A Growth Stage Centric Approach to Field Scale Corn Yield Estimation by Leveraging Machine Learning Methods From Multimodal Data

Lucas Waltz[1], Sushma Katari[1], Taylor Dill[3], Canaan Porter[2], Osler Ortez[3], Laura Lindsey[3], Arnab Nandi[2], Sami Khanal[1]

[1]Department of Food, Agricultural, and Biological Engineering, The Ohio State University, Columbus, Ohio, USA

[2]Department of Computer Science and Engineering, The Ohio State University, Columbus, Ohio, USA

[3]Department of Horticulture and Crop Science, The Ohio State University, Columbus, Ohio, USA

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

Field scale yield estimation is labor-intensive, typically limited to a few samples in a given field, and often happens too late to inform any in-season agronomic treatments. In this study, we used meteorological data including growing degree days (GDD), photosynthetic active radiation (PAR), and rolling average of rainfall combined with hybrid relative maturity, organic matter, and weekly growth stage information from three small-plot research locations to predict corn yield. Daily time-series data was transformed from the time domain to the growth stage domain and subsequently trained using a Long Short-Term Memory (LSTM) machine learning model. The results showed a mean absolute error (MAE) of 22.6 bu/acre on a five-fold cross validation set. When trained with location-specific data, the model achieved an MAE of as low as 19.2 bu/acre.

Keywords.

Precision Agriculture, Multimodal Data, Machine Learning, Unmanned Aerial Systems, Crop Phenotyping

1. Introduction

Corn is the largest crop in terms of acreage in the United States. It is extensively used in food, feed, and fuel products. Monitoring and estimating corn yield accurately in-season at the field scale level can help to inform the grain marketing efforts of farmers. Furthermore, estimates of yield during the growing season can inform the profitability of agronomic treatments such as nitrogen and fungicide applications.

Corn yield can be estimated manually using yield components such as the number of ears, number of kernel rows, and kernel weight (Licht, 2017) obtained from the field. Calculating these parameters manually is a time-consuming process. The labor-intensive nature of in-field yield estimation typically limits the sample size to a small portion and happens too late in the season to inform the profitability of any agronomic treatments.

However, with advancements in remote and proximal sensing capabilities and machine learning techniques, there is an opportunity to predict in-season crop yield based on the current year's weather and soil inputs which can provide in-season yield estimates sooner in the growing season and with greater accuracy. This paper takes a step towards the goal of in-season yield estimation by outlining an approach to accurately estimate yield from the combination of meteorological and phenological data from the prior year.

2. Related Works

In past studies, crop yield models mainly focused on estimating yield based on data acquired through optical, multispectral, and weather sensors. However, there are limited studies that estimate yield based on growth stage. The following sections briefly explain some commonly used methods for predicting crop yield.

2.1 Machine Learning Models

Various machine-learning strategies are used in monitoring within-field yield variability using satellite optical data namely Sentinel-2 data (Crusiol et al. 2022). High accuracy of crop yield was observed when the images at the R5 phenological stage of soybean were considered in support vector regression (SVR) and partial least square regression (PLSR) models demonstrating the importance of the crop phenological stage in crop yield estimation. However, the study did not account for the environmental conditions of the field which makes it difficult for generalization. Similarly, another study integrated a deep learning and machine learning model for estimating corn yield. This integrated network model achieved an RMSE of 6.298 in measuring the crop yield index indicating the capabilities of deep learning models (Kuwata and Shibasaki 2015).

Jiang et al. (2020) estimated corn yield at the county level using a combination of meteorological data and satellite-based vegetation indices. An LSTM model was developed and was found to outperform least absolute shrinkage and selection operator (LASSO) and random forest (RF). The LSTM model utilized five different growth phases as inputs to the LSTM model and was able to achieve RMSE of 1.48 Mg/ha (23.5 bu/acre) with sample size of n = 6,592 across 10 years.

Shook et al. (2021) estimated soybean yield using an LSTM with temporal attention approach based on data from Uniform Soybean Trust (UST) resulting in a dataset of n = 103,365 and achieving a mean absolute error (MAE) of 6.17 bu/acre using meteorological variables. This corresponds to an approximately 14% error in yield given that the range of yield in the study was between 33 and 55 bu/acre.

These studies indicate that machine learning models can be effective in estimating yield.

2.2 Process-based Models

There exist several process-based models such as Decision Support System for Agro-Technology Transfer (DSSAT), Agricultural Production Systems sIMulator (APSIM), and World Food Studies (WOFOST) for estimating crop yield under various crop and weather conditions

(Sargun and Mohan 2020)(Huang et al. 2019). These models simulate crop growth development and yield based on genetic features and meteorological conditions.

