model with the intrinsic dimensionality of the data that are used for calibration and validation of the model. Moreover, the intrinsic dimensionality can be regarded as a lower bound of the number of axes that are required for an adequate low-dimensional projection of a high-dimensional data set.

The identification of processes can be performed by applying ordination methods. One of the best known approaches in hydrology is the principal component analysis. It is a common tool for analyzing large data sets of groundwater (e.g., Fernandes et al., 2006; O'Shea and Jankowski, 2006; Lamouroux and Hani, 2006; Cruz and Amaral, 2004) or river water quality (e.g., Zeilhofer et al., 2006; Castañné et al., 2006; Singh et al., 2005; Ying, 2005; Haag and Westrich, 2002). The objective of these studies was usually to delineate different natural and anthropogenic processes that affect water quality, to determine mixing ratios, or to develop a perceptual model of the groundwater flow dynamics (e.g., Lamouroux and Hani, 2006; Thyne et al., 2004). In some studies, long-term trends of single principal components were investigated (e.g., Haag and Westrich, 2002).

It is remarkable that there are rather few geochemical studies that link groundwater and surface water bodies, and consider temporal as well as spatial patterns. The former has been done, e.g., by Thyne et al. (2004). However, in that study, the data set was subdivided into a baseflow data set and a spring runoff data set, limiting the potential of investigating the interplay between geochemical and hydrological processes.

On the other hand, most hydrological catchment studies focus on hydrological processes at the cost of biogeochemical processes. To that end, principal component analysis is often applied either directly to the precipitation, soil water, groundwater, and stream water quality data (e.g., Worrall et al., 2003) or within the end-member mixing analysis framework (Christophersen and Hooper, 1992; Burns et al., 2001; James and Roulet, 2006). This approach aims at determining different components that contribute to stormflow runoff generation in order to investigate the runoff generation process. To that end, usually only those solutes are considered that are assumed to behave conservatively at the scale of observation (Lischeid, 2008). This approach is well established in hydrology, but seems to have reached its limits now (Burns, 2002).

In contrast to that approach, however, the intention of this study is not to identify runoff components, but to identify the prevailing biogeochemical processes in general. Thus, it is not restricted to conservative tracers, but explicitly accounts for reactive solutes as well. However, then non-linear processes have to be accounted for. Moreover, often local effects are observed, i.e., relationships that hold only for a subsample of the total data set. For example, an impact of intense agricultural land use on groundwater nitrate concentration is likely to be seen only in oxic aquifers. In those cases, linear methods like the principal component analysis can map non-linear relationships only by piecewise approximation, spreading the variance across