Geophysical Research Abstracts
Vol. 21, EGU2019-14329, 2019
EGU General Assembly 2019

Extracting important features for crop yield prediction with convolutional neural networks on remote sensing and meteorological data

Aleksandra Wolanin (1), Luis Guanter (1), Gustau Camps-Valls (2), and Grégory Duveiller (3)

(1) GFZ, Potsdam, Germany (ola@gfz-potsdam.de), (2) Universitat de Valencia, Valencia, Spain, (3) European Commission Joint Research Centre (JRC), Ispra, Italy

Agriculture monitoring and yield estimation are important for food security, both regionally and globally. Understanding the year-to-year variability in crop yield and its relationship to meteorological conditions is particularly important for several regions where yield is highly dependent on changing environmental factors. For example, wheat yields in India have been steadily increasing since the 1960s and 1970s due to benefits of the Green Revolution, but in recent years the wheat yield has been unstable as major crop yield losses were attributed to unfavorable meteorological conditions. Modeling the effects of various environmental variables can be challenging, as their impact on the final yield is complex and varies depending on their intensity and the crop growth stage at which they occur (e.g., moderate rainfall is beneficial for crops, but extensive and untimely rainfalls can lead to huge yield losses).

In this work, we exploit interconnections between meteorological conditions and satellite data on vegetation during the whole growing season, and their simultaneous impact on wheat yield in the Wheat Belt in India. We use GLDAS 2.1 data as the meteorological input and MODIS data for the vegetation remote sensing input. Adding satellite information on crop is crucial for yield estimation, as it carries information on both crop phenology, as well as the crop response to the meteorological conditions. We apply machine learning algorithms (e.g., convolutional neural networks, CNNs) that can model non-linear processes and can extract important features in the multivariate time series automatically from data, without prior knowledge or human effort in feature design. By doing so, we do not force assumptions on which time is the most important for the final crop yield and we can include in the analysis the whole time series of multiple input variables at a high temporal resolution. Furthermore, we analyze the CNNs in terms of important features and crucial time windows for yield estimation, which shows that they the vary across space and time. By combining meteorological and satellite vegetation data with CNNs this work may help to disentangle the complex interactions between the features in the time series of the input data and the wheat yield.