Even though these models can be tuned and calibrated to predict crop yields accurately, they can be very crop and field-specific making it difficult to replicate them for other crop fields without further tuning and calibration. Furthermore, these models have not completely accounted for the relationship between meteorological conditions and crop growth stage timing which is an important factor for determining crop yield (Zhou et al. 2017). Considering the timing of growth stage in combination with meteorological conditions including temperature, precipitation, and photosynthetic active radiation can make yield predictions more robust and generalized.

2.3 Model-guided machine learning

Model-guided machine learning uses the outputs from a process-based model as inputs into a machine learning model. One of the studies implemented a framework for monitoring in-season crop phenology using a biophysical crop model (DSSAT) for guiding neural networks (Worrall et al. 2023). The results showed that neural networks guided with DSSAT estimated the progression of phenological stages better compared to the unguided and crop model-only method.

3. Data Sources and Types

Figure 1 shows data types collected in this study.

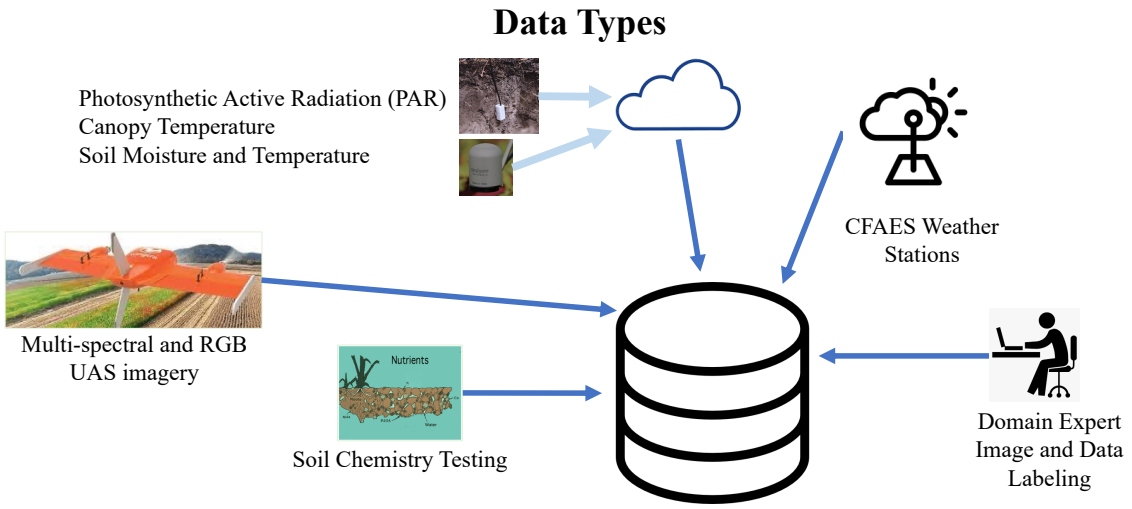

Figure 1: Summary of Data Types

3.1 Data Sources

The data sources underpinning this effort originate from three Ohio State agricultural research stations geographically dispersed across Ohio. They include Western Agricultural Research Station in Clark County, Northwest Agricultural Research Station in Wood County, and Wooster Campus in Wayne County. Each site included 80 plots. The experiment was a split-plot randomized complete block design with four replications of each treatment. Main plot factor included five planting dates spaced approximately every two weeks from mid-April to mid-June. The subplot-factor was four different hybrids (H1, H2, H3, H4) of varying relative maturities (100, 107, 111, and 115 days). Each replicate included a border plot on both ends of the block to reduce any edge-of-field effects on the measured plots. Furthermore, yield measurements were based on the center two rows (out of four). The research plots were managed according to agronomic best management practices (Thomison et al. 2017) outside of the main plot and subplot factors.

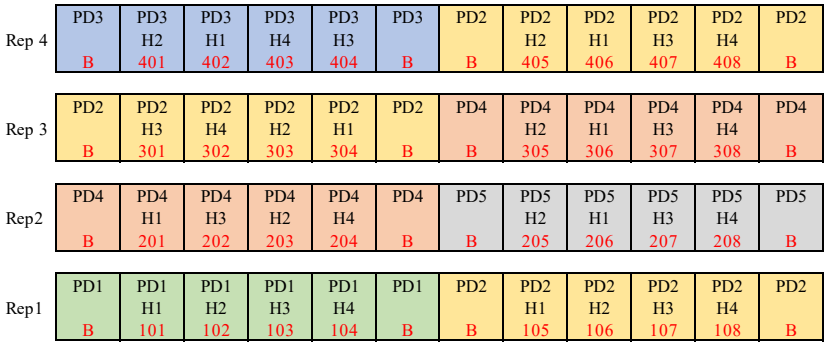

Figure 2: Plot Layout insert from the Western Research site (PD = Planting Date, H = Hybrid, B = Border Plot).

