Northern NY Agricultural Development Program
2023 Final Report

Satellite-Derived Yield Classification Maps and Yield Stability Zones for Management

Project Leader:

Quirine M. Ketterings, Ph.D., Cornell Nutrient Management Spear Program (NMSP), 323 Morrison Hall, Department of Animal Science, Cornell University, Ithaca, NY.

Collaborators:

- Northern New York dairy farms
- Crop Consultants and Nutrient Management Planners: Mike Contessa and Eric Beaver, Champlain Valley Ag, Peru, NY
- Cornell Campus: NMSP: Subhashree Srinivasagan, Nayana Venukanthan, Sunoj Shajahan; School of Integrative Plant Sciences: Louis Longchamps
- Rochester Institute of Technology: Jan van Aardt

Background:

Yield monitor systems have revolutionized collection of spatial yield data during harvesting, enabling the mapping of corn grain and silage yields at within-field scale. While this technology has greatly facilitated on-farm research in New York, it does have limitations. It is costly to implement, and equipment failures during harvesting can lead to the loss of valuable data, including whole harvest seasons. Additionally, the raw data gathered by yield monitor sensors contain errors and need post-harvest data cleaning. To address these challenges, we proposed exploring the use of remotely-sensed data obtained from satellites to estimate yield classes and create yield-based management zones. Our specific objectives included (i) determining the optimal timing for satellite data collection, (ii) identifying key predicting features that can be used to derive yield classes and develop yield stability zones, and (iii) determining the minimum number of fields or acres required to build a reliable model for the whole farm.

Methods:

Yield monitor data:

Four farms located in northern New York shared yield data for this project (Table 1). Yield monitor data were cleaned using the protocol developed at the NMSP (Kharel et al., 2019).

Table 1. The average yield (tons/acre) and area (acres) of corn silage fields in northern New York for each year; Satellite-Derived Yield Classification Maps and Yield Stability Zones for Management NNYADP project.

Farm	Field	Area (acres)	Yield in year 1	Yield in year 2	Yield in year 3
A	Field 1	231.7	20.4	21.3	25.8
	Field 2	70.9	20.8	17.9	28.2
	Field 3	116.2	23.7	29.5	23.6
	Field 4	236.0	21.8	24.0	24.9
B	Field 1	67.7	23.2	26.7	24.0
	Field 2	47.2	29.9	28.5	27.3
C	Field 1	45.7	23.6	27.0	20.9
	Field 2	30.4	20.4	21.9	16.4

Assigning yield classes:

Data from all three years and fields within Farm A were combined and then divided into five classes (Figure 1). The upper and the lower tails of the combined distribution were outliers and hence removed from the distribution. With the remaining distribution, the top and bottom 10% were identified as the high- and low-yielding areas and the remaining yield range was equally divided into three classes: medium-low, medium, and medium-high (Figure 1).

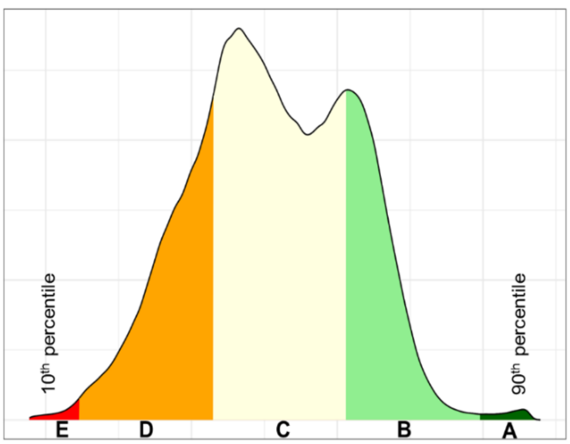

Figure 1: Categorizing yield data from Farm A across three years into different yield classes (A through E where A represents the top 10%, E is the bottom 10 percent, and B, C, and D are equally split into three classes between class A and E; Satellite-Derived Yield Classification Maps and Yield Stability Zones for Management NNYADP project.

Satellite data:

Satellite images for the fields in Table 1 were collected from the Planet's Dove constellation (https://www.planet.com/our-constellations/). The "Planet Explorer" plugin available within QGIS software allowed the download of the satellite images. The resolution of the downloaded images was 3x3 meters per pixel with red, green, blue, and near-infrared (NIR) spectral bands. These downloaded images were categorized into two types: bare soil images, captured within a 1-2 day window before planting, and canopy images, which were collected about weekly until the time of harvest. To streamline the process of extracting raw band data from these images, a Python plugin was developed. Once the raw band data became available, it was utilized for the computation of both soil and vegetation indices.

Retrieval of key predicting features:

- **Raw Satellite Bands**: The project includes raw satellite bands collected - the red, green, blue, and NIR bands, to explore their potential individual correlations with yield zones.

