Northern New York Agricultural Development Program
2022 Project Report

Satellite-Derived Yield Maps
and Yield Stability Based Management Zones

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

Collaborators
- Eric Beaver, Mike Contessa, Champlain Valley Agronomics, Peru, NY
- Mike Hunter, Kitty O'Neil, Cornell Cooperative Extension North Country Regional Ag Team Field Crops and Soils Specialists, Watertown and Canton, NY
- Cornell University campus collaborators: Sunoj Shajahan (NMSP), Joe Guinness (Department of Statistics and Data Science), Louis Longchamps (School of Integrative Plant Sciences),
- Jan van Aardt, Rochester Institute of Technology

Cooperating Producers
- 3 Northern New York farms with corn silage and/or grain data

Background
Yield monitors are increasingly used to determine yield of corn silage and grain across northern New York and throughout the state. Yield monitor systems allow farms to map their yields and, once at least three years of data are gathered, to also build their own yield potential database and yield-stability based management zones, and conduct on-farm research to learn how to manage zones, using an easy to conduct single strip experiment.

While yield monitor systems have advanced our ability to map yield, it is well known that collection of accurate data requires frequent calibration in the field as well as post-harvest yield data cleaning. If all goes well, the information gained is extremely valuable but if a sensor breaks during the harvest season, all yield data may be lost, with no opportunity to recover information. In addition, not all farms can afford yield monitoring equipment. Thus, we need to look for other, less risky, and more affordable, ways to obtain yield estimates and generate zone maps.

Drone and satellite imagery could be used to generate satellite-derived stability maps, making this concept usable for farmers even without yield monitor data. Prior work has focused on drone-collected yield data and, while very promising (*Sunoj et al., 2021*), this remains a labor-intensive approach, and it requires calibration areas within fields (Nrich strips).

This NNYADP-funded project was initiated to evaluate how effectively we can use the satellite imagery to obtain yield maps and yield stability-based zone maps.

Methods

Part A: Generating yield maps from satellite imagery

Yield monitor data

We collected three years (2020-2022) of yield data from six fields within three farms (two dairy corn silage, one cash corn grain) and cleaned them using the protocol developed at NMSP. Each farm consisted of two fields that were close to each other, with one field larger than the other. Only fields that had reliable yield monitor data for three years of corn grain or three years of corn silage were selected. The details of each field's average yield per year and area are shown in Table 1.

Table 1. The average yield and area of corn grain and silage for six fields that were part of the NNYADP Satellite-Derived Yield Maps and Yield Stability Based Management Zones project 2020-2022.

Farm	Field	Area (acres)	Yield unit	Yield in year 1	Yield in year 2	Yield in year 3
Silage A	Field 1	68	tons per acre	21.0	24.3	21.8
	Field 2	47	tons per acre	27.1	25.9	24.7
Silage B	Field 1	46	tons per acre	24.5	18.9	21.4
	Field 2	30	tons per acre	19.9	14.9	18.5
Grain	Field 1	80	bushels per acre	139.9	159.0	132.0
	Field 2	31	bushels per acre	144.7	157.4	133.5

Satellite imagery of soil and canopy

The satellite images from Planet's Dove constellation (one of two constellations operated by Planet; https://www.planet.com/our-constellations/) were downloaded using the Planet Explorer tool in QGIS, a free open-source Geographic Information System (https://www.qgis.org/en/site/). The resolution of the imagery is 3m per pixel with four spectral bands (red, green, blue, and near-infrared [NIR]). The soil imagery was obtained one or two days before planting, and the canopy imagery was obtained approximately 75 days after planting each year.

Yield classes

The yield data from each field were combined across years and divided into five yield classes (Figure 1). The upper and lower tails of the combined distribution are removed by applying an outlier filter. With the remaining distribution, the high and low-yielding areas are identified as the top and bottom 10 percentile yield data. The remaining yield range was equally split into three classes representing med-low, medium, and med-high yielding areas.

Feature extraction

- **Soil indices** (5 features): The soil imagery from each year was used to calculate five different soil indices: brightness index, saturation index, hue index, coloration index, and redness index. These indices provide optical signals of the soil.