Each plot was 10 feet wide, configured as four rows at 30-inch spacing spanning approximately 30 feet long or longer at each location. The plots were systematically designated using a 3-digit numbering system: 101-120, 201-220, 301-320, and 401-420. A visual representation of the plot layout for Western Corn is shown in **Figure 2**.

3.2 Data Types

3.2.1 In-Situ Soil and Meteorological Sensing Data

An array of soil sensors was deployed at two depths, specifically at 30 cm and 60 cm, within the plots for Planting Date 2 and 4 at all three research locations. Additionally, one Apogee SQ-521 photosynthetic active radiation (PAR) sensor and one Meter ATMOS 14 weather station were installed at each of these research sites. The weather station collected temperature, relative humidity, vapor pressure, and barometric pressure in the crop canopy.

The data collected by these sensors was aggregated by a total of six ZL6 data loggers, with two loggers allocated at each research site. These loggers were connected to the Meter Group's Zentra Cloud, a data management and visualization platform. The ZL6 data loggers were configured to record sensor data at 30-minute intervals and upload data to the Zentra cloud hourly.

Data visualization was available through user-configurable dashboards on the website and data was also accessible via an application programming interface (API). A Python script was employed to interface with the Zentra Cloud application programming interface (API) to retrieve the data and aggregate it into a local database.

3.2.2 Weather Station Data

At each of the research locations, an Ohio State University (OSU) weather station collects precipitation, wind speed, and air temperature at multiple heights, which is accessible at weather.cfaes.osu.edu. In addition, the website also provides calculated daily values such as Growing Degree Days (GDD). The accumulation of GDD over the growing season is widely used in predicting corn growth and development.

3.2.3 Manually Labeled Data

Site visits were conducted weekly at each location by personnel from the Department of Horticulture and Crop Science (HCS). These individuals possessed expertise in the classification of crop growth stages. Furthermore, the final yield of each plot was collected at harvest.

3.3 Analysis of Yield Data

Table I provides a summary of yield information across the three research locations organized by planting date and hybrid. Note that Hybrid 1 for Planting Dates 1,2, and 3 for Northwest Research Station were removed from the dataset due to raccoon infestation that affected yield measurements.

Table 1. Summary of yield in bu/acre by location, hybrid, and planting date

	Western Research Station					Northwest Research Station					Wooster - Snyder Farm				
Planting Date	h1 yield	h2 yield	h3 yield	h4 yield	Avg yield by Planting Date	h1 yield	h2 yield	h3 yield	h4 yield	Avg yield by Planting Date	h1 yield	h2 yield	h3 yield	h4 yield	Avg yield by Planting Date
pd1	149	192	207	194	185	NA	235	245	249	243	188	212	214	203	204
pd2	162	189	179	142	168	NA	240	236	247	241	228	229	214	224	224
pd3	135	144	124	156	140	NA	246	247	254	249	190	201	198	180	192
pd4	218	268	216	237	235	198	203	195	202	199	199	203	196	193	198
pd5	225	224	223	237	227	186	176	166	173	175	174	189	200	187	188

4. Data Preparation

Figure 3 summarizes the data processing steps involved:

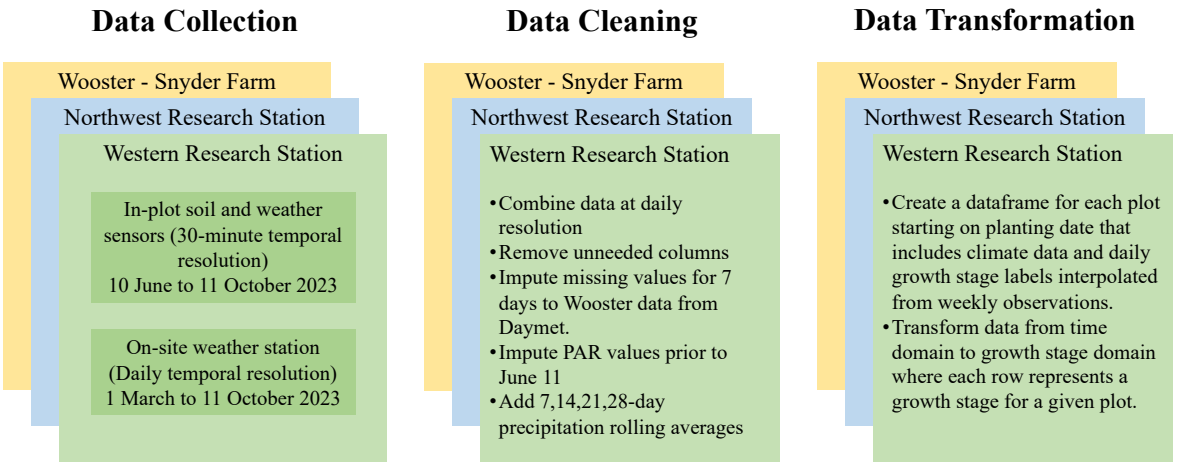

Figure 3: Summary of data processing steps

The data cleaning step involved combining the 30-minute resolution in-plot sensors with the daily resolution on-site weather station resolving two missing data issues:

