Using Machine Learning to Identify Cultivar x Site Interaction and Environmental Variable Affecting Aboveground Biomass

Li Zhang[1]*, Kathrine D. Behrman[1], Jason Bonnette[1], Thomas E Juenger[1]

[1]University of Texas at Austin, Austin TX

Project Goal:

Multi-Scale Modeling will define conditions of a sustainable biofuel system and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. The aim is to use multiscale modeling to predict switchgrass performance and sustainability under future climate change.

Abstract:

Switchgrass, a perennial grass native to North American, is a promising second generation biofuel crop. Previous studies have showed that switchgrass yields are sensitive to climatic variability associated with temperature and precipitation in space and time. Switchgrass yields are often cultivar-dependent and exhibit cultivar x environment interactions as well. In this study, we grew six switchgrass cultivars (two lowlands: Alamo and Kanlow; two uplands: Blackwell and Cave-in-Rock; and two hybrids: Liberty and Carthage) at three locations across US (TMPL: Temple TX; CLMB: Colombia, MO; and FRMI: Fermin Lab, IL) for three consecutive years (2017-2019). Seasonal data on destructive leaf area index (DesLAI), tiller count (TC), plant height (HT) was collected and aboveground biomass (DesBIO) was measured each year. Our preliminary statistical analyses show that end of season aboveground biomass is strongly affected by site and genotype but not year. To further investigate these interactions, multivariate time series clustering was used to determine how two phenotypes, seasonal TC and DesBIO, are related for all cultivars for all locations. Not surprisingly, the two upland cultivars and the two lowland cultivas make two distinct clusters regardless of site. The hybrid Liberty always clutters with the lowland cultivar, indicating that phenotypically it is a lowland. On the other hand, the other hybrid, Carthage, phenotypically resembles an upland in the northern sites (CLMB and FRMI) and a lowland in the southern site (TMPL), thus indicating this cultivar is displaying strong cultivar x environment interactions. Next, machine learning (i.e., random forest algorithm) was used to identify the environmental variable(s) affecting these phenotypes. Average temperature between each sampling interval was identified as the major factor influencing aboveground biomass. There is a positive linear relationship between temperature and biomass when average temperature is between 15 to 25°C. At average temperature values between 25 to 30°C, biomass plateaus and is constant. This study is an example of how intra-annual time series data and machine learning methods (multivariate time-series clustering and random forest) can help identify cultivar by site interaction and the shape between environmental variables and phenotypes.

Funding Statement:

This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), award number DE-SC0014156 to TEJ.