Comparison of machine learning methods for data infilling in hydrological forecasting

Juan Carlos Chacon Hurtado, Leonardo Alfonso, and Dimitri Solomatine

UNESCO-IHE Institute for water education, Integrated Water Systems and Governance, Delft, Netherlands
(j.chacon@unesco-ihe.org)

The continuous measurement of hydrological variables requires sensors that must be deployed in the field, increasing the risk of failure due to natural or anthropic conditions inherent to its deployment. The failure of these sensors will interrupt the data stream, which in operational hydrological systems might lead to unsatisfactory performance of forecasting models, or biases due to lack of information in simulation models. To mitigate this, various techniques to fill these missing values can be used, varying from simple regression techniques, to more complex machine learning methods. This research aims at exploring the performance of the latter, considering particular properties of the measurements, length of the missing data series and particular properties of the missing variable. This study is carried out in two different European catchments which differ in geographical conditions, mechanisms and monitoring frequency.