- The photosynthetic active radiation (PAR) sensors were not available for installation until mid-June meaning that data from the early part of the growing season was not available for these sensors. However, the on-site weather station included a measurement for solar radiation. Using the data from mid-June to the end of the growing season where PAR and solar radiation were measured concurrently, a linear relationship was established which was then used to impute PAR values prior to mid-June.
- The Wooster on-site weather station had a gap in recorded data where values from 24-30 August 2023 were not recorded. These were imputed from the Daymet dataset (Thornton, et al. 2023).

Furthermore, additional calculated columns were created with previous 7, 14, 21, and 28-day rolling averages of precipitation to be used as a proxy for plant available water.

The growth stage labels were originally recorded as alphanumeric values corresponding to vegetative and reproductive growth stages commonly used by agronomists (Abendroth et al. 2011). These were converted to integers ranging from -1 to 16 via a Python dictionary and growth stages ‘V9’ to ‘V18’ were combined in order to achieve growth stages of similar duration as shown in **Figure 4**. The integer (-1) represented the period from planting date to ‘VE’.

```
# Dictionary to map alphanumeric growth stages to integers
gs_corn_dict = {
    'VE':  0, 'V1': 1, 'V2': 2, 'V3': 3, 'V4': 4, 'V5': 5, 'V6': 6,
    'V7': 7, 'V8': 8, 'V9': 9, 'V10': 9, 'V11': 9, 'V12': 9,
    'V13': 9, 'V14': 9, 'V15': 9, 'V16': 9, 'V17': 9, 'V18': 9,
'VT': 10,
    'R1': 11, 'R2': 12, 'R3': 13, 'R4': 14, 'R5': 15, 'R6': 16
}
```

Figure 4

With knowledge of the planting date for each plot, the data transformation step involved creating a dataframe for each plot at each location where the meteorological data began on the planting date. An additional column for growth stage was added and populated as an integer on each date where a growth stage observation was made, typically weekly as illustrated in **Table II**. Daily growth stage values were achieved by linearly interpolating between the weekly growth stage observations and converting resulting values to integers.

Table 2. Excerpt of growth stage labels

Planting Date Number	PD1	PD1	PD1	PD1	PD2	PD2	PD2
Hybrid	H1	H2	H3	H4	H1	H2	H3
date	101	102	103	104	105	106	107
5/5/23	VE	VE	VE	VE			
5/10/23	V1	V1	V1	V1	VE	VE	VE
5/17/23	V2	V2	V2	V2	V1	V1	V1
5/24/23	V3	V4	V4	V4	V3	V3	V3
5/31/23	V5	V5	V5	V5	V4	V5	V5
6/7/23	V6	V7	V7	V7	V5	V6	V6
6/14/23	V7	V7	V7	V7	V6	V7	V7
6/21/23	V7	V8	V8	V8	V7	V7	V8
6/26/23	V8	V9	V9	V10	V8	V8	V8
7/6/23	V13	V14	V13	V13	V10	V11	V11
7/13/23	R1	R1	R1	VT	R1	V16	V15
7/19/23	R2	R2	R2	R2	R1	R1	R1
7/24/23	R2	R2	R2	R1	R2	R2	R2
8/2/23	R4	R4	R3	R3	R3	R3	R3
8/8/23	R4	R4	R4	R4	R4	R4	R4
8/15/23	R5	R5	R4	R4	R4	R4	R4
8/23/23	R5	R5	R5	R5	R5	R5	R5
9/1/23	R5	R6	R6	R6	R5	R6	R6
9/6/23	R6	R6	R6	R6	R6	R6	R6
9/14/23	R6	R6	R6	R6	R6	R6	R6
9/21/23	R6	R6	R6	R6	R6	R6	R6
9/27/23	R6	R6	R6	R6	R6	R6	R6

With a dataframe for each plot consisting of input values of GDD, PAR, several rolling averages of precipitation, and daily growth stage values, the next step was to transform each dataframe from the time domain to the growth stage domain. This would ensure that each plot had identical sequence lengths of 18 rows (-1 to 16) and would be conducive to machine learning approaches.

