Title: Assessment of tree-based boosting machine learning techniques for forecasting high-impact drought events across dry and hot future climate scenarios: A case study from Italy

Majid Niazkar[1, 2*], Jeremy Pal[1, 2], Jaroslav Mysiak[1, 2]
[1] *Risk Assessment and Adaptation Strategy division, Euro-Mediterranean Center of Climate Change, Venice, Italy*
[2] *Ca' Foscari University of Venice, Venice, Italy*
* *e-mail: majid.niazkar@cmcc.it*

Thematic relevance of the poster:
Climate Predictions and Extremes – Machine learning techniques for forecasting high-impact events and their attribution

Abstract:
In the realm of climate predictions and extremes, machine learning (ML) techniques serve as versatile tools not only for capturing climate patterns but also for climate change impact assessments. Generally, each ML method comes with a substantial predictive skill, which demands for comparing performances of different models when selecting the most robust one. Among ML models, decision tree is a basic approach with a tree structure, whose performance was enhanced by implementing boosting techniques and ensemble learning. The improvement led to several tree-based boosting techniques including Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), Light Gradient-Boosting Machine (LightGBM), and Category Boosting (CatBoost). In this study, performances of these ML models were compared for predicting extreme drought under climate change impacts in the upper part of Reno basin, northern Italy. For this purpose, 39 bias-adjusted EURO-CORDEX CMIP5 climate models with 71 future projections (rcp2.6, rcp4.5, and rcp8.5) were extracted for the study area. The projections were classified into four groups (dry vs. wet and hot vs. cool) based on changes in precipitation and temperature, while the focus of this study was merely on the hot and dry group. Furthermore, reanalysis precipitation and temperature data (1971-2005) were utilized as common baseline to calculate 6-month Standardized Precipitation Evapotranspiration Index (SPEI) values, which were used to train ML models. For ML modelling, monthly mean precipitation, temperature, and potential evapotranspiration were input data. The ML performances were evaluated by ground-based observed data (2001-2023) and the corresponding SPEI values. The comparative analysis indicates that tree-based boosting techniques have a high accuracy in predicting SPEI for both training and testing dataset. Moreover, the most robust ML model was selected for forecasting SPEI for each climate projection separately. Finally, forecasted SPEI values under hot and dry projections provide valuable insights for high-impact drought events in future, while future studies can consider other three groups of climate projections to explore risk shift under different climate change scenarios.

Acknowledgements: This research work was carried out as part of the TRANSCEND project with funding received from the European Union Horizon Europe Research and Innovation Programme under Grant Agreement No. 10108411.