EGU24-15950, updated on 09 Oct 2024
https://doi.org/10.5194/egusphere-egu24-15950
EGU General Assembly 2024

Leveraging Information Flow for Data-Driven Subseasonal Forecasting of Sahelian Hot Extremes

Victoria M. H. Deman, Daniel F. T. Hagan, Damián Insua-Costa, Akash Koppa, and Diego G. Miralles
Hydro-Climate Extremes Lab (H-CEL), Ghent University, Belgium (victoria.deman@ugent.be)

The semi-arid Sahel region has witnessed an increase in extreme weather conditions such as repeated drought cycles, desertification, heatwaves and floods in recent decades. These events pose existential threats to the already vulnerable population and natural ecosystem. Addressing the underexplored potential of subseasonal forecasting in the Sahel, data-driven models offer an alternative to traditional dynamical approaches. These models – distinguished by enhanced computational efficiency, reduced sensitivity to initial conditions, the ability to learn intricate relationships from data, and the ability to capture nonlinear dynamics – represent an asset in building resilience in the region.

This study investigates the potential of employing a rigorous causality framework based on the Liang–Kleeman information flow for predictor selection. Previous research has underscored the pitfalls of using correlations for predictor selection when forecasting using machine learning models, as spurious correlations may lead to the selection of predictors without any physical connection. In response, our research investigates the potential of this information flow causality to select predictors within a vast array of predefined variables, including coupled ocean–atmospheric oscillation indices, sea-surface temperatures, vegetation indices and soil moisture. Subsequently, our focus is directed towards predicting summer maximum temperature extremes with lead times of 2, 4, 8 and 16 weeks using the selected predictors and a variety of deep learning techniques. Despite the challenge of predicting short-lived heatwaves in a region characterised by the high baseline temperatures, our results indicate that the information flow causality effectively reduces dimensionality, and enables a selection of features with causal relationships that facilitates subsequent forecasting. In the following, the causal knowledge from the predictor selection step will be quantitatively transferred into the machine learning models themselves, thereby providing an interpretable framework for the prediction of the hot extremes in the region.