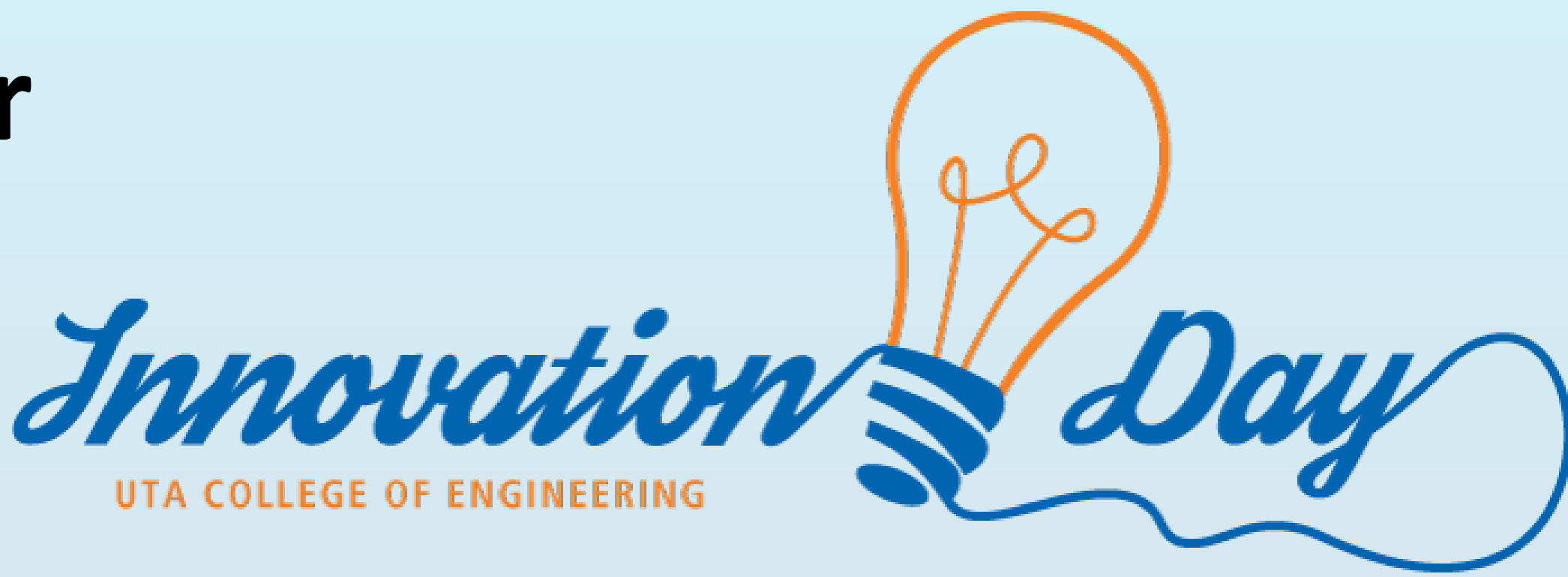


# #62: Monitoring and Predicting Crop Yield for Site-Specific Zone Management

Jennifer Nguyen<sup>1</sup>, Sami Rezwan<sup>1</sup>, Long Xuan Hoang Duong<sup>2</sup>  
Darya Abbasi, Ph.D. Candidate<sup>1</sup>, Kwok Tsui, Ph.D.<sup>1</sup>, Yuan Zhou, Ph.D.<sup>1</sup>

- <sup>1</sup> Department of Industrial, Manufacturing, & Systems Engineering
- <sup>2</sup> Department of Computer Science & Engineering



## Summary

Monitoring and predicting crop yield dynamics at fine spatial scales are essential for supporting site-specific zone delineation and management in precision agriculture (PA). When optimizing management choices, such as seed density and fertilizer application rates, the spatial and temporal heterogeneity of characteristics in a specific field presents challenges. Instead of managing whole fields as a single unit, PA management is customized for small areas (zones) within fields via observing, measuring, and responding to multiple agricultural variables such as crop yield, soil properties, terrain attributes, and weather.<sup>1</sup>

This research employs multiple data-driven analytical approaches to uncover crop yield patterns and identify the key factors influencing agricultural productivity.

## Background

The United States is one of the largest producers and consumers of agricultural commodities in the world. In 2023, nearly 0.8% of U.S. GDP was contributed by the output of American farms, and about 5.5% of GDP is contributed by industries that rely on agricultural commodities.<sup>2</sup> Therefore, it is extremely important to continuously grow agricultural productivity.

### Research Objectives:

- Perform a spatiotemporal variability analysis to monitor within-farm yield dynamics over time.
- Apply machine learning (ML) algorithms to predict crop yield productivity and identify the most influential predictive factors.

