Northern NY Agricultural Development Program
2020 Final Report

Using Satellite and Automated Drone Imagery to Give All Farmers Access to Yield Estimates for Zone-Based Field Management

Project Leader:
Quirine M. Ketterings, Cornell Nutrient Management Spear Program (NMSP),
323 Morrison Hall, Department of Animal Science, Cornell University, qmk2@cornell.edu

Collaborators:
- Champlain Valley Agronomics: Eric Beaver and Mike Contessa
- Cornell campus collaborators: Sunoj Shajahan, Jason Cho, Greg Godwin (NMSP), Joe Guinness (Statistical Sciences), Karl Czymmek (PRO-DAIRY)
- Northern NY dairy and cash grain producers with yield monitor data

Background:

Yield monitors are now increasingly used to determine yield of corn silage and grain on farms across Northern New York. With yield data, farmers can build their own yield potential database for nutrient management planning, engage in the adaptive nitrogen (N) management process, and participate in meaningful on-farm research. However, not all farms can afford yield-monitoring equipment, data collection requires calibration (and farm scales), and post-harvest data cleaning.

This project was initiated to evaluate if/how data sources such as aerial images obtained with unmanned aerial systems (drones) and satellites can be used to estimate corn grain and silage yields. If accurate predictions are possible, an approach can be designed to give all corn growers access to reliable yield data and we can avoid issues with calibration, sensor breakdown during harvest, and post-harvest data cleaning. Such an image-based approach can therefore be time-saving and potentially result in more accurate estimates. Furthermore, grain and silage yields can be estimated for previous years, based on the availability of imagery data, which allows farmers to more quickly develop yield stability zones for zone-based management decisions.

Methods:

Part A: Drone Imagery:

We evaluated the timing of drone flight (growth stage) and vegetation index type on accuracy of image-based yield predictions using data collected in 2019 at Aurora, NY. We used a Quantix unit (unmanned aerial vehicle mounted with a multispectral camera), a unit being used by Champlain Valley Agronomics in Peru, NY. The treatments included N sidedress at V4, V6, V8, and V10 along with no-N and N-rich strips. We collected drone images at weekly intervals starting from emergence (VE) to maturity (R5) to answer two questions:

(1) what is the best timing for image collection (growth stage), and
(2) what vegetation index gives us the most accurate estimations.

Part B: Satellite Imagery:

Datasets and indices calculations:

- ***Yield monitor data:*** We obtained yield monitor data from six farms in NNY, of which four had more than four years of yield data and a minimum of two fields with N-rich strips in 2019. The dataset included grain and silage yield data. Raw data were cleaned of errors using the standardized data cleaning protocol developed for NY corn fields.

- ***Remote sensing data:*** We downloaded satellite imagery obtained from PlanetLab's CubeSAT platform. This platform offers free imagery captured daily at 3 m/pixel resolution with four spectral bands (red, green, blue, and near-infrared [NIR]). In this study, we used the crop canopy reflectance signals before the senescence stage (before leaves started to turn yellow), and soil reflectance signals before planting for the same fields. For example, for one of the farms, the downloaded crop imagery was captured on September 25, 2019, and the soil imagery was captured on May 22, 2019. From both these images, we used the individual spectral band values from four bands as a variable to estimate yield (2 images × 4 bands = 8 features).

- ***Vegetation indices*** (5 features): From the satellite image of crop canopy, we calculated five vegetation indices that are commonly being used to estimate yield, crop health, and biomass content. The five indices are normalized difference vegetation index (NDVI), two enhanced vegetation index (EVI and EVI2), green normalized difference vegetation index (GNDVI), and excess green (EXG). We also used the individual spectral band values from four bands.

- ***Soil indices*** (5 features): Previous literature on developing yield estimation models did not use soil indices as a variable, but we believed the soil indices will also have an influence on yield estimations. Therefore, in this study, we calculated five different soil indices from the soil reflectance imagery obtained before planting. The five indices were brightness index, coloration index, hue index, saturation index, and redness index. Also, we used individual spectral band values from four bands.

Digital elevation model (1 feature): Along with features, we also used the digital elevation model (DEM) data obtained from the New York State Geographic Information Systems (GIS) Clearinghouse (https://gis.ny.gov/) at 1 m/pixel resolution.