This transformation was accomplished through functions available in the pandas library in Python. For all the rows in a dataframe corresponding to a particular growth stage, the sum of GDD and PAR was calculated along with the mean of the 7, 14, 21, and 28-day rolling averages of precipitation. **Figure 5** shows an example of one of the dataframes that has been transformed to the growth stage domain.

The resulting plot-level dataframes stored in Python dictionaries were then flattened into dataframes by location (Northwest, Western, Wooster) and into a combined dataframe that included all 3 locations. Each location had different organic matter levels which were introduced as an additional column (independent variable) into the dataset for Western (4.36%), Northwest (3.24%), and Wooster (2.23%). Planting date labels (pd1-pd5), hybrid labels (1-4), and hybrid values (100, 107, 111, 115) were introduced as columns (independent variables) in the flattened

dataframes both for training inputs and to separate the dataset into training and test sets. Finally, the yield in bu/acre was included as a column (dependent variable) in each dataframe. The dataset included a total of 228 sequences of data (Western - 80, Wooster - 80, Northwest – 68).

	growth_stage	GDD (F)	PAR_sum	7_day_avg_precip	14_day_avg_precip	21_day_avg_precip	28_day_avg_precip	num_days	cum_days
0	-1	98.300000	430270.16094	0.142792	0.192338	0.200325	0.199432	22	22
1	0	60.600000	126142.33447	0.060857	0.125286	0.140286	0.127286	5	27
2	1	98.700000	156875.03128	0.306939	0.181939	0.194490	0.183010	7	34
3	2	89.300000	217608.13158	0.136939	0.221939	0.166939	0.180102	7	41
4	3	48.300000	128833.11982	0.026786	0.116964	0.167143	0.131786	4	45
5	4	60.600000	91583.06438	0.000000	0.021667	0.134286	0.125357	3	48
6	5	155.300000	211897.24597	0.002041	0.008673	0.051429	0.115306	7	55
7	6	93.700000	166419.36543	0.158776	0.080408	0.058707	0.078265	7	62
8	7	217.522625	269261.34003	0.525831	0.494404	0.337936	0.255684	12	74
9	8	229.113750	218649.00000	0.375194	0.291404	0.421228	0.363064	10	84
10	9	89.080875	86897.00000	0.980621	0.584654	0.417014	0.543029	4	88
11	10	71.272500	70195.00000	0.251581	0.590248	0.416883	0.440890	3	91
12	11	146.055000	144038.00000	0.104624	0.421217	0.425954	0.338985	6	97
13	12	238.540875	260754.00000	0.164374	0.145506	0.292100	0.362131	10	107
14	13	88.843875	117041.00000	0.463550	0.298904	0.231019	0.317954	4	111
15	14	287.432625	305680.00000	0.357275	0.325455	0.296985	0.253651	13	124
16	15	457.468500	548572.00000	0.289131	0.355188	0.355160	0.356755	22	146
17	16	470.495625	632276.00000	0.029754	0.032294	0.060131	0.090014	35	181

Figure 5: Dataframe from Plot 101 at Western Research Station

5. Model Architecture and Configuration

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) commonly used for time-series prediction tasks. LSTM networks were created to solve the vanishing/exploding gradient issue that presents itself in basic RNNs. In this case, the data has been formatted where each sequence is a growth stage series of 18 rows and thus makes it conducive for training by an LSTM network.

The dataset includes columns for GDD, PAR, and rolling averages of precipitation which indicates the quantity of temperature, PAR, and water available for the creation of biomass and ultimately grain yield via photosynthesis. Each of the input features was scaled using StandardScaler in the scikit-learn library to facilitate improved training by removing the mean and scaling to unit variance. The dataset was subsequently converted into various combinations of training and testing tensors for model training and inference outlined in the results section.

Each model utilized learning rate = .001, hidden layer size = 25, number of layers = 2, loss function = mean absolute error (MAE), and dropout = 0.5. These values were determined through initial experimentation with the dataset.

6. Results

6.1 Batch Size Evaluation

The first evaluation was based on batch size of how data is presented to the LSTM network. **Figure 6** shows that batch sizes 6, 8, and 10 were clustered together and stabilized at a mean absolute error (MAE) of approximately 20, while batch sizes 4, 12, and 16 stabilized at an MAE of approximately 28. Based on these results, a batch size of 8 was used for the remainder of the paper. Each batch size was run for 2,500 epochs.