## Study Setting and Agricultural Data

**Study Setting:** A farmer’s field (28.6 hectares) in northeast Oklahoma. All the agricultural data used in this research were provided by our collaborators from USDA Agricultural Research Service (ARS).

**Yield Monitoring Data (2016-2020):** Yield volumes (bushels/acre) and other crop attributes (e.g., moisture, location), derived from the areas covered by the tractor every second. (\* area = distance travelled per second by swath width)

**Soil Properties:** Texture (sand, silt, clay), normalized difference vegetation index (NDVI), chemical elements (e.g., C, Ca, N, Zn), electrical conductivity (EC), pH, etc.

**Terrain Attributes:** Aspect, surface elevation, flow accumulation, slope, openness indices (positive, negative), soil water index, topographic position index, topographic wetness index, etc.

**Weather Data (2016-2020):** Precipitation, solar radiation, temperature (min, max).

## Literature Review

**Table 1.** Summary of the literature review on crop yield prediction.

|                          |                                                                                                                           |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------|
| <b>Study Setting</b>     | Counties, crop regions, farms; typically sectioned into points                                                            |
| <b>Data Used</b>         | Historical crop yield, weather (rainfall, min/max temperatures), NDVI, MODIS satellite images, soil properties (salinity) |
| <b>ML Models</b>         | Random Forest (RF), Polynomial Regression, Support Vector Regressor (SVR), Deep Neural Networks (DNN), etc.               |
| <b>Model Performance</b> | RMSE and R <sup>2</sup> are common performance metrics                                                                    |

## Methodological Overview

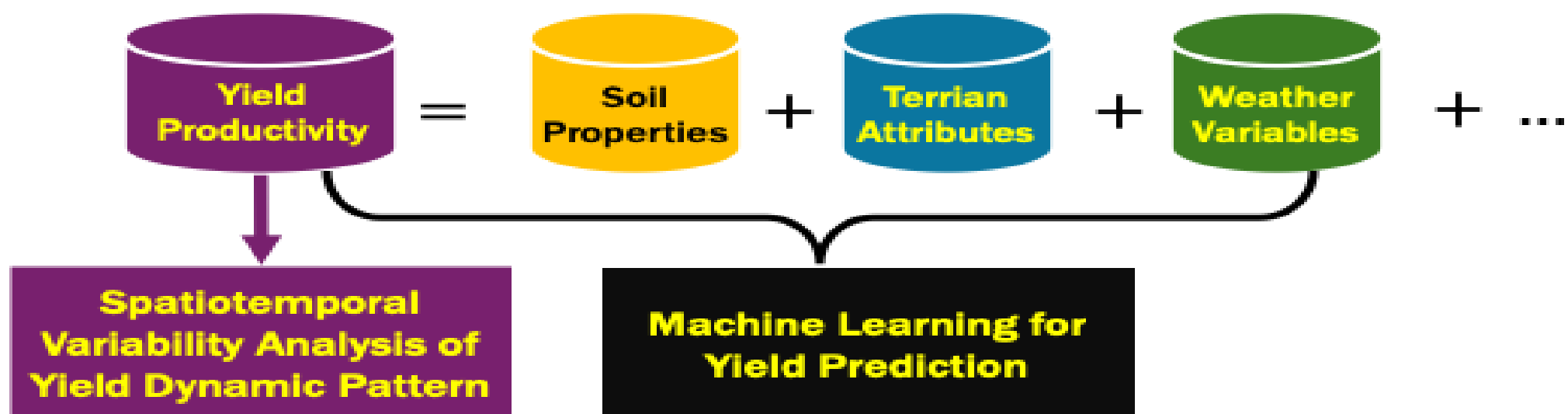


**Fig. 1.** Overview of research methods

## Data Pre-Processing

**Data Cleaning:** (1) Unit conversion from bu/ac to t/ha, (2) Identify outliers using  $\pm 3$  standard deviations (SD), (3) Impute outliers using inverse-distance weighted (IDW)3 method.

**Grid Sampling:** (1) Divide the whole farm field into a set of 10m-by-10m grids, (2) Calculate yield of individual grids (using mean).

**Data Normalization:** Use the min-max normalization method to rescale the grid-level yield data to a 0-100 range.

**Data Aggregation:** Integrate the normalized yield data and other agricultural data—including soil properties, terrain attributes, and weather variables—at the grid level.