Model Development:
Yield variability within field is large and the relationship between the image features and yield is not linear. Machine learning approaches are proven to be accurate in developing relationships with this kind of non-linear data. Therefore, in our study, we used machine-learning approaches to build yield prediction models using the image features.

One of the common and first-hand approaches in machine learning is the "random forest" model, predominantly used for classification applications. Before building a model, it is necessary to identify and select the best features that will be important to develop that model and eliminate the rest of the features. We used a feature selection approach called the "Boruta" method, which is the most recommended method for random forest models. To make our data suitable for this machine-learning model, we converted the yield data from numerical variable (measurable quantities) to categorical variable (as classes). We divided the entire range of yield data into the following five yield classes: high, medium-high, medium, medium-low, and low. Each class represents a range of yield values. The yield data were divided into two sets: one set with 80% of the data used for training the model and second set with the remaining 20% used to test the model. Only the selected five features were used to develop models.

Results:

Part A: Drone Imagery:
Of all vegetation indices tested, NDVI consistently was best indicator for estimating yield.

We also discovered that flights at R4 gave most accurate predictions, better than flights after leaf senescence or flights at earlier vegetative stages.

Timing of sidedressing impacted the accuracy of the yield estimations, with less reliable predictions when sidedressing occurred after V6.

We developed an exponential model to estimate the yield (as numeric value) using the reflectance signals from the no-N, N-rich, and sidedress at V4 treatments. The resulting estimated yield map was similar to the yield monitor data points (Figure 1). Furthermore, it showed clear differences in yield at the three no-N strip locations.

The findings of these analyses showed that our satellite work should focus on images collected later in the season, at or close to R4, with NDVI as vegetation index.