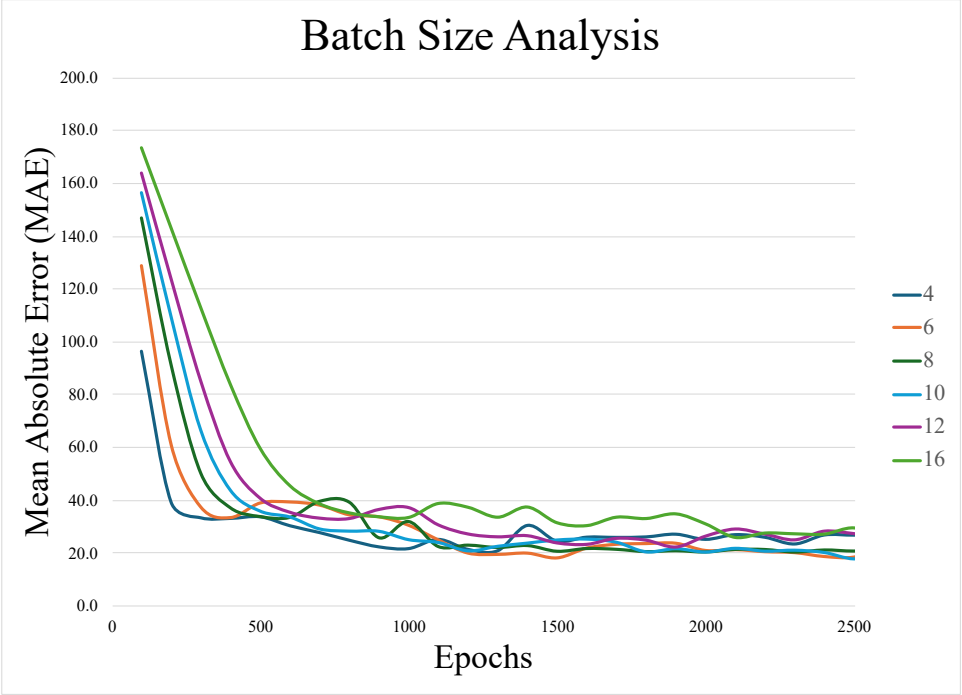

Figure 6. Comparison of Loss Function (MAE) vs. Training Epochs for Various Batch Sizes (4,6,810,12,16)

6.2 Input Data Evaluation

While each growth stage series included values from -1 to16, it is recognized that in the later growth stages, the weather has a diminishing role in yield. Additionally, four rolling averages were calculated to assess which representations of precipitation most positively influenced the model. To understand which combinations of input data and input sequences yielded the best results, the data was partitioned into a five-fold cross validation set by planting date, using data from four planting dates as training data and the remaining one as a validation set. This approach was chosen to ensure that all replications stayed in the same partition to avoid overfitting. Additionally, the availability of five planting dates was conducive for five-fold cross validation.

```
gs1 =[-1,…,13]
gs2 =[-1,…,12]
gs3 =[-1,…,11]
gs4 =[-1,…,10]
gs5 =[0,…,13]
gs6 =[0,…,12]
gs7 =[0,…,11]
gs8 =[0,…,10]
gs9 =[1,…,13]
gs10=[1,…,12]
gs11=[1,…,11]
gs12=[1,…,10]
```

```
X_1 =
['GDD','PAR','7d_precip']
X_2 =
['GDD','PAR','7d_precip','14d_precip']
X_3 =
['GDD','PAR','7d_precip','14d_precip', '21d_precip']
X_4 =
['GDD','PAR','7d_precip','14d_precip', '21d_precip',
'28d_precip]
```

Figure 7: Combinations of growth stage sequence and input columns

Twelve different variations of growth stage sequences were tested with four different variations of input columns for a total of 48 different combinations. The variations that were evaluated are shown in **Figure 7**.

Table 3. Comparison of Average and Standard Deviation of MAE for 5 cross-validation runs

Average MAE for 5 Cross-Validation Runs

	X_1	X_2	X_3	X_4
gs1	30.5	28.2	27.1	29.5
gs2	28.9	26.8	24.7	23.9
gs3	25.7	29.3	25.6	26.1
gs4	26.0	22.6	24.7	28.4
gs5	27.5	30.3	25.5	28.4
gs6	29.6	27.0	29.2	24.6
gs7	26.2	29.4	27.5	26.6
gs8	27.5	27.8	26.8	30.6
gs9	25.5	25.8	27.7	31.8
gs10	25.6	27.4	27.7	24.9
gs11	24.0	23.2	27.3	28.9
gs12	25.9	29.5	31.8	30.7

Highlighted in Table IV in green is the best combination of growth stage sequence and training input columns that yield the lowest MAE (gs4, X_2). The subsequent sections use this input configuration. Figure 10 shows the predicted vs. actual yield for the input configuration gs4, X_2 that was experimentally determined to have the lowest MAE.

