Assessing the Impact of Biophysical Parameters on Corn Yield Potential in the Upper Midwest

Team Data Deere

Ryan Leahy, Ben Schwartz, Haleigh Gronwold, Seung-Taek (Stan) Lee

JOHN DEERE

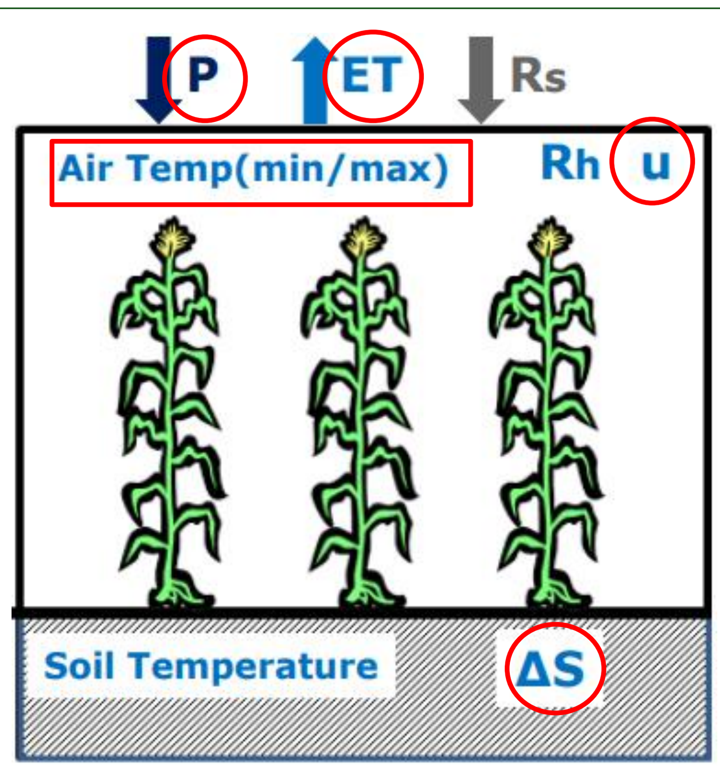
Abstract

Our first research objective was to analyze the impact of similar biophysical parameters on corn yield in the Upper Midwest United States. We hypothesized that precipitation was the driving factor influencing corn yield in this region. We created a Python script that gathered U.S. weather station and crop yield data and analyzed it through manipulation in R Studio, plotting the parameters and their respective corn yields from each county in the region from 2002 to 2012. With our analysis, we found that mean monthly temperatures had on average a higher positive correlation with corn yield than precipitation. The second objective of our research was to predict corn yield in the region based on an input vector of biophysical parameters. To achieve this, we created a predictive model in Python using Keras, a high-level neural networks API running over TensorFlow. With our model, we predicted corn yield with an absolute percentage error between 18% and 21%. These programs could be used in the future to predict crop growth in other regions, with its accuracy dependent on the data given.

Introduction

- Corn is a prevalent crop in the Upper Midwest region of the United States and its production is a cornerstone of the region's economy.
- Climate change has had detrimental effects on the agricultural industry and the ability to predict environmental changes in future growing seasons based on data analysis is crucial to John Deere [1].
- Variations in climate can affect corn yield through biophysical properties, including precipitation, surface temperature, soil evaporation, soil moisture, and wind speed.
- We gathered data on the parameters and corn yield in Michigan, Wisconsin, Iowa, and Minnesota between 2002 and 2012, a time frame that experienced both wet and dry years, offering a variety of biophysical conditions for analysis.

Figure 1: Parameter Visual and Selection



Hypothesis

Total monthly precipitation is the key variable that influenced corn yield in the Upper Midwest region of the United States between 2002 and 2012.

