EGU22-5464
https://doi.org/10.5194/egusphere-egu22-5464
EGU General Assembly 2022

Using interpretable machine learning to identify compound meteorological drivers of crop yield failure

Lily-belle Sweet and Jakob Zscheischler
Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany (lily-belle.sweet@ufz.de)

Extreme weather events, such as droughts, floods or heatwaves, severely impact agricultural yield. However, crop yield failure may also be caused by the temporal or multivariate compounding of more moderate weather events. An example of such an occurrence is the phenomenon of 'false spring', where the combined effects of a warm interval in late winter followed by a period of freezing temperatures can result in severe damage to vegetation. Alternatively, multiple weather events may impact crops simultaneously, as with compound hot and dry weather conditions.

Machine learning techniques are able to learn highly complex and nonlinear relationships between predictors. Such methods have previously been used to explore the influence of monthly- or seasonally-aggregated weather data as well as predefined extreme event indicators on crop yield. However, as crop yield may be impacted by climatic variables at different temporal scales, interpretable machine learning methods that can extract relevant meteorological features from higher-resolution time series data are desirable.

In this study we test the ability of adaptations of random forest models to identify compound meteorological drivers of crop failure from simulated data. In particular, adaptations of random forest models capable of ingesting daily multivariate time series data and spatial information are used. First, we train models to extract useful features from daily climatic data and predict crop yield failure probabilities. Second, we use permutation feature importances and sequential feature selection to investigate weather events and time periods identified by the models as most relevant for crop yield failure prediction. Finally, we explore the interactions learned by the models between these selected meteorological drivers, and compare the outcomes for several global crop models. Ultimately, our goal is to present a robust and highly interpretable machine learning method that can identify critical weather conditions from datasets with high temporal and spatial resolution, and is therefore able to identify drivers of crop failure using relatively few years of data.