Tropentag, September 11-13, 2024, hybrid conference

"Exploring opportunities …
for managing natural resources and a better life for all"

Forecasting grain maize yield in sub-Saharan Africa: A hybrid modelling approach

Amit Kumar Srivastava[1], Krishnagopal Halder[2], Jan Timko[3], Lily-belle Sweet[4], Manmeet Singh[5], Thomas Gaiser[6], Frank Ewert[6], Radwa Elshawi[3]

[1] *Leibniz Center for Agricultural Landscape Research (ZALF), Germany*
[2] *Vidyasagar University, Dept. of Remote Sensing and GIS, India*
[3] *University of Tartu, Institute of Computer Science, Estonia*
[4] *Helmholtz Centre for Environmental Research, Dept. of Computational Hydrosystems, Germany*
[5] *University of Texas at Austin, United States*
[6] *University of Bonn, Inst. Crop Sci. and Res. Conserv. (INRES), Germany*

Abstract

The dual challenges of climate change and a burgeoning global population, projected to surpass 9 billion by the year 2030, present unprecedented hurdles for the agricultural sector. These challenges underscore the need for developing precise and timely crop yield forecast (CYF) models. To enhance CYF, various methodologies have been explored, ranging from process-based models (PBM) to data-driven statistical approaches. Process-based models' are hindered by uncertainties in structure, inputs, and parameters, exceeding observed yield variations over time/space, ehereas, Machine Learning (ML) models are often a black box needing explanations.Therefore, this paper delves into the transformative potential of hybrid model approach where we tried to use the outputs of PBM as features to improve the prediction accuracy of ML for maize yield predicitons across sub-Saharan Africa inclusing 54 countries. As input features into ML, the climate data was obtained from ERA5-Land consisting of a daily time series from 1981 to 2016 including solar radiation, precipitation, minimum and maximum temperature. Grain maize yield data from 1981–2016 were obtained from the PANGAEA (https://pangaea.de) database. The relevant soil parameters from African soil property maps (https://www.isric.org). We adopted time-based cross validation strategy, characterised by iterative training of the XGBoost model using twenty years of historical data. The results indicate an R^2 score of around 0.72 on the five-year forecast, along with MAE and RMSE values of 0.61 and 1.07 tons per hectare (t ha^{-1}) respectively under pure ML approach, however, the hybrid model approach indicated slightly better forecast with 4.1 %, 5.0 %, and 3.8 % in R^2, MAE and RMSE respectively. We also moved beyond prediction and interpreted the outputs of our model approach using SHAP and force plots which provided key insights in explaining the yield forecast results (importance of variables by time). We found soil type and water stress factor as the most critical features in grain maize yield forecasting.

Keywords: Hybrid model, maize, sub-Saharan Africa, yield forecast

Contact Address: Amit Kumar Srivastava, Leibniz Center for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany, e-mail: AmitKumar.Srivastava@zalf.de