Methods

Data Retrieval:

Initial datasets collected: precipitation (PRCP), average temperature (TAVG), minimum temperature (TMIN), maximum temperature (TMAX), evaporation (EVAP), wind speed (AWND)

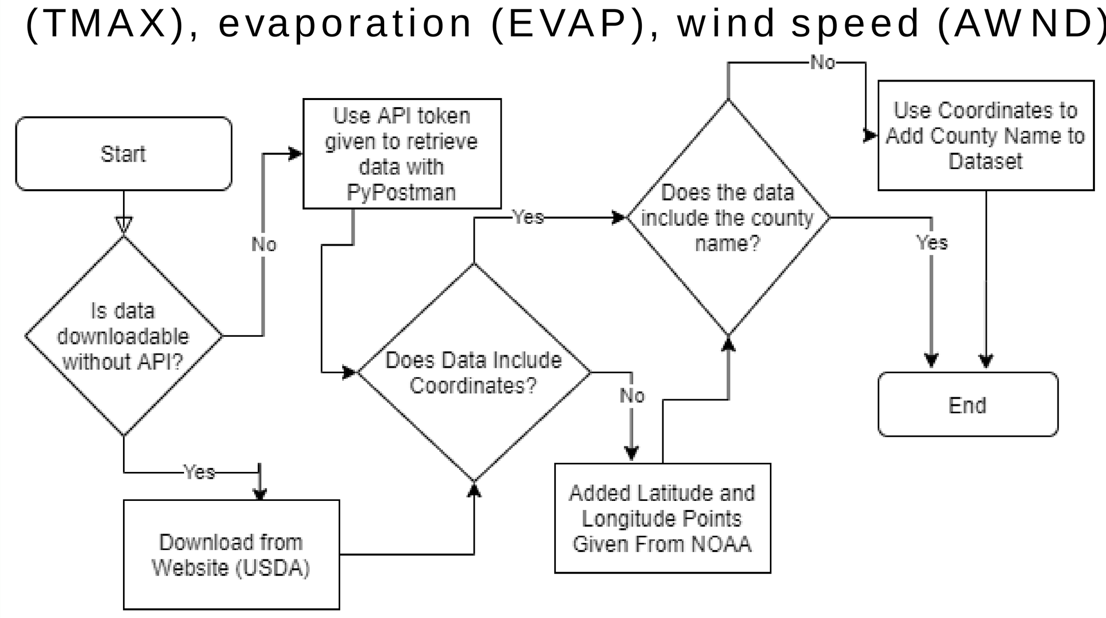


Figure 2: Parameter Retrieval Flowchart

Correlation Analysis:

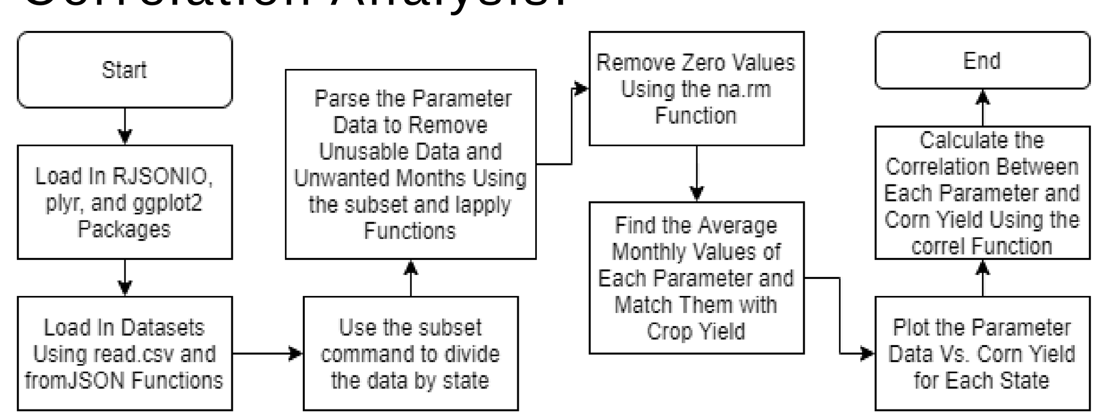
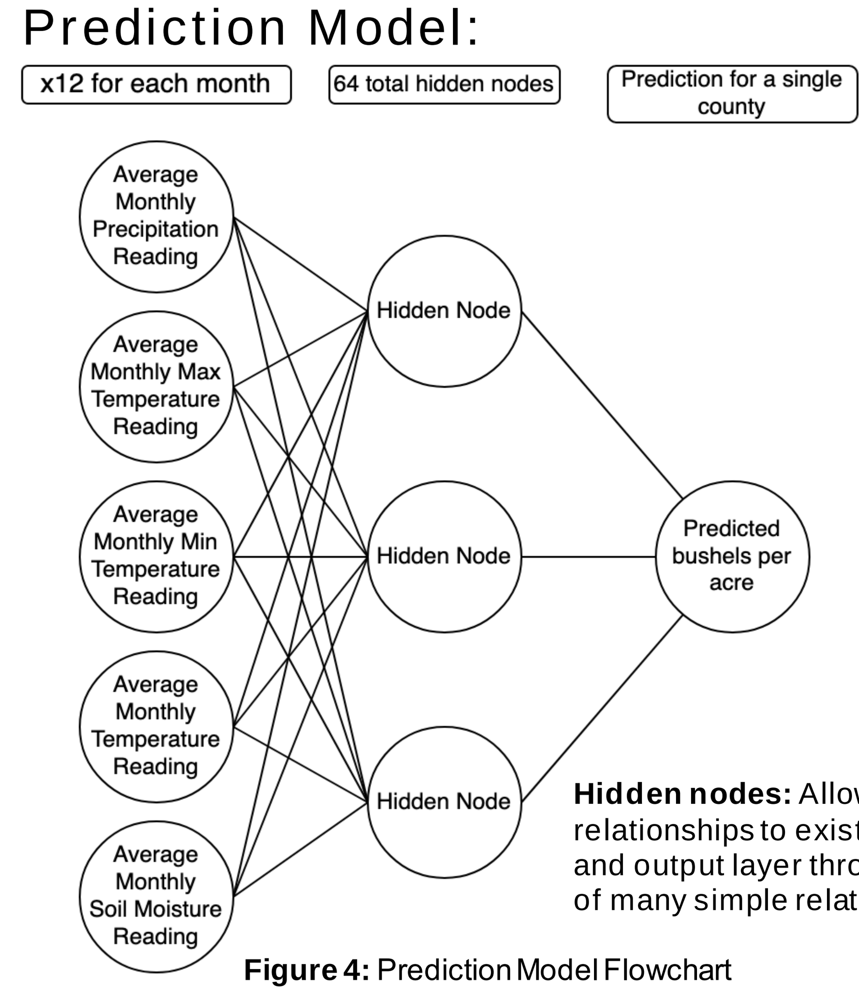