## Yield Monitoring

### Yield Data Visualization and Descriptive Statistics:

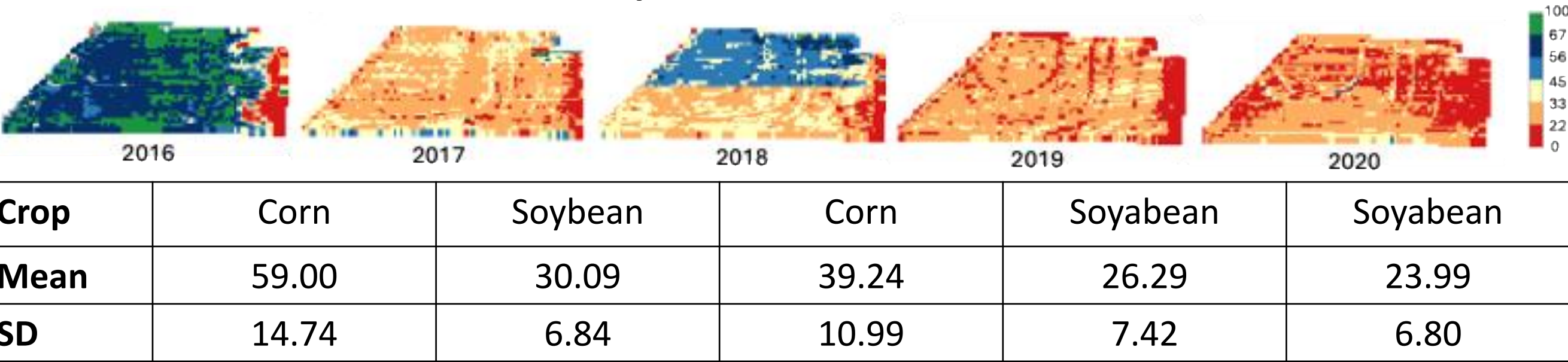


**Fig. 2.** Yield maps of the study farm (unit = tons/hectare or t/ha)

### Yield Dynamic Pattern: Yield Stability<sup>4</sup>

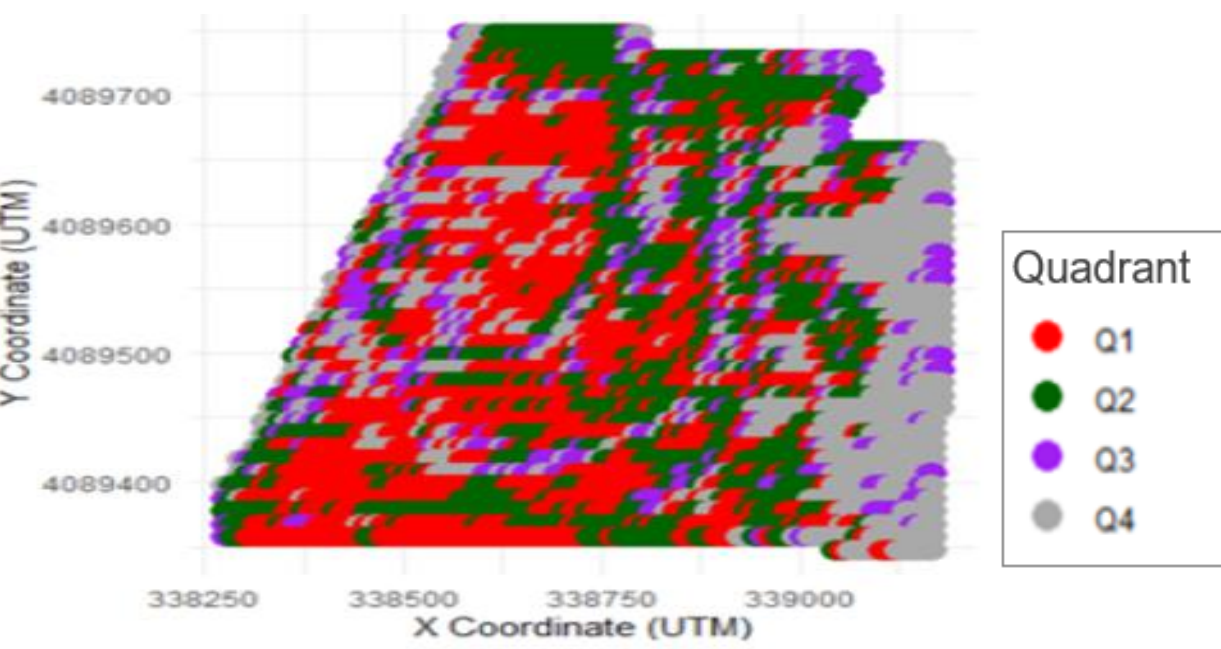


**Fig. 3.** Yield stability visualization

|           |                              |
|-----------|------------------------------|
| <b>Q1</b> | High Yield; Low Variability  |
| <b>Q2</b> | High Yield; High Variability |
| <b>Q3</b> | Low Yield; High Variability  |
| <b>Q4</b> | Low Yield; Low Variability   |