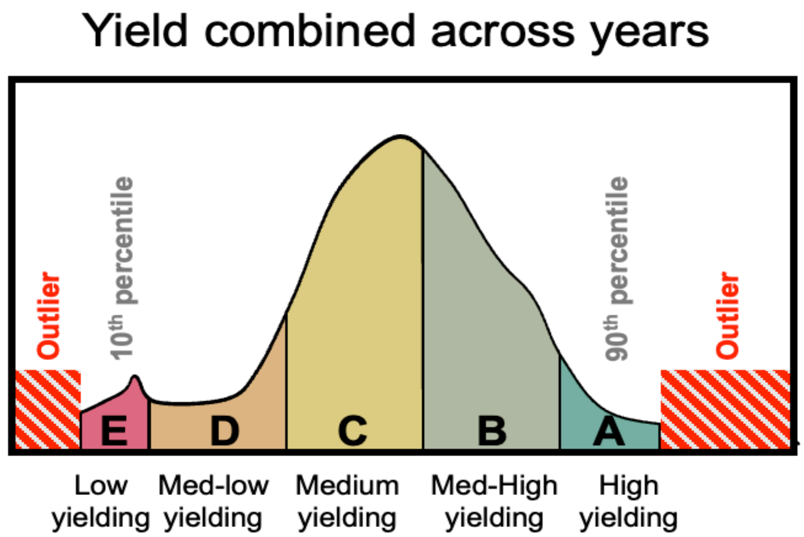

Figure 1: Categorizing yield data across three years into different yield classes (A through E where A represents the top 10%, E is the bottom 10 percent, and B, C, and D are equally split into three classes between class A and E.

- **Vegetation indices** (7 features): The canopy imagery from each year was used to calculate seven vegetation indices: normalized difference vegetation index (NDVI), green normalized difference vegetation index (GNDVI), enhanced vegetation index-2 (EVI2), simple ratio (SR), green chlorophyll vegetation index (GCVI), excess green (EXG), and triangular greenness index (TGI). These indices are canopy signals that represent crop health.

- **Digital elevation model** (1 feature): Along with the feature indices, 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 1m per pixel resolution.

- **Landform classes** (6 features): The DEM imagery was used to categorize each pixel in the image into either of the six landform classes: valley, lower slope, flat area, middle slope, upper slope, and ridge. This is a nominal variable, so they cannot be arranged or ranked in any specific order. To use them in the machine learning models, we performed "one-hot encoding," which transforms each class into a feature and assigns either a 1 or 0 value to that feature.

Machine learning modeling for yield maps

- **Feature selection**: We employed three feature selection methods to select the important features among the 19 features influencing the yield classes. The feature selection methods were Boruta, sequential lasso selection (SLS), and recursive feature elimination (RFE). The top seven features were selected based on the unified ranking from all the methods.
- **Model development**: We trained three machine-learning models: random forest classifier, logistic regression, and support vector machine classifier. For each model, hyperparameters (number of estimators, c-value, solver, etc.) were tuned to identify the best parameter values and model.

- **Validation approaches**: With the identified hyperparameters and models, the yield class predictions were made based on the following two approaches: (i) Year fit – training with two years of data and testing the left-out year; and (ii) Farm fit – training with one field and predicting the adjacent field in the same farm.

Part B: Generating yield stability maps from satellite imagery

The datasets we used for training yield models consisted of three years of yield data and satellite imagery for each field. The same datasets were used to develop yield stability maps from satellite imagery. The NDVI was calculated from the canopy imagery captured approximately 75 days after planting from each year (Figure 2A). Using the three years of data, each pixel's temporal average and temporal standard deviation were computed (Figure 2B). These NDVI values data were plotted to obtain the "weighted average" and "weighted standard deviation (SD)" of NDVI values to delineate the pixels into either of the four zones (Figure 2C), similar to yield-based stability zones.

Results:

Part A: Satellite-derived yield maps

Modeling results

- **Selected features**: Among the 19 features, the top seven features were GCVI, GNDVI, elevation, SR, brightness index, EVI2, and NDVI. Thus, top ranked features include soil (brightness index) and topography (elevation) features, as well as five vegetation indices. These top features were the same for grain and silage fields, but they ranked somewhat differently across farms.

- **Best performing machine learning model**: The random forest produced the best accuracy with a specific set of hyperparameters, followed by the logistic regression model and support vector machine classifier. Moreover, the random forest model was faster than the other two models in training.

Validation results

- **Year fit** – Training the models with two years of data and testing the left-out year resulted in accuracies ranging from 66% to 73% for grain fields and 62% to 78% for silage fields. These accuracies indicate that, as far as the training data cover the variability of the testing field, high accuracy can be attained. The potential application of this approach will be to use the available years of yield data to train the model and predict the missing years' or past years' data.

- **Farm fit** – The accuracy of the models trained with one field each year and testing the adjacent field resulted in comparable accuracy. The accuracy ranged from 73% to 76% for the grain farm and 71% to 82% for silage farms. The higher testing accuracies compared to the year fit approach are likely because the training and testing data were obtained within the same year, which would have the same weather and reflectance conditions. In most cases, training with a bigger field and testing the smaller adjacent field resulted in higher accuracy than the other way. These results show the potential of expanding the yield class prediction to the entire farm by training with only a few major fields with yield monitor data.

