EGU24-10153, updated on 28 May 2024
https://doi.org/10.5194/egusphere-egu24-10153
EGU General Assembly 2024

A Spatio-Temporal Optimization- Based Feature Selection Framework for detecting drivers of heatwaves

César Peláez-Rodríguez[1], Jorge Pérez-Aracil[1], Ronan McAdam[2], Antonello Squintu[2], Enrico Scoccimarro[2], and Sancho Salcedo-Sanz[1]
[1]Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Spain
[2]Climate Variability and Prediction Division, Fondazione Euro-Mediterraneo sui Cambiamenti Climatici - CMCC, Bologna, Italy.

Machine Learning (ML) encompasses various techniques and algorithms that have proven highly effective in addressing complex climate science tasks. In particular, using ML to detect and forecast extreme events has gained much attention recently. Considering the vast volume of spatial and temporal data available, the employment of data-driven methodologies becomes indispensable for effectively uncovering potential drivers of these events. This study arises with the ambition of proposing a comprehensive and general framework that provides insights about interactions between heatwaves and potential physical drivers across multiple spatio-temporal scales. A novel feature-selection methodology is presented to advance the detection of heatwaves. It is based on a two-step procedure. In the first stage, a non-supervised task is developed for spatially clustering the different variables. Thus, there is a reduced initial pool of driver candidates. In the second stage, a wrapper methodology is applied to determine which time periods are representative for each of the clusters in the occurrence of heatwaves. This algorithm discerns which clusters are selected as drivers and which are discarded, along with the time period, relative to the heatwave occurrence, during which each cluster should be investigated. Thus, the feature selection is developed based on the spatio-temporal distribution of the different variable clusters.

Experiments have been developed for detecting heatwaves in the Lake Como region, a region of key agricultural activity in Northern Italy. A wide range of ocean and atmospheric variables taken from ERA5 are used (e.g. sea ice concentration, precipitation). The framework allows the identification of the time lag for each variable from short-term to seasonal time scales (up to 180 days). Results have spotted drivers on the subseasonal to seasonal timescale. Important drivers have been identified for short-term periods (less than one week). Local variables are shown to be of much significance for these periods. The method allows for identifying the strong influence of local variables whilst identifying correlations between the heatwave occurrence and different variables and spatial locations. The importance stages of the variable candidates can be established by running the model while removing the most critical variables.