Figure 3: Correlation Analysis Flowchart

Prediction Model:



Hidden nodes: Allow for complex relationships to exist between the input and output layer through the combination of many simple relationships

Figure 4: Prediction Model Flowchart

Results

Legend: Iowa, Michigan, Minnesota, Wisconsin

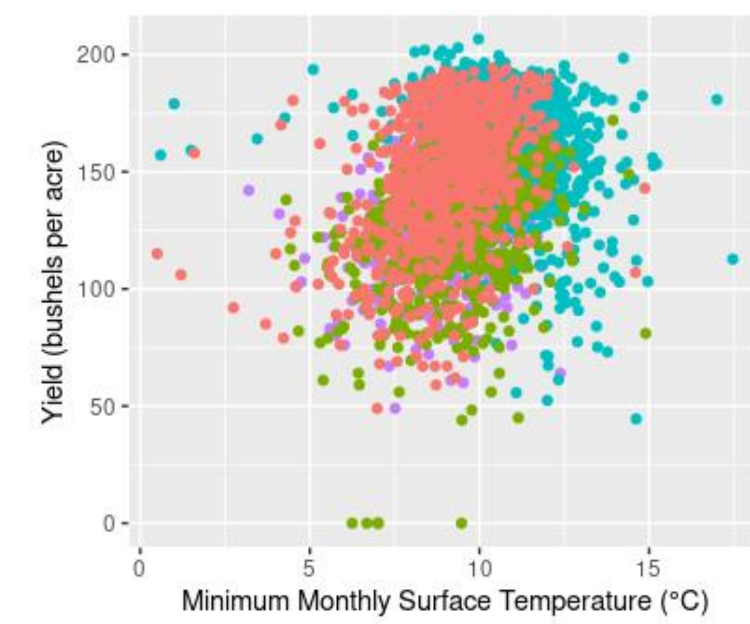


Figure 5: Correlation between Corn Yield and Minimum Monthly Surface Temperature (R = 0.298)