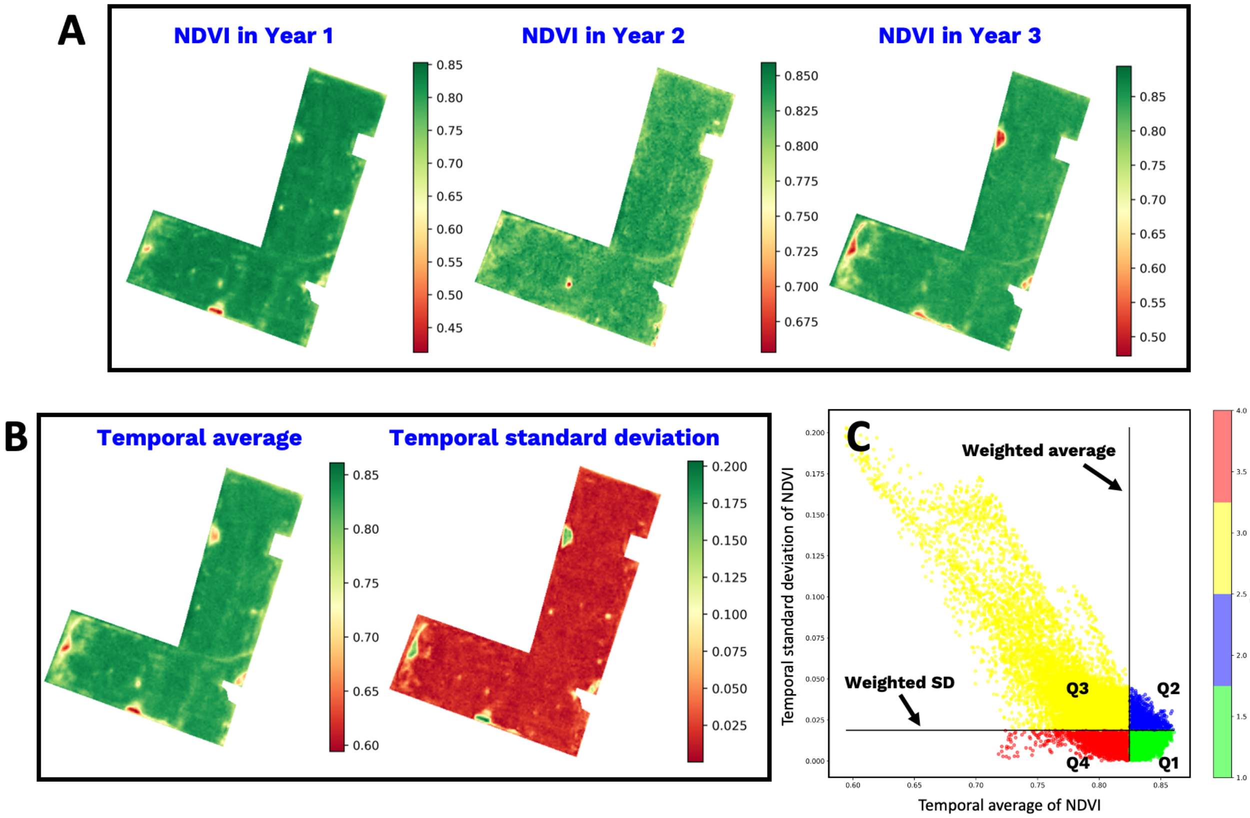

Figure 2: Process of obtaining stability maps from satellite imagery. (A) NDVI maps from satellite imagery obtained approximately 75 days after planting; (B) the temporal average and standard deviation maps calculated from three years of NDVI data; and (C) scatter plot showing the distribution of temporal average NDVI and temporal standard deviation to obtain the weighted average and SD and delineating satellite-derived stability zones.

Part B: Satellite-derived yield stability maps

The satellite-derived yield maps looked similar to the actual yield stability maps (see Figure 3 for an example). The accuracy of the classification of stability zones by our developed approach at approximately 75 days after planting was about 84% (Figure 3). Most of the areas representing Zone 1 (green) and 3 (yellow) were correctly classified. Misclassification was mainly from Zone 2 (blue) and 4 (red).

We also used earlier (approximately 65 days after planting) and later-stage (approximately 83 days after planting) images as a preliminary trial. We saw slight improvements in the accuracy (to 85%) for later-stage images and a reduced accuracy of 78% for the earlier-stage image. Selecting the right imagery time remains a potential research question to be answered. The Planet's satellite images have a daily revisitation rate, making it a great data source for deriving more inferences for satellite-derived yield stability maps.

