Geophysical Research Abstracts
Vol. 15, EGU2013-4635, 2013
EGU General Assembly 2013

Discovering temporal patterns in water quality time series, focusing on floods with the LDA method

Alice Hélène Aubert (1), Romain Tavenard (2), Rémi Emonet (2), Simon Malinowski (3), Thomas Guyet (3), René Quiniou (4), Jean-Marc Odobez (2), and Chantal Gascuel-Odoux (1)

(1) Inra, UMR 1069, Soil Agro and hydroSystem, Rennes, France (alice.aubert@rennes.inra.fr), (2) Idiap Research Institute, Martigny, Switzerland, (3) Agrocampus Ouest, UMR 6074, IRISA, Rennes, France, (4) Inria, UMR 6074, IRISA, Rennes, France

Studying floods has been a major issue in hydrological research for years. It is often done in terms of water quantity but it is also of interest in terms of water quality. Stream chemistry is a mix of solutes. They originate from various sources in the catchment, reach the stream by various flow pathways and are transformed by biogeochemical reactions at different locations. Therefore, we hypothesized that reaction of the stream chemistry to a rainfall event is not unique but varies according to the season (1), and the global meteorological conditions of the year (2). Identifying a typology of temporal chemical patterns of reaction to a rainfall event is a way to better understand catchment processes at the flood time scale.

To answer this issue, we applied a probabilistic model (Latent Dirichlet Allocation or LDA (3)) mining recurrent sequential patterns to a dataset of floods. The dataset is 12 years long and daily recorded. It gathers a broad range of parameters from which we selected rainfall, discharge, water table depth, temperature as well as nitrate, dissolved organic carbon, sulphate and chloride concentrations. It comes from a long-term hydrological observatory (AgrHys, western France) located at Kervidy-Naizin. A set of 472 floods was automatically extracted (4). From each flood, a document has been generated that is made of a set of "hydrological words". Each hydrological word corresponds to a measurement: it is a triplet made of the considered variable, the time at which the measurement is made (relative to the beginning of the flood), and its magnitude (that can be low, medium or high). The documents are used as input data to the LDA algorithm. LDA relies on spotting co-occurrences (as an alternative to the more traditional study of correlation) between words that appear within the flood documents. It has two nice properties that are its ability to easily deal with missing data and its additive property that allows a document to be seen as a mixture of several flood patterns. The output of LDA is a set of patterns that can easily be represented in graphics. These patterns correspond to typical reactions to rainfall events. The patterns themselves are carefully studied, as well as their repartition along the year and along the 12 years of the dataset.

The novelties are fourfold. First, as a methodological point of view, we learn that hydrological data can be analyzed with this LDA model giving a typology of a multivariate chemical signature of floods. Second, we outline that chemistry parameters are sufficient to obtain meaningful patterns. There is no need to include hydro-meteorological parameters to define the patterns. However, hydro-meteorological parameters are useful to understand the processes leading to these patterns. Third, our hypothesis of seasonal specific reaction to rainfall is verified, moreover detailed; so is our hypothesis of different reactions to rainfall for years with different hydro-meteorological conditions. Fourth, this method allows the consideration of overlapping floods that are usually not studied. We would recommend the use of such model to study chemical reactions of stream after rainfall events, or more broadly after any hydrological events. The typology that has been provided by this method is a kind of bar code of water chemistry during floods. It could be well suited to compare different geographical locations by using the same patterns and analysing the resulting different pattern distributions.

(1) Aubert, A.H. et al., 2012. The chemical signature of a livestock farming catchment: synthesis from a high-frequency multi-element long term monitoring. HESSD, 9(8): 9715 - 9741.
(2) Aubert, A.H., Gascuel-Odoux, C., Merot, P., 2013. Annual hysteresis of water quality: A method to analyse the effect of intra- and inter-annual climatic conditions. Journal of Hydrology, 478(0): 29-39.
(3) Blei, D. M.; Ng, A. Y.; Jordan, M. I., 2003. Latent Dirichlet allocation. Journal of Machine Learning Research, 3(4–5): 993–1022.
(4) de Lavenne, A., Cudennec, C., Streamflow velocity estimation in GIUH-type approach: what can neighbouring

basins tell us? Poster Presentation - EGU General Assembly, 22-27 April 2012, Vienna, Austria.