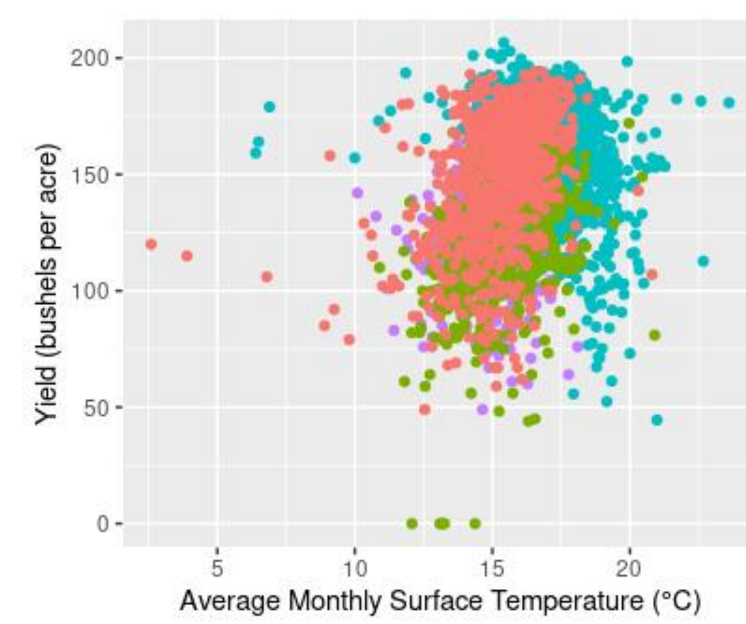


Figure 6: Correlation between Corn Yield and Average Monthly Surface Temperature (R = 0.282)

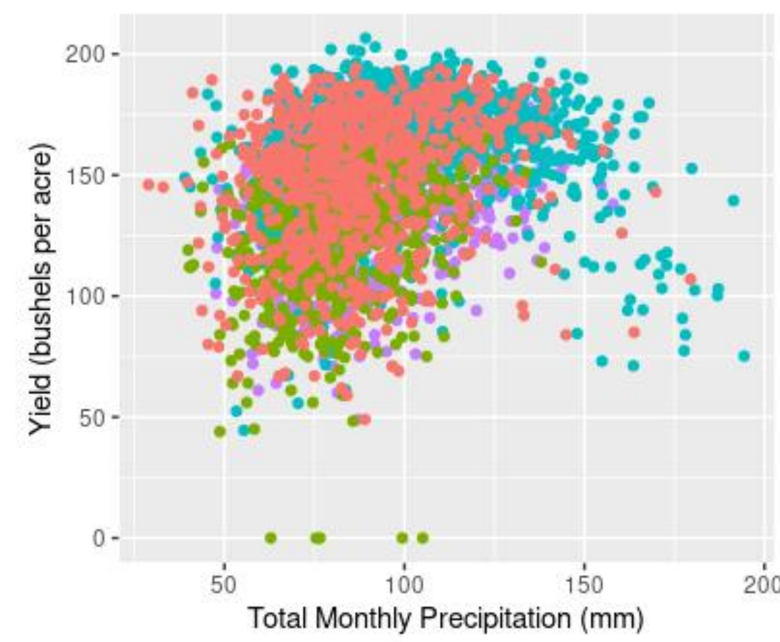


Figure 7: Correlation between Corn Yield and Total Monthly Precipitation (R = 0.264)