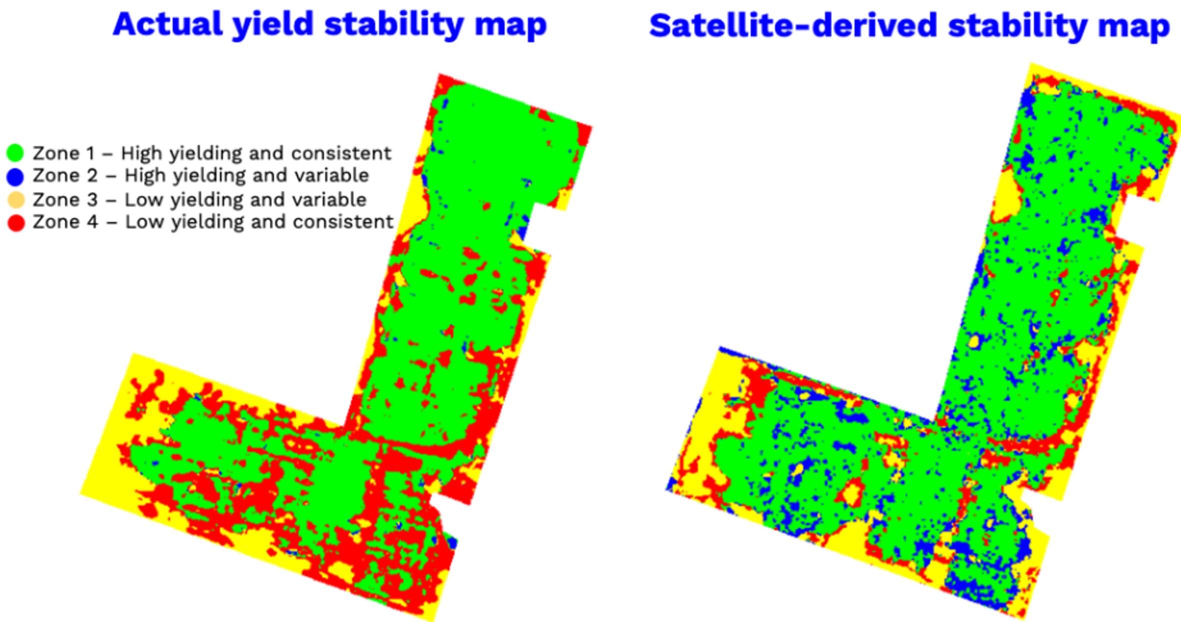

Figure 3: Example of a comparison between satellite-derived stability map and the actual yield stability map.

Conclusions/Outcomes/Impacts

The "farm fit" approach gave an accuracy ranging from 71% to 82%, which shows promise in predicting the yield classes of all fields on a farm by training with only a few larger fields. The "Year fit" approach gave an accuracy of 62% to 78%, which suggests that determining yield for missing years or past years where yield was not collected is feasible as well, but estimations will be slightly less accurate.

The classification accuracy of yield stability maps was about 84% from the satellite imagery captured on approximately 75 days after planting which shows potential for developing an alternative technique to derive stability maps for the farms that do not have yield monitor data.

Work is ongoing to answer further questions about year fit, and farm fit including the number of years of data and the number of fields in a farm should be used for training of machine-learning models. We are also exploring if machine-learning models can be employed to improve the accuracy of satellite-derived stability maps. In addition, we initiated work to determine accuracy of satellite imagery-derived yield (not classifications as we did here, but actual yield levels per 6 x 6 ft grid cell.

Outreach

Project updates were shared with farmers and farm advisors at various meetings including the 2022 Northeast Region Certified Crop Advisor and Field Crop Dealer Meeting, and Northern New York Crop Congress. Additional presentations are scheduled for the winter of 2023. The farms that shared data received farm-specific yield reports.

Next Steps

With the help of interns, we aim to expand on our current analyses to improve the yield class prediction models, as well as answer some key questions of how many years of data and how many fields in a farm should be used to train the model. We also aim to develop a pipeline for data analysis. We are actively pursuing collaborations and funding to expand current findings to statewide evaluations.

Acknowledgments

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

Reference

Sunoj, S., J. Cho, J. Guinness, J. van Aardt, K.J. Czymmek, and Q.M. Ketterings (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.

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

- 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), Department of Animal Science, Cornell University, qmk2@cornell.edu, 607-255-3061, http://nmsp.cals.cornell.edu.