- **Vegetation Indices:** Vegetation indices (VIs) serve as indicators of canopy health, and different VIs are specialized in capturing distinct aspects of vegetation characteristics and well-being. We computed seven vegetation indices utilizing satellite bands (green, blue, NIR, and red) obtained from peak canopy images captured every week. A few of the VIs considered in our project include the normalized difference vegetation index (NDVI), excess green (ExG), green chlorophyll vegetation index (GCVI), and triangular greenness index (TGI), among others (Figure 2).

- **Soil Indices:** To determine soil health and condition, we utilized the original bands from bare soil imagery captured 1-2 days before planting. Five soil indices, including hue, saturation, brightness, coloration, and redness, were calculated to provide insights into soil properties.

- **Climate Variables:** Spatial data for 14 climate variables, including temperature, precipitation, wind patterns, and growing degree days, were procured from NASA Power (https://power.larc.nasa.gov/) and incorporated into our analysis.

- **Elevation:** We obtained elevation data from a digital elevation model (DEM) accessible through the New York State Geographic Information System (GIS) at a 1-meter per pixel resolution.

- **Landform Classes:** Using the DEM, we categorized landforms into six classes: 1) valley, 2) lower slope, 3) flat area, 4) middle slope, 5) upper slope, and 6) ridge. Since this feature consists of non-numerical classes, we used target encoding to convert to a numerical feature.

Machine learning (ML) for predicting yield classes:

- Feature Selection: In total, approximately 32 features (see Figure 4 for 18 of these features) were considered for predicting the assigned yield classes. Feature selection is a critical process for learning influential features and removing any irrelevant or redundant features before developing and training an ML model. Boruta feature selection method was employed and all the important features were included in the training of the ML model.

- Model Development: Three ML classification models such as the random forest, support vector machine, and k-Nearest Neighbor (kNN) are trained using the selected features. Each model undergoes hyperparameter tuning to optimize its performance and achieve the highest accuracy attainable for that particular model.

- Validation: The best-tuned model will be used to identify the number of fields required for achieving a reliable prediction model. For this, the model will be initially developed with 2 fields and more fields will be added based on the performance accuracy.

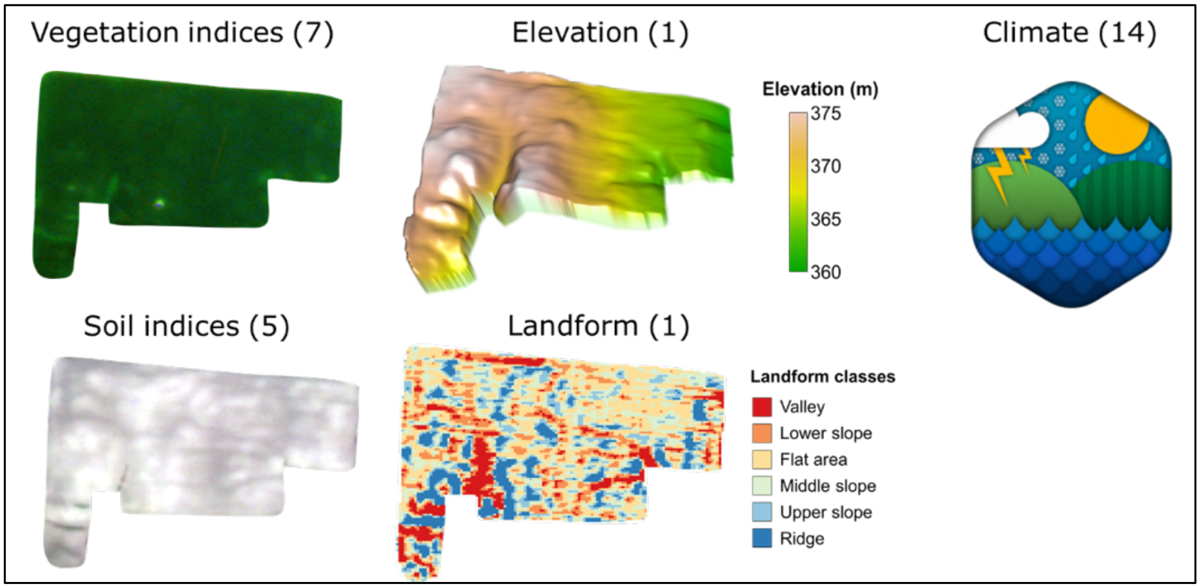

Figure 2. Features collected from satellite images, digital elevation maps, and climate for predicting yield stability zones; Satellite-Derived Yield Classification Maps and Yield Stability Zones for Management NNYADP project.