High/Low Yield:  $>/<$  Whole-Field Mean

High/Low Variability:  $>/<$  Whole-Field SD

## Yield Prediction

- Response variable [**Y**]: normalized yield productivity (t/ha) of a specific crop, e.g., corn, soybean, derived across all years during the data collection period.
- Predictor variables [**X**<sub>1</sub>, **X**<sub>2</sub>, ...]: soil properties, terrain attributes, and weather data.
- ML methods [**Y** = *f*(**X**<sub>1</sub>, **X**<sub>2</sub>, ... )]: Decision Tree, Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost) Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost).

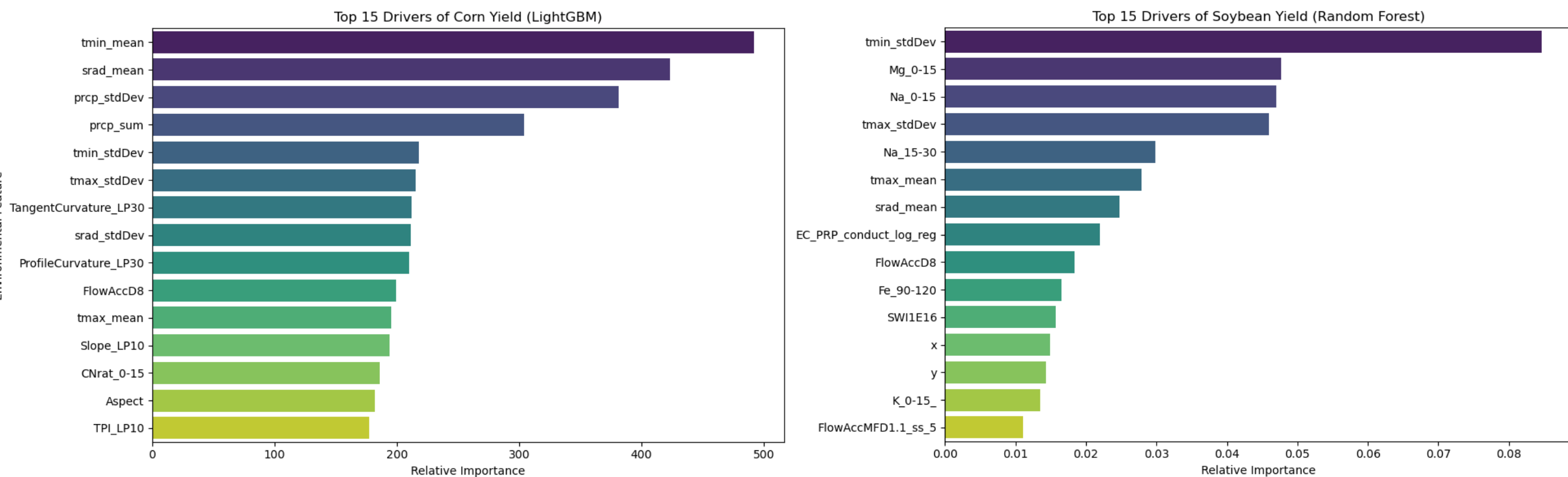


**Fig. 4.** Important predictive factors for predicting corn (*left*) and soybean (*right*) based on LightGBM for corn and Random Forest for soybean.

## Conclusions and Future Work

### Conclusions:

- Developed a process for preparing agricultural data to support yield monitoring and prediction for site-specific zone delineation and management.
- Conducted a grid-level spatiotemporal variability analysis to assess yield stability.
- Performed multiple ML algorithms to predict crop yield productivity and identify the most influential predictors.

### Future Work:

- Improve the methods for identifying management zones such that more meaningful information can be obtained for informed decision-making.
- Explore the use of spatially explicit modeling techniques to better account for the spatial autocorrelation and geographic dependencies in grid-level yield data.

## References

- Stafford, J. V. (2000). Implementing precision agriculture in the 21st century. Journal of agricultural engineering research, 76(3), 267-275.
- U.S. Department of Agriculture’s Economic Research Service (ERS)
- Panigrahi, N. (2023). Inverse distance weight. In Encyclopedia of mathematical geosciences (pp. 666-672). Cham: Springer International Publishing.
- Kharel, T. P., Maresma, A., Czymmek, K. J., Oware, E. K., & Ketterings, Q. M. (2019). Combining spatial and temporal corn silage yield variability for management zone development. Agronomy Journal, 111(6), 2703-2711.

*This project was supported by USDA Agricultural Research Service (ARS), which provided the agricultural data used in this research.*