



## **Integration of multi-modal proximal remote sensing data for improved prediction of agronomically important crop traits**

**Erin Farmer<sup>1</sup>**, Ruyu Yan<sup>5</sup>, Peter Michael<sup>5</sup>, Nick Lepak<sup>4</sup>, Cinta Romay<sup>2</sup>, Ed Buckler<sup>1,2,4</sup>, Noah Snavelly<sup>6</sup>, Ying Sun<sup>7</sup>, Kelly Robbins<sup>1</sup>, Joe Gage<sup>3</sup>, Abe Davis<sup>5</sup>, Michael Gore<sup>1</sup>

<sup>1</sup>*Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, New York 14853, USA*

<sup>2</sup>*Institute for Genomic Diversity, Cornell University, Ithaca, New York 14853, USA*

<sup>3</sup>*Department of Crop and Soil Sciences, North Carolina State University, Raleigh, North Carolina 27695, USA*

<sup>4</sup>*United States Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, New York 14853, USA*

<sup>5</sup>*Department of Computer Science, Cornell University, Ithaca, New York 14853, USA*

<sup>6</sup>*Cornell Tech, Cornell University, Ithaca, New York 14853, USA*

<sup>7</sup>*Soil and Crop Sciences Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA*

**ORCID:** 0000-0002-4841-8284

**Keywords:** remote sensing, machine learning, lidar, multispectral, uav, drone, ugv, rover

Recent advancements in proximal remote sensing have increased the spatial and temporal resolution of data collection, as well as the availability of these technologies for applications to precision agriculture. These sensors have allowed the collection of new and large quantities of data, which have been used to successfully determine phenotypes and parametrize crop growth models. So far, these data streams have been mostly used separately, though they contain unique structural, spatial, and spectral information. Thus, this research aims to integrate these disparate data sources to improve estimations of agronomically important crop traits. In this study, we examine two high-throughput and relatively inexpensive remote platforms: unoccupied ground vehicles (UGV) and unoccupied aerial vehicles (UAV). Data were collected on maize hybrids from the Genomes to Fields initiative over 5 years, from 2018 to 2022, in Aurora, NY. We used ground rovers to collect lidar scans, which were converted to point clouds, to construct the three-dimensional sub-canopy architecture of maize plants. Multispectral sensors, covering red, green, blue, red-edge, and near infrared (NIR) were deployed on a UAV platform to characterize maize canopies. Machine learning methods, including autoencoders, will be used to extract latent phenotypes from the lidar point clouds and multispectral images. Ultimately, these will be used



to predict manually measured traits, such as yield, in order to compare the prediction accuracies of models using these measurements separately and jointly.