Machine Learning to Predict Biomass Sorghum Yields under Future Climate Scenarios

Tyler Huntington[1,2], Xinguang Cui[1,2,3], Umakant Mishra[1,2,4]*, **Corinne D. Scown**[1,2,5,6]*

[1]Biosciences Area, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 94720, USA

[2]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA, 94608, USA

[3]Department of Aerospace Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei, 430074, China

[4]Environmental Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL, 60439, USA

[5]Energy Technologies Area, 1 Cyclotron Road, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

[6]Energy & Biosciences Institute, University of California, 2151 Berkeley Way, Berkeley, CA, 94704, USA

Project Goals: Predict future bioenergy sorghum yields across the continental United States, accounting for market and farmer behavior shifts, using a machine learning model trained on historical data to identify most promising regions for cultivation.

Crop yield modeling is critical in the design of national strategies for agricultural production, particularly in the context of a changing climate. Forecasting yields of bioenergy crops at fine spatial resolutions can help to evaluate near- and long-term pathways to scaling up bio-based fuel and chemical production, and to understand the impacts of abiotic stressors such as severe droughts and temperature extremes on potential biomass supply. We used a large dataset of 28,364 *Sorghum bicolor* yield samples (uniquely identified by county and year of observation), environmental variables, and multiple approaches to analyze historical trends in sorghum productivity across the U.S. We selected the most accurate machine learning approach (a variation of Random Forest) to predict future trends in sorghum yields under four greenhouse gas emission scenarios and two irrigation regimes. We identified irrigation practices, vapor pressure deficit, and time (a proxy for technological improvement) as the most important predictors of sorghum productivity. Our results showed a decreasing trend of sorghum yields over future years (on average 2.7% from 2018 to 2099), with greater decline under a high greenhouse gas emissions scenario (3.8%) and in the absence of irrigation (4.6%). Geographically, we observed the steepest predicted declines in the Great Lakes (8.2%), Upper Midwest (7.5%), and Heartland (6.7%) regions. Our study demonstrates the use of machine learning to identify environmental controllers of sorghum biomass yield and predict yields with reasonable accuracy. These results can inform the development of more realistic biomass supply projections for bioenergy if sorghum production is scaled up.