Application of Spatially Adjusted Machine Learning Approaches to Improve Sorghum Biomass Prediction Using Unmanned Aerial Vehicles

Sebastian Varela[1]* (sv79@illinois.edu), **Andrew D. B. Leakey**[1,2]

[1] DOE Center for Advanced Bioenergy and Bioproducts Innovation; and [2] Department of Plant Biology, University of Illinois at Urbana Champaign, Urbana, IL

https://cabbi.bio/

Project Goals: In the context of design and engineer environmentally resilient crops with maximum yield and resource use efficiency, this project implements the use of UAV-acquired multispectral and thermal imagery to screen water use efficiency (WUE) and productivity traits in sorghum/Miscanthus field trials at scale.

Abstract Text:

Sorghum bicolor is a model C4 feedstock grass that has significant potential for improved production of biofuels and bioproducts. Phenotyping biomass production is a key element of field trials for breeding, quantitative genetics, and bioengineering that will increase productivity, resilience, sustainability, and value of crops. Traditional harvest methods for evaluating biomass production (yield) are very labor intensive and destructive, limiting the scale and speed of research. Remote sensing using cameras mounted to unmanned aerial vehicles (UAV) allows rapid, non-destructive, high-frequency data collection. Nevertheless, analytical methods are needed to exploit the high spatial and temporal resolution in images to predict final yield. Data has an explicit spatial nature that suggests that the true underlying relationship among dependent and independent variables can be spatially varying. Models that do not take this under consideration can draw limited inference. The application of Geographically Weighted Regression (GWR) as a spatially adjusted approach to improve yield prediction has been tested in wheat [1] and rice [2].

In this study, we introduce the Geographical Random Forest (GRF) concept. The principle idea is similar to that of GWR [3], in which we move to local computation (plot vicinity) rather than global (whole extent of experimental area) one, but relaxing normality assumption and non-linear dependencies between UAV-based features, and yield prediction. First, we identify the most informative features in image data from UAVs to predict yield of 870 diverse bioenergy sorghum accessions. Second, we then test GWR and GRF as local spatially adjusted approaches, and compare these spatially adjusted approaches with Principal Component Regression (PCR) as an “aspatial” global benchmark. Yield of biomass across a diverse population of accessions of biomass sorghum was predicted from UAV imagery with high precision and accuracy (r=0.93, RMSE = 3.06). The most important data descriptors predicting yield were: (1) UAV-derived height of the canopy 49 days after planting (DAP) associated with a period of fast vertical growth in mid-season; (2) canopy ground cover prior to canopy closure 43 DAP; and (3) canopy nitrogen status via Normalized Difference Red edge (NDRE) prior to canopy closure 43 DAP. PCR performance

was moderate, but with significant bias given its “aspatial” global fit. It overestimates low yield and underestimated high yield plots. GWR predictability outperforms PCR, it slightly improves accuracy and reduces bias, suggesting the benefit of a “locally” computed model. GRF outperforms GWR, suggesting the superiority of GRF as a spatially corrected, and non-linear model calibrated and computed locally rather than globally. This methodological advance will be used to advance CABBI research goals by evaluating field trials for breeding, quantitative genetics, and bioengineering that will increase productivity, resilience, sustainability, and value of crops.

1. Haghighattalab, A.; Crain, J.; Mondal, S.; Rutkoski, J.; Singh, R.P.; Poland, J. Application of geographically weighted regression to improve grain yield prediction from unmanned aerial system imagery. Crop Sci. 2017, 57, 2478–2489.
2. Shiu, Y.-S.; Chuang, Y.-C. Yield Estimation of Paddy Rice Based on Satellite Imagery: Comparison of Global and Local Regression Models. Remote Sens. 2019, 11, 111.
3. Fotheringham AS, Brunsdon C, Charlton M. 2003. Geographically weighted regression: the analysis of spatially varying relationships. Hoboken, NJ: John Wiley & Sons.

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.