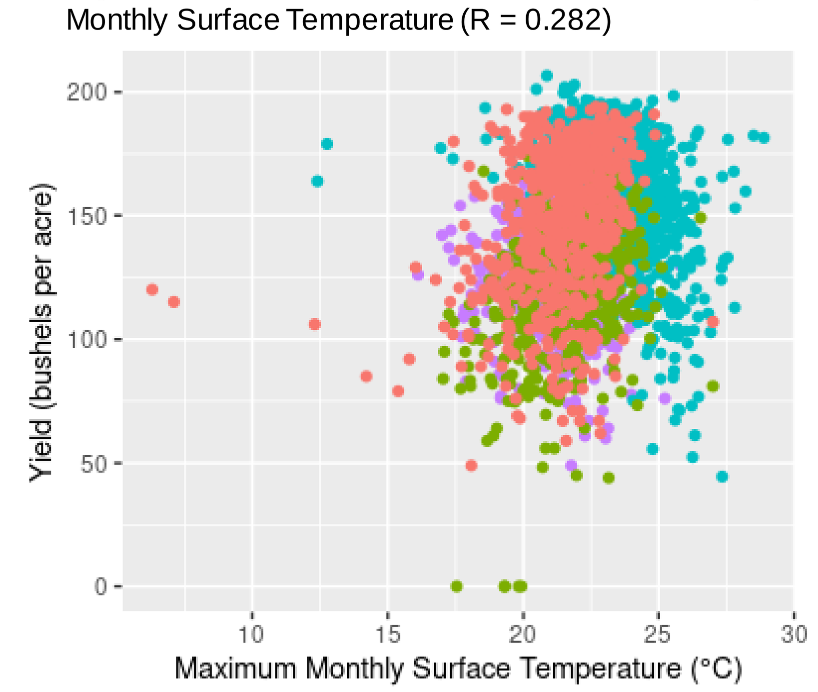


Figure 8: Correlation between Corn Yield and Minimum Monthly Surface Temperature (R = 0.244)

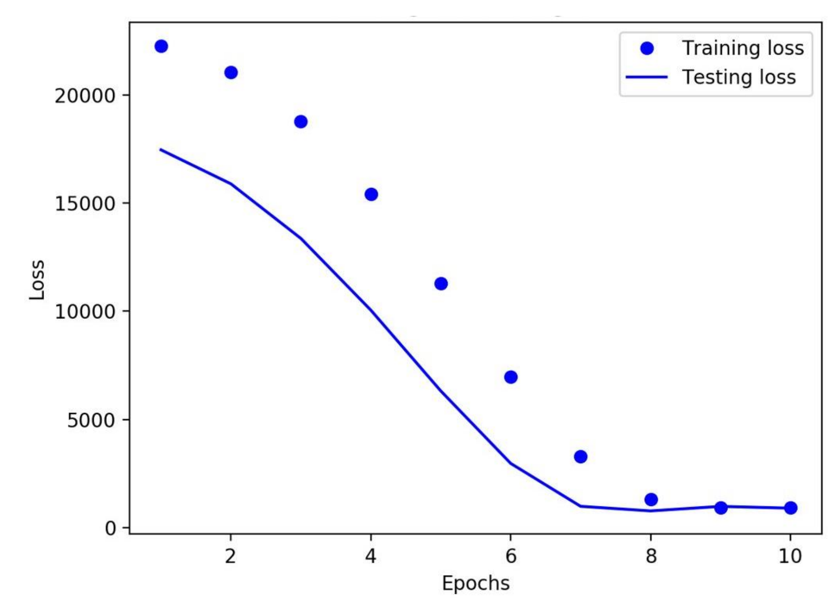


Figure 9: Training and Testing Loss of Prediction Model

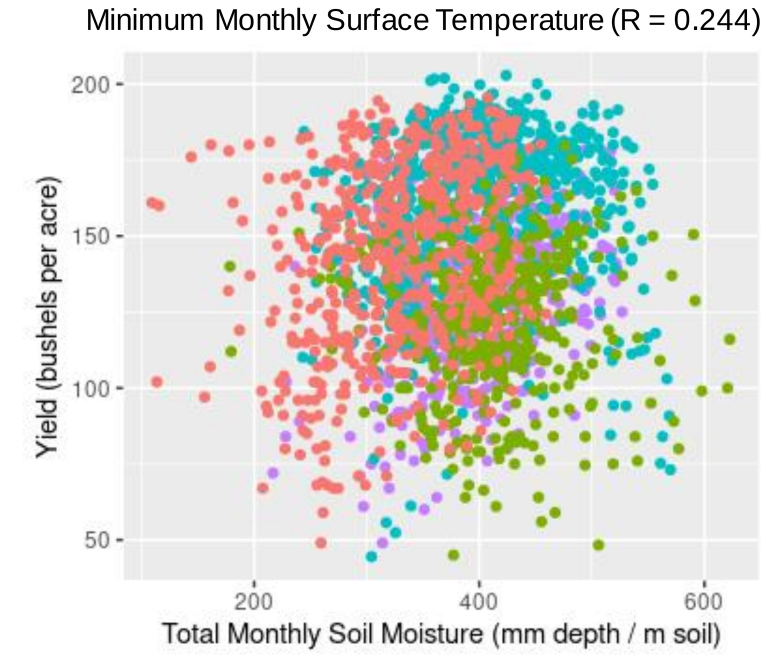


Figure 10: Correlation between Corn Yield and Total Monthly Soil Moisture (R = 0.113)

