EGU24-2028, updated on 15 Mar 2024
https://doi.org/10.5194/egusphere-egu24-2028
EGU General Assembly 2024

Improving early crop yield and price predictions using satellite imagery with machine and deep learning techniques

Florian Teste[1,2], Hugo Gangloff[1], Philippe Ciais[3], and David Makowski[1]
[1]Université Paris-Saclay, INRAE, AgroParisTech, UMR MIA 518, 91120, Palaiseau, France (florian.teste@agroparistech.fr)
[2]Atos France, Technical Services, 95870, Bezons, France
[3]CEA-CNRS-UVSQ, Université Paris-Saclay, IPSL, LSCE, Gif-sur-Yvette, 91190, France

Early and accurate crop yield predictions and prices are crucial for food security management and planning. However, the lack of pre-harvest data poses significant challenges, undermining the reliability and effectiveness of existing methods.
This study introduces an innovative approach that addresses these challenges using satellite data products—specifically, Gross Primary Production (GPP) (0.05° spatial resolution) and dimension-reduction techniques to forecast corn yield and price variation across various regions. We predict national corn yield and price variations by leveraging these satellite-derived products. The value of the approach is demonstrated in three case studies conducted for corn in the US (Corn Belt region), Malawi, and South Africa.
The predictors are derived from GPP year-on-year variation of each region at the peak growing season, i.e., in July for the US Corn Belt (harvest in October) and March for Malawi and South Africa (harvest in May).
We compute the spatial average and Principal Components (PCs) of the GPP year-on-year variations through Empirical Orthogonal Function (EOF) analysis. Additionally, we explore neural network architectures, including Autoencoder (AE) and Variational Autoencoders (VAEs), and extract latent features to reduce the dimension of the GPP data from several thousand to a dozen synthetic features. The PCs, the AE and VAE latent features are used as predictors in Generalized Linear Models (GLM) and Least Absolute Shrinkage and Selection Operator (LASSO) models for predicting year-to-year corn yield and price variation. A neural network is also trained to predict yield and price variations from the latent features for comparison. All models are evaluated using year-to-year cross-validation with three metrics, i.e., Area Under Curve (AUC), the Brier Skill Score (BSS), and the Matthew Correlation Coefficient (MCC).
Our results demonstrate the superior predictive performance of PCs for US corn yield variations with an AUC of 0.97 (95% CI 0.92-1), a BSS of 0.75, and an MCC of 0.83.
This approach outperforms alternative methods in performance, simplicity, and execution speed. The EOF approach also yields superior results for yield variation prediction in South Africa with an AUC of 0.88 (95% CI 0.75-0.99), a BSS of 0.47, and an MCC of 0.61, while the autoencoder approach is most effective for Malawi with an AUC of 0.98 (95% CI 0.93-1), a BSS of 0.75 and an MCC of 0.83.
For price, our results indicate that the spatial averages of GPP year-on-year July variation in the US Corn Belt can be used to forecast the forthcoming increase or decrease in global corn price at

harvest with an AUC of 0.92 (95% CI 0.75-0.99), a BSS of 0.5 and an MCC of 0.66. However, in South Africa and Malawi, the most accurate price predictions are obtained with the VAE approach. With VAE, the AUC is 0.75 (95% CI 0.59-0.92), the BSS is 0.2, and the MCC is 0.27 in South Africa, while these metrics reach 0.94 (95% CI 0.59-0.92), 0.63, and 0.7 in Malawi.
This study highlights the value of combining satellite data with dimension-reduction methods for large-scale prediction of crop yields and price variations several months before harvest.