Gross primary productivity and yield estimation of winter wheat based on machine learning

OUSAYD Lahcen [(1)(*)], BELAQZIZ Salwa [(1) (2)], KHABBA Said [(2) (3)], ERRAKI Salah [(2) (4)], EL HACHIMI Chouaib[(2)], CHEHBOUNI Abdelghani [(2) (5)]

(1) *Ibn Zohr University, Faculty of Science, LabSIV Laboratory, Department of Computer Science, Agadir, Morocco.*
(2) *Mohammed VI Polytechnic University, Center for Remote Sensing Applications (CRSA), Benguerir, Morocco.*
(3) *Cadi Ayyad University, Faculty of Sciences Semlalia, LMFE, Department of Physics, Marrakesh, Morocco.*
(4) *Cadi Ayyad University, Faculty of Sciences and Technology, ProcEDE, Department of Physics, Marrakesh, Morocco.*
(5) *Université de Toulouse, Centre d'Etudes Spatiales de la Biosphère (CESBIO), Toulouse, France.*

(*) Corresponding author: lahcen.ousayd@gmail.com

Keywords: Crop yield forecasting; light use efficiency ; machine learning; remote sensing, vegetation indices; weather data; Winter wheat

Population increase, urbanization, and climate change have an impact on the food security of a country's population. This food security is related to cereals, particularly wheat, as it represents the main ingredient in foods in several countries, including Morocco. Thus, monitoring and early forecasting of wheat yield are of great relevance in this regard. Recently, the use of machine learning and deep learning algorithms to help decision-makers to have a reliable forecasts of crop production became an interesting topic for research. In this context, this study aims to use different regression models for predicting wheat yield. As machine learning algorithms usually need large volume of observations for model training, due to the absence of sufficient data, we have first applied a satellite data-driven light use efficiency (LUE) model to estimate gross primary productivity (GPP), and then combined this model with the harvest index (ratio of aboveground biomass to yield) to convert the estimated wheat GPP to yield in 250 wheat fields in R3 irrigated zone (about 40 Km east of Marrakech city). We feed the dataset to three machine learning models and a deep learning model (Long Short Term Memory (LSTM)). Overall, The calculated wheat are consistent with the observed grain yield. However, The (LSTM) method exhibited the best metrics, with an R^2 = 0.90 and RMSE of 0.20 t ha^{-1}.