Geophysical Research Abstracts
Vol. 13, EGU2011-4790, 2011
EGU General Assembly 2011

Predictive modeling of global and regional runoff to the Baltic Sea using machine learning methods

Federico Montesino Pouzols

Environmental and Industrial Machine Learning Group, Dept. of Information and Computer Science, Aalto University, Espoo, Finland (federico.pouzols@tkk.fi)

Climate change and environmental issues in the Baltic Sea as well as their socio-economic implications have a profound and direct impact on 9 European countries and more than 85 million inhabitants. The Baltic Sea is highly influenced by freshwater runoff. An average yearly river runoff contribution of 14151 m^3 has been estimated for the last century, excluding the Kattegat and Danish Sounds. Hence, the total volume of water, 21200 km^3 would be replaced in 48 years if no other factors are considered.

Understanding runoff mechanisms in the Baltic basin is specially relevant since it has a crucial impact on aspects such as salinity and nutrient concentration and can induce deep ecological changes. In this work, 2 datasets of monthly values of freshwater discharges to the Baltic Sea are studied: 1) values reconstructed from 1921, and 2) monthly measurements starting on 1970. These series cover 7 sub-basins defined following the HELCOM (Helsinki Commission, Baltic Marine Environment Protection Commission) Baltic sub-drainage basins. The data are provided by the Swedish Meteorological and Hydrological Institute.

Several specific sources of nonstationarity and nonlinearity as well as difficulties for predictive modeling have been identified in these time series, such as river regulation, hydropower developments, the presence of large lakes in the region and a considerable variability of snow cover periods. Here we adopt a data-driven approach to developing short- and long-term models for global and regional time series. Different machine learning techniques are applied within three modeling approaches: a) purely autoregressive, b) models with only exogenous predictors or transfer functions (integrating climate related indexes as exogenous inputs), and c) autoregressive exogenous models. Results are compared in terms of computational cost and accuracy considering different performance measures.

Diverse machine learning techniques and specific time series prediction methodologies are analyzed, including Gaussian processes, neuro-fuzzy methodologies for time series prediction, support vector regression, double-regularized linear models and adaptive kernel smoothing regression. Among different climate indexes, the Baltic Sea Index (BSI) is found to lead to more accurate results and more parsimonious models. A clear geographical relation is found between climate indexes and the different sub-basins analyzed.

A number of general conclusions can be drawn from the analyses in this work: 1) satisfactory performance can be attained using purely autoregressive models provided proper feature selection procedures are followed. This is possible for prediction horizons of up to 24 months and even greater. 2) Climate indexes as exogenous inputs can improve results, confirming previous works that have established the link between some climate indexes and runoff in the Baltic region using transfer function models. However, these are not enough to obtain satisfactory results in general as compared with autoregressive models. 3) Significant variations for sub-basins can be observed, and the results suggest the use of tailored, regional climate indexes, and 4) Ensemble models combining different machine learning models can further improve the prediction accuracy and consistently show a more robust behavior.