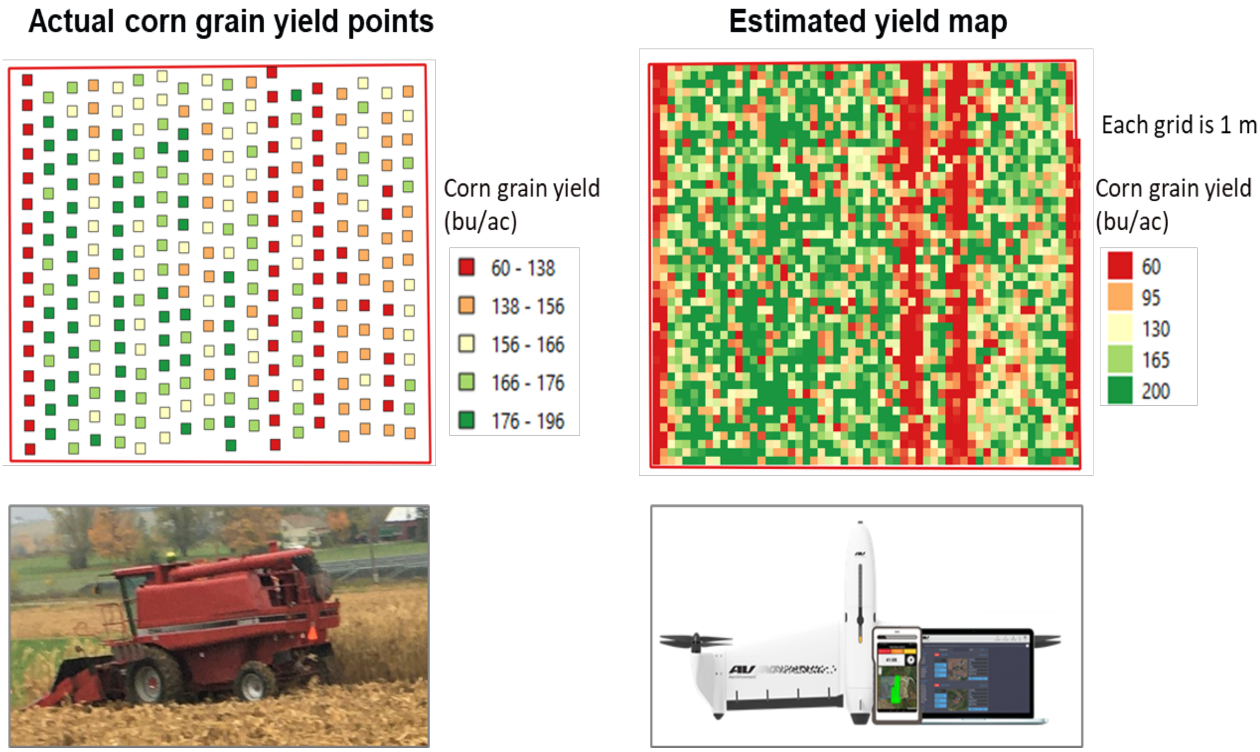

Figure 1. Cleaned yield monitor data and the estimated yield map from the developed exponential model.

Part B: Satellite Imagery:

The top five features identified by the Boruta method were digital elevation (DEM), crop greenness, soil NIR, NDVI, and soil redness. The performance of the random model that included these five features was validated with the test data (20%), which produced a classification accuracy of 58% for grain and 68% for silage.

Figure 1 shows the cleaned yield monitor data and the estimated yield map generated from the trained model with five yield classes. To generate the estimated map, the selected five features were extracted from the satellite images and the DEM and passed into the trained model to classify into different yield classes.

Even with somewhat low accuracy (58%), the estimated yield map looked similar to the actual yield map (Figure 2). Specifically, the low-yielding headland regions and the medium- and medium-high yielding non-headland regions were appropriately identified. A few patches of low-yielding portions of the field in the non-headland region were also identified. Reducing the number of yield classes to three and developing the model improved the accuracy to 77% for grain and 84% for silage, but the yield map had less information when only three yield classes were used.

