



Using data mining techniques for the analysis of gridded global hydrological data

Gerald A. Corzo P., Marjolein H.J. van Huijgevoort, and Henny A.J. van Lanen

Hydrology and Quantitative Water Management Group, Centre for Water and Climate, Wageningen University, the Netherlands (gerald.corzoperez@wur.nl)

Global change encouraged the research community to a relative recent development of a new generation of hydrological models that simulate the global water cycle, incl. hydrological extremes. These new, so-called Global Hydrological Models (GHMs) or Land Surface Hydrological Models (LSHMs) require and generate considerable amount of data for each hydrometeorological variable generated by the model. The analysis of this type of results becomes a challenge not only in the data management, but also in the analysis of multiple variables that also considers the geo-spatial representation of the system. This work presents partial results of using data-mining techniques (e.g. clustering, classification and regression) for the whole globe. The research aims at the data mining extraction procedures to help understanding historical droughts, in particular spatio-temporal characteristics. The MySQL engine is used to setup a database and to organize the forcing gridded information compiled through the EU-FP6 WATCH project (0.5°, daily, 1958-2001). Clustering of spatial patterns of k-means is used to generate a graphical representation of spatially-distributed hydrometeorological variables. This is done for overall and seasonal statistical information. The patterns found in the clusters of different continents are compared in terms of area and drought intensity. Based on these methodologies some principles are taken to develop methods to evaluate meteorological and subsequently hydrological drought at the global scale using different types of large-scale grid-based models (e.g. RCMs, LSHMs, GHMs).