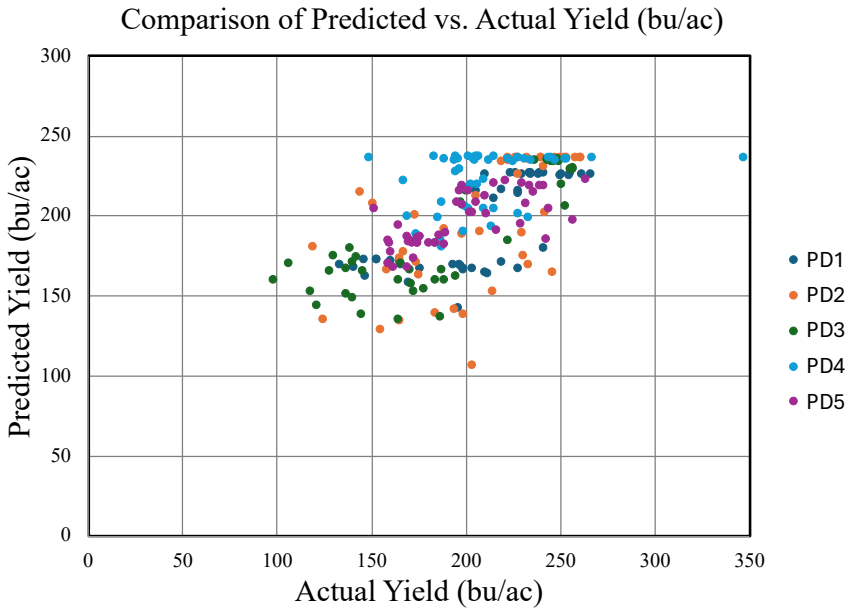

Figure 8. Results shown from five-fold cross validation. Each planting date shown (PD1-5) is from the test set and was trained from the other 4 planting dates using input configuration: gs4, X_2.

6.3 Using location specific data

In this section, the effect of keeping location-specific data segregated and training the data for each location separately using gs4 and X_2 as the input data configuration was evaluated. With a smaller dataset, training was conducted for 6,000 epochs to enable training losses to stabilize. This input data configuration was selected based on the results of Section 6.2. Table VI shows that a marginally improved MAE is achieved during testing at Northwest and Wooster (19.2 and 21.2 respectively) while Western (36.4) does not appear to benefit from location specific training

and testing. The yield measurements at Western had considerably higher standard deviation which may contribute to the reduced performance in the LSTM model.

Table 4. Analysis of model accuracy (MAE) using location specific data

	Western Research Station		Northwest Research Station		Wooster Snyder Farm	
Cross Validation	**train_loss**	**test_loss**	**train_loss**	**test_loss**	**train_loss**	**test_loss**
Train (PD2,3,4,5), Test (PD1)	10.5	45.5	7.8	12.8	8.3	21.9
Train (PD1,3,4,5), Test (PD2)	9.7	32.1	8.2	12.2	9.6	31.8
Train (PD1,2,4,5), Test (PD3)	14.0	27.2	9.7	14.0	7.2	20.9
Train (PD1,2,3,5), Test (PD4)	8.4	59.0	9.4	41.5	5.9	14.7
Train (PD1,2,3,4), Test (PD5)	10.6	18.1	10.2	15.3	9.1	16.7
Location Average		**36.4**		**19.2**		**21.2**

7. Future Work

The results from this paper indicate that high quality field scale data could provide the basis for accurate field-scale yield estimation. Furthermore, it highlights the possibility of creating agricultural foundation models drawn from the largest dataset possible which can then be fine-tuned on field scale data to provide the most accurate results for farmers. Of course, farmers who don't have the interest or capability in collecting high-fidelity on-farm data could still benefit from the results of generalized, less accurate models.

While this paper demonstrates promising results for field scale yield estimation, future work centered around five areas are important to make this a viable approach in production agriculture:

1. **Growing the dataset**. Machine learning models generally improve as their underlying dataset grows. Given the accuracy reported in this paper is from a dataset much smaller than other machine learning approaches referenced in this paper, it shows promise that a larger dataset that covers many more years and soil types could further increase the accuracy using this approach.
2. **Evaluating alternative models**. While LSTM models are commonly used for time-series data, XGBoost and various types of Transformer architectures have also proven to perform well on time-series data. It would be beneficial to evaluate alternative model architectures on this dataset.
3. **High spatial resolution soil testing**. Including more soil characteristics in addition to organic matter such as CEC, pH, and slope at higher spatial resolution holds the potential to improve the model prediction accuracy.
4. **Growth Stage inference from Unmanned Aerial Systems (UAS) imagery**. While this paper relied on manual observations of growth stage weekly in the field, future versions of this work are envisioned to leverage imagery from Unmanned Aerial Systems (UAS) to infer growth stage using machine learning techniques such as Vision Transformers (ViT).
5. **Splicing weather data scenarios in-season to provide daily yield predictions.** While this paper retrospectively looked at yield estimation, the goal is to provide yield predictions and ranges during the growing season that splices elapsed weather data with future meteorological scenarios for the current growing season to provide in-season yield estimation.