Results:

Optimum timing for satellite image collection

The normalized difference vegetation index (NDVI) metric is commonly used for assessing vegetation health and density based on satellite data, as it is calculated using the near-infrared (NIR) and red bands. We computed the average NDVI for each date in the dataset and created a time-series plot. **The higher NDVI values coincided with growth stages between R2 and R4, indicating peak canopy development** (Figure 3, highlighted in red). This pattern was consistent across all fields under consideration. This observation aligns with a prior study, which demonstrated that sensing between the R2-R4 growth stage leads to more accurate yield predictions (Sunoj et al., 2021). Therefore, we selected data from three peak canopy dates for feature selection and building ML models.

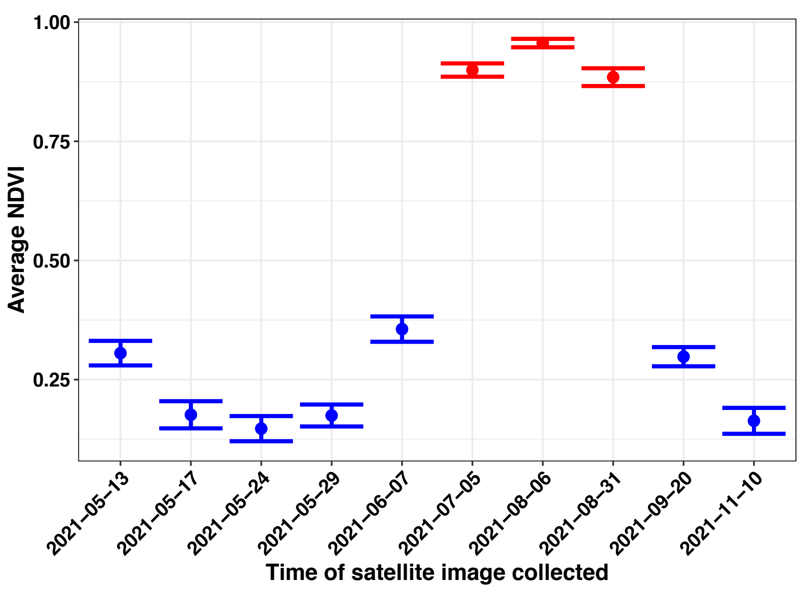

Figure 3. Time-series plot of average NDVI values showing the peak canopy stage between R2 and R4 vegetation stage in field 1 of Farm A; Satellite-Derived Yield Classification Maps and Yield Stability Zones for Management NNYADP project.

Yield Classification Maps and Yield Stability Zones for Management NNYADP project.

Important features driving yield stability zones

For Farm A, out of the 32 features, the Boruta feature selection method identified topographic information: elevation and landform as the most relevant feature (Figure 4). The heightened importance of topographic data may be attributed to its connection with soil temperature, moisture levels, and texture. Likewise, for Farms B and C, elevation consistently ranked among the top three important features.

Across all farms, several soil indices, including the saturation index, hue index, redness index, brightness index, and coloration index extracted from bare soil imagery, emerged as highly relevant for predicting yield classes. Among the vegetation indices, GCVI and GNDVI consistently stood out as the top relevant features for all silage farms. Both of these indices rely on the near-infrared (NIR) and green bands, highlighting the strong correlation between these individual bands and yield zones. None of the climate variables were relevant across farms, suggesting that the spatial resolution at which climate variables are gathered is inadequate for capturing field-scale variability. Zone maps derived from the yield classes showed high accuracy, suggesting satellite imagery can be used to derive yield stablity zones.