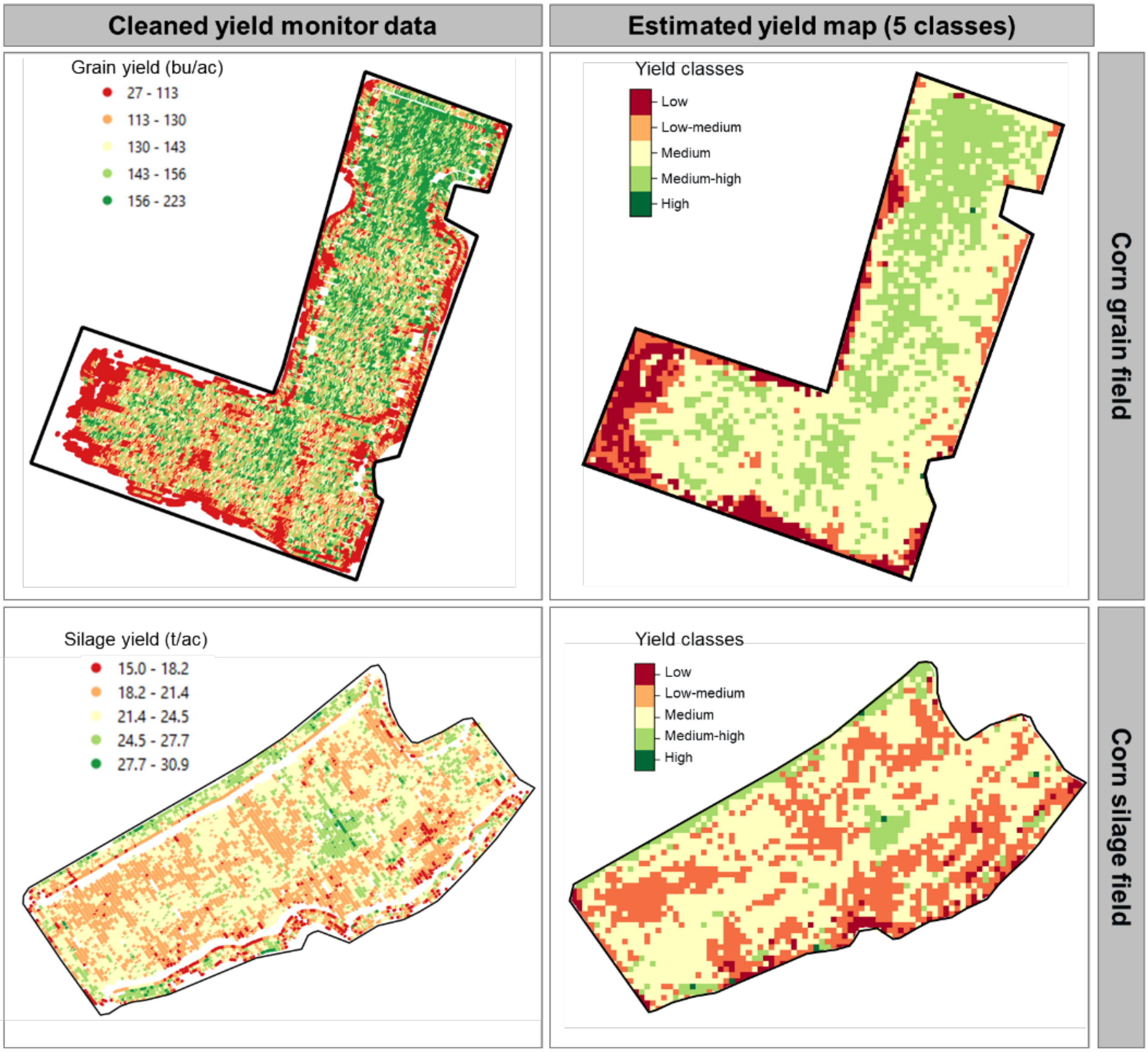

Figure 2. Cleaned yield monitor data (left) visualized after processing and cleaning and the estimated yield map (right) obtained from the satellite imagery and digital elevation model from a grain field (top) and a silage field (bottom).

Conclusions/Outcomes/Impacts:

The estimated yield map using the developed models showed promising results, demonstrating the potential to use free data layers (satellite images and DEM) for mapping the yield without the yield monitor data.

Work is currently ongoing to evaluate other fields for which imagery and yield monitor data were obtained, and to evaluate the importance of field calibration (knowing the actual yield of a known area in the field to predict yield in other parts of the field). We hope to continue this work with expansion of fields, in the hope to develop yield stability zone maps for farmers.

Our current focus is on improvement of the models with the future goal to develop a standalone tool, which takes the data layers and automatically produces a yield and/or zone map. Such a tool will give all farmers access to yield estimates for zone-based field management, which will benefit allocation of resources to areas that need the extra inputs (such as nutrients) or benefit from inputs (seed populations, etc.) and away from areas where yields are limited due to other reasons.

Outreach:

This project was described at winter meetings including the Northeast Region Certified Crop Advisor annual meetings where we gave a presentation on "Creating and managing zones within fields; a pathway forward", attended by approximately 125 people, mostly certified crop advisors. The farms that shared data received their farm-specific yield reports.

We held meetings that were attended by consulting firms to get feedback on zone management (practices, barriers, need for research) that indicated the importance of evaluation tools and yield assessment tools that are more easily accessible to farmers and farm advisors.

Additional talks and extension articles will be developed once analyses of the 2019 initial dataset is completed (anticipate by March 31).

Next Steps:

With the help of two computer science/statistics students we aim to expand on our current analyses to include more acres and years of data and to evaluate accuracy when image collection is done earlier in the growing season. We also plan to write an agronomy fact sheet on the use of satellite imagery for precision agriculture, with a focus on different satellite platforms and information such as resolution of data, frequency of flights, where to download the imagery for various satellites, and how to use the data. We are actively pursuing collaborations and funding to expand current findings to statewide evaluations.

Acknowledgments:

We thank the regional farmers and crop consultants who shared yield data with us that allowed us to get a large database in place for use in this project funded by the Northern New York Agricultural Development Program. Co-funding was obtained from the New York Farm Viability Institute (Aurora farm data collection in 2019) and USDA-NIFA (Federal Formula Funds).

Reports/articles in which results of this project have been published:

Website updates:

- Drones and Satellite Imagery for Yield Predictions and N Management Decisions: http://nmsp.cals.cornell.edu/NYOnFarmResearchPartnership/DronesforMgt.html.

For More Information:

Quirine M. Ketterings, Cornell Nutrient Management Spear Program (NMSP), Dept. of Animal Science, Cornell University, qmk2@cornell.edu, 607-255-3061, http://nmsp.cals.cornell.edu.