8. Conclusion

This paper demonstrated that using a growth stage centric approach to field scale yield estimation can achieve MAE of 22.6 bushels / acre with a relatively small dataset (228 sequences) across 3 locations. On this dataset, it also demonstrated that the MAE could be improved to 19.2 bu/acre

by training for a single location. This corresponds to an error of approximately 10% in yield estimation. Based on these results, it indicates that transforming meteorological data into a growth stage centric dataset could be a promising approach to provide in-season field scale yield prediction.

Acknowledgments

The authors would like to thank the farm managers, Lynn Ault, Joe Davlin, and Matt Davis, and their teams at each research site for executing these small-plot research trials and their support in collecting the data that went into this study. This study was funded by the Nationwide AgTech Innovation Hub, a collaboration between Nationwide Mutual Insurance Company, Ohio Farm Bureau and The Ohio State University College of Food, Agricultural, and Environmental Sciences.

References

Abendroth, L.J., Elmore, R.W., Boyer, M.J., and Marlay, S.K. (2011). Corn growth and development. PMR1009. Iowa State University Extension.

Casteel, S., Conley, S., Holshouser, D., Lee, C., Licht, M., Lindsey, L., et al. (2022). The Soybean Growth Cycle: Important Risks, Management and Misconceptions. Soybean Research & Information Network. https://soybeanresearchinfo.com/wp-content/uploads/2022/01/Science-for-Success-Soybean-Growth-Stages-V3.pdf. Accessed 5 June 2024

Crusiol, L. G. T., Sun, L., Sibaldelli, R. N. R., Junior, V. F., Furlaneti, W. X., Chen, R., et al. (2022). Strategies for monitoring within-field soybean yield using Sentinel-2 Vis-NIR-SWIR spectral bands and machine learning regression methods. Precision Agriculture, 23(3), 1093–1123. https://doi.org/10.1007/s11119-022-09876-5

Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., et al. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. Agricultural and Forest Meteorology, 276–277, 107609. https://doi.org/10.1016/j.agrformet.2019.06.008

Kuwata, K., & Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 858–861). Presented at the IGARSS 2015 - 2015 IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy: IEEE. https://doi.org/10.1109/IGARSS.2015.7325900

Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., et al. (2020). A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. Global Change Biology, 26(3), 1754–1766. https://doi.org/10.1111/gcb.14885

Licht, M. (2017). Estimating Corn Yields Using Yield Components | Integrated Crop Management. https://crops.extension.iastate.edu/cropnews/2017/08/estimating-corn-yields-using-yield-components. Accessed 5 June 2024

Sargun, K., & Mohan, S. (2020). Modeling the crop growth - A review. MAUSAM, 71(1), 103–114. https://doi.org/10.54302/mausam.v71i1.10

Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2021). Crop yield prediction integrating genotype and weather variables using deep learning. PLOS ONE, 16(6), e0252402. https://doi.org/10.1371/journal.pone.0252402

Thornton, M.M., Shrestha, R., Wei, Y., Thornton, P.E., Kao, S-C, and Wilson, B.E. (2023). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4 R1 (Version 4.1). ORNL Distributed Active Archive Center. https://doi.org/10.3334/ornldaac/2129

Thomison, P., Michel, A., Tilmon, K., Culman, S., Paul, P. (2017). Corn Production. In L. E. Lindsey & P. R. Thomison (Eds.), Ohio Agronomy Guide (15th ed., Bulletin no. 472, pp. 32–55). The Ohio State University Extension.

University of Missouri. (2012) Arrested Development in the Soybean Field. Integrated Pest Management University of Missouri. Arrested Development in the Soybean Field Part2 // Integrated Crop and Pest Management News Article // Integrated Pest Management, University of Missouri

Worrall, G., Judge, J., Boote, K., & Rangarajan, A. (2023). In-season crop phenology using remote sensing and model-guided machine learning. Agronomy Journal, 115(3), 1214–1236. https://doi.org/10.1002/agj2.21230

Zhou, B., Yue, Y., Sun, X., Ding, Z., Ma, W., & Zhao, M. (2017). Maize kernel weight responses to sowing date-associated variation in weather conditions. The Crop Journal, 5(1), 43–51. https://doi.org/10.1016/j.cj.2016.07.002