



Leveraging climate data at different spatial scales via machine learning to improve sub-seasonal drought predictions

Matteo Giuliani¹, Francesco Bosso¹, Claudia Bertini², Dimitri Solomatine^{2,3}, and Schalk Jan van Andel²

¹Politecnico di Milano, Politecnico di Milano, Dept. Electronics, Information, and Bioengineering, Milano, Italy

²IHE Delft Institute for Water Education, Hydroinformatics and Socio-Technical Innovation Department, Delft, Netherlands

³Delft University of Technology, Water resources section, Delft, Netherlands

Droughts are one of the most dangerous natural hazards affecting today societies, with an economic impact amounting to over 9 billion euros per year in Europe. Drought events usually originate from a precipitation deficit, which can then cause water shortages, agricultural losses, and environmental degradation. Despite the numerous efforts and recent advances in extreme events forecasting, the sub-seasonal time scale still represents a challenging lead time for state-of-the-art hydroclimatic predictions. In this case, the reference period is short enough for the atmosphere to retain a memory of its initial conditions, but also long enough for oceanic variability to affect atmospheric circulation. However, the relative contribution of climate teleconnections and local atmospheric conditions to the genesis of total precipitation at sub-seasonal scale remains unclear.

In this work, we aim to address this gap by using Machine Learning (ML) to combine the information extracted from teleconnection patterns, global climate variables, and local atmospheric conditions to produce sub-seasonal drought forecasts. Specifically, we implemented a first ML pipeline that uses correlation maps to select relevant grids of global Sea Surface Temperature, Mean Sea Level Pressure, and geopotential height at 500 hPa from the ERA5 reanalysis dataset, which are spatially aggregated via Principal Component Analysis and combined with a set of local variables in the considered region. The second ML approach extends our analysis by explicitly considering the potential role of teleconnection patterns, including North Atlantic Oscillation, Scandinavian oscillation, East Atlantic oscillation, and El Niño Southern Oscillation, to identify different forecast models - in terms of both input variables and model parameters - for the different phases of the climate oscillations and for each month of the year. The resulting combination of global and local variables is then used as input in different ML models, including both feedforward neural networks and extreme learning machines.

Our framework is developed within the CLimate INTelligence (CLINT) project and tested in the task of predicting the total monthly precipitation in the Rijnland area (Netherlands). The resulting ML-based forecasts are then benchmarked against state-of-the-art dynamic forecast products, i.e. the ECMWF Extended Range forecasts. Our findings indicate that combining global and local climate

information into ML-based forecast models significantly improves state-of-the-art drought forecast accuracy, thus representing a promising option to timely prompt anticipatory drought management measures.