COMPARISON OF MACHINE LEARNING MODELS FOR WITHIN-FIELD COTTON YIELD PREDICTION: ASSESSING THE VALUE OF TEMPORAL DATA AND GENERALIZABILITY

Gonzalo J. Scarpin[a], [b]; Luke Fuhrer[c]; Wesley Porter[d]; Amandeep Kaur Dhaliwal[a]; Anish Bhattarai[a]; Amrinder Jakhar[a]; Leonardo M. Bastos[a]*

[a]Department of Crop and Soil Sciences, University of Georgia, Athens, Georgia, USA.

[b]Instituto Nacional de Tecnología Agropecuaria (INTA), Estación Experimental de Reconquista, Santa Fe, Argentina.

[c]Digital Ag Innovation Lab, Iowa State University, Ames, IA, USA

[d]Department of Crop and Soil Sciences, University of Georgia, Tifton, Georgia, USA.

***Presenting author: lmbastos@uga.edu**

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.

Cotton (*Gossypium hirsitum* L.) yield prediction is challenging due to the complex interactions between static environmental factors, such as soil properties and topography, and dynamic variables, including weather and crop management. These factors generate significant spatial and temporal variability within and across fields. While remote sensing technologies, particularly satellite-derived vegetation indices, provide spatially explicit data to evaluate crop health, translating this complex, multi-layered data into accurate, operational yield forecasts remains a challenge. Machine learning has emerged as a powerful tool to address this, yet critical questions remain regarding the specific value of temporal data, the computational feasibility of complex algorithms, and the true generalizability of models to unseen environments.

This study aimed to develop and evaluate a suite of machine learning models for predicting seed cotton yield, systematically assessing the incremental value of incorporating early, middle, and late-season remote sensing data. Static soil and topographic variables were integrated with dynamic Sentinel-2 vegetation indices across 32 cotton fields in Georgia, United States, over six growing seasons (2019-2024). Seed cotton yield spatial maps were generated via kriging and spatially matched with the predictors. Subsequently, the trade-off between predictive accuracy - measured by the coefficient of determination and root mean squared error—and computational demand was explicitly quantified. Furthermore, model generalizability was assessed utilizing a rigorous leave-one-field-out cross-validation framework.

The findings demonstrated that relying solely on pre-season static variables resulted in uniformly poor predictive performance across all algorithms. The most significant improvement in accuracy occurred with the integration of early-season vegetation indices, while mid- and late-season data provided progressively smaller gains in predictive ability. Among the tested algorithms, the stacked Ensemble model and the Extreme Gradient Boosting (XGBoost) model exhibited outstanding capabilities. The Ensemble model achieved the highest overall predictive accuracy (coefficient of determination of 0.949 and root mean squared error of 237 kilograms per hectare) when utilizing the full temporal dataset. However, this accuracy came at an extreme computational cost, requiring over 122 hours of processing time, which limits its operational deployment. Conversely, the XGBoost model proved to be the most practical solution, delivering highly competitive accuracy (coefficient of determination of 0.935 and root mean squared error of 268 kilograms per hectare) while completing its training in less than two hours.

Finally, the leave-one-field-out cross-validation revealed a substantial reduction in predictive performance compared to standard cross-validation, highlighting that traditional validation methods often overestimate a model's real-world applicability. Finally, this study underscores the necessity of in-season dynamic data for reliable yield forecasting and identifies XGBoost as offering the optimal balance between high predictive accuracy and computational efficiency for operational precision agriculture systems.

Keywords.
cotton yield prediction; machine learning; XGBoost; ensemble model; remote sensing; precision agriculture.

INTRODUCTION

Cotton (*Gossypium hirsutum* L.), cultivated primarily for its soft and versatile fiber, is the single most important natural textile material worldwide, serving as a critical economic engine for both developing and industrialized nations. Beyond its primary role in the textile industry, cotton is also a notable source of oilseed for human consumption and meal for animal feed, making it a multifaceted agricultural commodity with a profound impact on global commerce and food systems (Oosterhuis, 1990).

Cotton yield is the integrated outcome of complex interactions between genetics, environmental characteristics, and management practices (Snider et al., 2013). These factors generate significant yield variation—not only from one year to the next but also within a single field. Understanding and quantifying the sources of this variability is the central goal of precision agriculture and the prerequisite for developing effective yield prediction models (Filippi et al., 2024).

