EGU2020-21329
https://doi.org/10.5194/egusphere-egu2020-21329
EGU General Assembly 2020

Multisensor crop yield estimation with machine learning

Laura Martínez Ferrer, Maria Piles, and Gustau Camps-Valls
Universitat de València, Image Processing Laboratory, Paterna, Spain (laumarf3@alumni.uv.es)

Providing accurate and spatially resolved predictions of crop yield is of utmost importance due to the rapid increase in the demand of biofuels and food in the foreseeable future. Satellite based remote sensing over agricultural areas allows monitoring crop development through key bio-geophysical variables such as the Enhanced Vegetation Index (EVI), sensitive to canopy greenness, the Vegetation Optical Depth (VOD), sensitive to biomass water-uptake dynamics, and Soil Moisture (SM), which provides direct information of plant available water. The aim of this work is to implement an automatic system for county-based crop yield estimation using time series from multisource satellite observations, meteorological data and available in situ surveys as supporting information. The spatio-temporal resolution of satellite and meteorological observations are fully exploited and synergistically combined for crop yield prediction using machine learning models. Linear and non-linear regression methods are used: least squares, LASSO, random forests, kernel machines and Gaussian processes. Here we are not only interested in the prediction skill, but also on understanding the relative relevance of the covariates. For this, we first study the importance of each feature separately and then propose a global model for operational monitoring of crop status using the most relevant agro-ecological drivers.

We selected the Continental U.S. and a four-year time series dataset to perform the research study. Results reveal that the three satellite variables are complementary and that their combination with maximum temperature and precipitation from meteorological stations provides the best estimations. Interestingly, adding information about crop planted area also improved the predictions. A non-linear regression model based on Gaussian processes led to best results for all considered crops (soybean, corn and wheat), with high accuracy (low bias and correlation coefficients ranging from 0.75 to 0.92). The feature ranking allowed understanding the main drivers for crop monitoring and the underlying factors behind a prediction loss or gain.