A dimensionality reduction-feature selection approach to streamflow reconstruction using dendrochronological data

Marco Obertelli, Alessandro Amaranto, and Andrea Castelletti
Politecnico di Milano, DEIB - Department of Electronics, Information and Bioengineering, Italy (marco.obertelli@mail.polimi.it)

A dimensionality reduction-feature selection approach to streamflow reconstruction using dendrochronological data

Marco Obertelli\textsuperscript{1}, Alessandro Amaranto\textsuperscript{1}, Andrea Castelletti\textsuperscript{1}

\textsuperscript{1}Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy

Instrumental gauge records are the basis for understanding streamflow variability and for informing water management plans accordingly. Yet, these records seldom exceed 100 years and might thus be insufficient for capturing the whole range of streamflow variability for a given river section. Tree ring width from climatically sensitive trees provides a means for developing long-duration chronologies that extend beyond instrumental recording. Considering their dependence on temperature and surface water availability in a specific year, dendrodata have been recognized in the literature as representative climate proxies; and have been therefore widely used in recent years for reconstructing chronologies of hydrological variables as precipitation frequency, drought severity and streamflow variability. Paleoclimate reconstructions are usually carried by combining a dimensionality reduction-feature selection technique with linear regression methods. The goal of this study is to better understand long-term hydrological variability in the Rhine and the Po river basins by reconstructing for the first time their streamflow trajectories from paleoclimatic data. We apply Principal Component Analysis (PCA) for dimensionality reduction and Multiple Linear Regression (MLR) as a reconstruction model. Palmer Drought Severity Index (PDSI) trajectories, inferred from tree-rings chronologies are employed as paleoclimate proxy. Numerical results show a good accuracy in the reconstruction approach, especially in the Rhine basin (average $R^2 = 0.60$). The accuracy decreases in the Po basin, probably due to the Alpine hydrologic regimes which includes complex nonlinear phenomena (e.g. solid precipitation and snowmelt) not fully described by the PDSI drought index. Historical evidence of the reconstructed 1817 drought in the Po river basin has been found as a proof of the reliability of this approach. The variety of the morphological and hydrological characteristics reflected in the two river basins considered in this study, allows to explore how some of their peculiarities reflect on the streamflow reconstruction, increasing the possibility to replicate the approach in other areas of the world.