EGU21-14590
https://doi.org/10.5194/egusphere-egu21-14590
EGU General Assembly 2021

Cereal yield forecasting combining satellite drought-based indices, regional climate and weather data using machine learning approaches in Morocco

El houssaine Bouras[1,2], Lionel Jarlan[2], Salah Er-Raki[1,3], Riad Balaghi[4], Abdelhakim Amazirh[3], Bastien Richard[5], and Saïd Khabba[3,6]

[1]ProcEDE, Department of Applied Physique, Faculty of Sciences and Technologies, Cadi Ayyad University, Marrakech, Morocco. (bouras.elhoussaine@gmail.com)

[2]CESBIO, University of Toulouse, IRD/CNRS/UPS/CNES, Toulouse, France

[3]Center for Remote Sensing Applications (CRSA), University Mohammed VI Polytechnic (UM6P), Benguerir, Morocco

[4]National Institute for Agronomic Research (INRA), Rabat, Morocco

[5]G-EAU, University Montpellier, AgroParisTech, CIRAD, IRD, INRAE, Institut Agro, Montpellier, France

[6]LMFE, Department of Physics, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco.

Cereals are the main crop in Morocco. Its production exhibits a high inter-annual due to uncertain rainfall and recurrent drought periods. Considering the importance of this resource to the country's economy, it is thus important for decision makers to have reliable forecasts of the annual cereal production in order to pre-empt importation needs. In this study, we assessed the joint use of satellite-based drought indices, weather (precipitation and temperature) and climate data (pseudo-oscillation indices including NAO and the leading modes of sea surface temperature -SST- in the mid-latitude and in the tropical area) to predict cereal yields at the level of the agricultural province using machine learning algorithms (Support Vector Machine -SVM-, Random forest -FR- and eXtreme Gradient Boost -XGBoost-) in addition to Multiple Linear Regression (MLR). Also, we evaluate the models for different lead times along the growing season from January (about 5 months before harvest) to March (2 months before harvest). The results show the combination of data from the different sources outperformed the use of a single dataset; the highest accuracy being obtained when the three data sources were all considered in the model development. In addition, the results show that the models can accurately predict yields in January (5 months before harvesting) with an $R^2 = 0.90$ and RMSE about 3.4 $Qt.ha^{-1}$. When comparing the model's performance, XGBoost represents the best one for predicting yields. Also, considering specific models for each province separately improves the statistical metrics by approximately 10-50% depending on the province with regards to one global model applied to all the provinces. The results of this study pointed out that machine learning is a promising tool for cereal yield forecasting. Also, the proposed methodology can be extended to different crops and different regions for crop yield forecasting.