EGU21-1305, updated on 19 Sep 2021
https://doi.org/10.5194/egusphere-egu21-1305
EGU General Assembly 2021


# Applying machine learning for drought prediction using a large ensemble of climate simulations

**Elizaveta Felsche**[1,2,3] and Ralf Ludwig[1]
[1]Department of Geography, Ludwig Maximilians University of Munich, Munich, Germany
[2]Center for Digital Technology and Management, Munich, Germany
[3]Technical University of Munich, Munich, Germany

There is strong scientific and social interest to understand the factors leading to extreme events in order to improve the management of risks associated with hazards like droughts. Recent events like the summer 2018 drought in Germany already had severe und unexpected impacts, e.g. forest fires and crop failures; in order to increase preparedness robust prediction tools are urgently required. In this study, machine learning methods are applied to predict the occurrence of a drought with lead times of one to three months. The approach takes into account a list of thirty atmospheric and soil variables as predictor input parameters from a single regional climate model initial condition large ensemble (CRCM5-LE). The data was produced the context of the ClimEx project by Ouranos with the Canadian Regional Climate Model (CRCM5) driven by 50 members of the Canadian Earth System Model (CanESM2) for the Bavarian and Quebec domains.

Drought occurrence was defined using the Standardized Precipitation Index. The training and test datasets were chosen from the current climatology (1955-2005) for the Munich and Lisbon subdomain within the CRCM5-LE. The best performing machine learning algorithms managed to obtain a correct classification of drought or no drought for a lead time of one month for around 60 % of the events of each class for the both domains. Explainable AI methods like feature importance and shapley values were applied to gain a better understanding of the trained algorithms. Physical variables like the North Atlantic Oscillation Index and air pressure one month before the event proved to be of high importance for the prediction. The study showed that better accuracies can be obtained for the Lisbon domain, due to the stronger influence of the North Atlantic Oscillation Index on Portugal's climate.