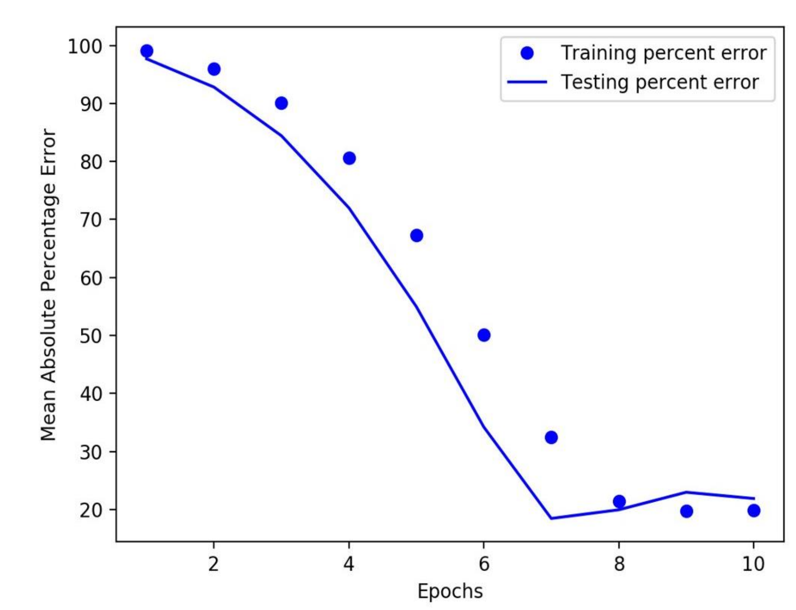


Figure 11: Accuracy of Prediction Model

Discussion & Analysis

Correlation Analysis

- Removed Monthly Average Wind Speed and Total Monthly Evaporation from analysis due to lack of data.
- The order of correlation from highest to lowest is Minimum Monthly Surface Temperature, Monthly Average Temperature, Monthly Average Precipitation, and Maximum Monthly Average Temperature. After adding in Total Monthly Soil Moisture data, it had the lowest correlation.
- Temperature correlation seems reasonable, since the minimum temperature has a large impact on kernel size and success in early growth stages [3].
- Low correlations due to high variability of monthly data and vagueness of yield data.

Prediction Model

- Trained data on years 2002-2010 Upper Midwest data.
- Tested our model on 2011-2012 Upper Midwest data.
- Model had a mean absolute percentage error of 20%, with variances caused by the weights in the neural network that are randomized at beginning of the process.
- Validated model on 2002-2012 Illinois data, with mean absolute percentage error of 15% between the predicted county yield and the recorded yield for that year.
- When adding soil moisture data, model did not improve
- Benefit of using a neural network is its ability to define complex relationships with multiple variables.
- Drawback of using a neural network is since it is black box model, it is difficult to understand how each input node affects the output.

Conclusions & Next Steps

- Hypothesis was incorrect
- There does not appear to be one key parameter driving corn yield in the Upper Midwest
- Prediction model had substantial error (around 20%)
- Due to broad geographical data and high variability of parameters
- Took longer to train with the addition of soil moisture

Next Steps:

- More geographically accurate data
- Use Python package for in depth analysis of model's interpretation of the data
- Test accuracy of model on each individual county

References

[1] Agriculture in the Midwest. (2017). Retrieved from https://www.climatehubs.usda.gov/hubs/midwest/topic/agriculture-midwest

[2] Keras overview. (n.d.). Retrieved from https://www.tensorflow.org/guide/keras/overview

[3] Lindsey, A., & Thomison, P. (n.d.). Agronomic Crops Network. Retrieved from https://agcrops.osu.edu/newsletter/corn-newsletter/2018-24/night-temperatures-i

Acknowledgements

Our team would like to thank both The Data Mine and John Deere for the opportunity to work on this project. Specifically, the feedback we received from Dr. Mark Ward, Margaret Betz, Ellen Gundlach, Dr. Angela Bowman, David Glass, and Gaylia Ostermeier was critical to our success.