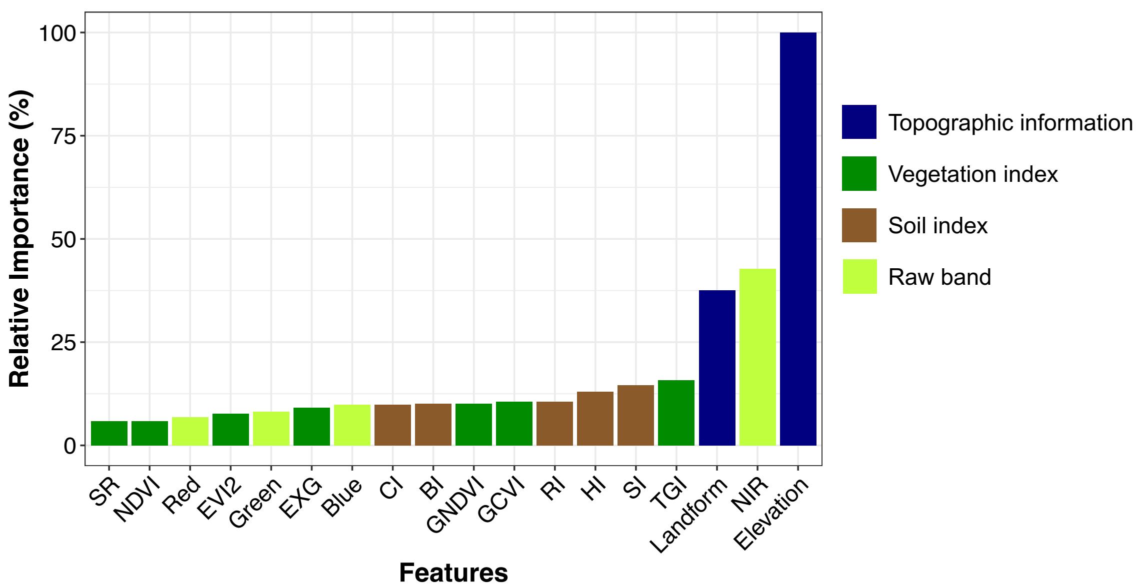

Figure 4. Relative importance of the relevant features selected using Boruta method for Farm A; Satellite-Derived Yield Classification Maps and Yield Stability Zones for Management NNYADP project. Key (alphabetical): BI – Brightness Index, CI – Coloration Index, EXG – Excess Green, EVI2 – Enhanced Vegetation Index (2 bands), GCVI – Green Chlorophyll Vegetation Index, GNDVI- Green Normalized Difference Vegetation Index, HI – Hue Index, NDVI – Normalized Difference Vegetation Index, NIR – Near infrared, RI – Redness Index, SI – Saturation Index, TGI – Triangular Greenness Index.

Determining the number of fields required to build a reliable model

Work is ongoing to develop ML classification models including Random Forest (RF), Support Vector Machine (SVM), and k-Nearest Neighbor (kNN). These models will undergo hyperparameter tuning and subsequent comparison to determine the most effective ML model.

For this objective in the project, the models will be trained using data from two silage fields and then tested on the remaining fields within one farm to assess classification performance. If the performance is not satisfactory, the model will be retrained using data from additional fields or until satisfactory performance is achieved.

Conclusions:

- Satellite image collection between R2 and R4 growth stages, representing peak canopy development, resulted in the best yield class estimations.
- The machine learning feature selection process revealed that topographic information (elevation and landform) were the most significant predictors of yield classes.
- Nearly all the soil indices considered proved to be important in predicting yield zones.
- Vegetation indices such as GCVI and GNDVI, which utilize green and NIR bands, were also influential.
- None of the climate variables emerged as top relevant features, likely due to insufficient spatial variability.

Further investigations are underway to utilize these selected relevant features to determine the number of fields required for building a reliable prediction model. Three classification models are being considered, and comparing their predictive performance across various farms will provide additional insights into the accuracy of the model.

Also see Determining the number of fields required to build a reliable model section (Results section) and Next Steps below.

Outreach:

- Ketterings, Q.M. (2024). Connecting the Dots: Dairy Sustainability, Value of Manure, Yield Stability Zones. NNYADP Research Update Meetings, Chazy, NY, March 13, 2024. 24 people.
- Ketterings, Q.M. and S. Srinivasagan (2023). The Power of Sensor Technology. Certified Crop Adviser and Field Crop Dealer Meeting 2023. November 29, 2023. Syracuse, NY. 50 min. ~100 people.
- Marcaida, M., S.N. Srinivasagan, S. Sunoj, and Q.M. Ketterings (2023). Using Precision Technologies to Improve Nutrient Efficiency and Save Fertilizer Costs. CCE Weekly Webinar Series on Maximizing Fertilizer Efficiency with Peak Fertilizer Prices. March 30, 2023. On-line. 1 hr. ~40 people.

Next Steps:

In addition to ongoing work noted earlier (Conclusions), next steps for this research include expanding on our current analyses to improve the yield class prediction models for individual fields and then evaluate how many fields of data in a farm should be used to train an ML classification model that can then be used to derive yield stability zones across the entire farm.

Acknowledgments:

We thank the farmers and crop consultants who shared yield data with us, allowing us to get a large database in place for this project. Co-funding was obtained from USDA-NIFA (Federal Formula Funds).

References:

- Kharel, T.P., S.N. Swink, A. Maresma, C. Youngerman, D. Kharel, K.J. Czymmek, and Q.M. Ketterings (2019). Yield monitor data cleaning is essential for accurate corn grain/silage yield determination. Agronomy Journal 111: 509-516. doi:10.2134/agronj2018.05.0317.
- Sunoj, S., Cho, J., Guinness, J., van Aardt, J., Czymmek, K. J., & Ketterings, Q. M. (2021). Corn grain yield prediction and mapping from Unmanned Aerial System (UAS) multispectral imagery. Remote Sensing, 13(19), 3948. https://doi.org/10.3390/rs13193948.

For More Information:

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