The primary drivers of this variability can be categorized as static or dynamic. Static environmental factors, such as soil properties and topography, create a persistent spatial pattern that influences crop performance. Soil texture governs critical functions like water holding capacity, nutrient retention, and soil aeration (Celik et al., 2023), while topography affects water movement, soil erosion, and microclimatic conditions across a field (Li et al., 2001). Dynamic environmental factors, primarily weather and in-season vegetation indices (VIs) derived from satellite remote sensing, introduce temporal variability and are widely used to monitor vegetation responses to environmental stresses (İrik et al., 2024). Management variables, including nitrogen fertilization, irrigation scheduling, and cultivar selection, add another layer of human-controlled variability directly impacting final cotton yield (Dhaliwal et al., 2022).

The vast and composite datasets generated by remote sensing are often too complex for traditional statistical methods. Machine Learning (ML), a branch of artificial intelligence, has emerged as a powerful paradigm for integrating these multi-source data streams to predict agricultural outcomes (van Klompenburg et al., 2020). ML algorithms excel at identifying complex, non-linear patterns characteristic of biological systems, making them ideally suited for yield prediction (Paudel et al., 2021). Recent literature demonstrates the superiority of ML models, particularly tree-based ensembles, over conventional linear regression for cotton yield prediction (Leo et al., 2021; Dhaliwal et al., 2022; Celik et al., 2023). Kashyap et al. (2024) found that stacked generalized ensemble models predicted cotton yield with higher efficiency and lower error compared to other models, while Quddus et al. (2025) reported a Deep Neural Network model achieving an R^2 of 97.9% and an RMSE of 28.3 kg ha^{-1}.

Despite high reported accuracy, the operational adoption of ML models remains limited. Several gaps persist: the specific contribution of data from different phenological stages has not been systematically quantified; the computational trade-off between performance and efficiency has received little attention; and most studies use standard random cross-validation, which overestimates generalizability. Leave-One-Field-Out (LOFO) cross-validation better emulates deploying a pre-trained model in a new environment (Croci et al., 2025).

Therefore, the objectives of this study are: (i) to develop and evaluate Ensemble and XGBoost models for predicting seed cotton yield (SCY), systematically assessing the incremental value of adding temporal remote sensing data from early, middle, and late phenological stages; (ii) to quantify the trade-off between model performance and computational demand; and (iii) to determine model generalizability under both standard cross-validation and LOFO validation strategies.

MATERIALS AND METHODS

Field Studies

This study was conducted across 32 cotton fields in the state of Georgia, USA, during six growing seasons (2019–2024). Two selected farmers contributed to multiple fields each year, with variation in location, area, and production level (Table 1). Fields were located in two distinct regions: Screven County (Coastal Plain, southeastern Georgia; ~1220 mm annual rainfall, ~18°C mean temperature; predominantly loamy and sandy soils) and Miller County (southwestern Georgia; ~1320 mm annual rainfall, ~20°C mean temperature; a mix of sand, loam, and clay soils). A total of 21 and 11 fields were located in Screven and Miller counties, respectively.

Table 1. Summary of field-level yield statistics across multiple sites and years in Miller and Screven counties, Georgia. For each field, the cultivated area (ha), minimum, maximum, and average yield are reported, along with the coefficient of variation (CV) of yield.

Field N	County	Field Name	Year	Area (ha)	Minimum ($kg\ ha^{-1}$)	Maximum ($kg\ ha^{-1}$)	Average ($kg\ ha^{-1}$)	CV (%)
1	Miller	Shy	2019	67.6	160.8	6644.6	3615.7	37.5
2	Miller	Big Half	2019	39.5	63.3	6482.2	3625.4	38.9
3	Miller	Big Half	2020	39.5	584.3	4008.6	2509.3	21.5
4	Miller	Fire Tower	2020	18.8	804.7	5079.7	2665.2	31.5
5	Miller	Hog House	2020	27.9	1156.8	3756.6	2320.4	16.5
6	Miller	Weaver S.	2021	19.2	594.1	3685.7	2621.5	21.4
7	Miller	Weaver N.	2021	20.0	247.9	2749.3	1850.9	30.6
8	Miller	Shy	2021	30.2	924.8	4953.5	3522.6	25.7
9	Miller	45 Dry	2022	18.2	932.6	4660.5	3040.8	22.1
10	Miller	Hog House	2022	22.7	637.4	5082.5	3342.9	21.7
11	Miller	Matt Barn	2022	14.5	366.3	5134.1	3190.7	35.0
12	Miller	Shy	2022	67.6	982.8	5428.8	3308.5	27.3
13	Screven	Beach Field	2023	26.1	1672.4	3634.0	2832.7	13.1
14	Screven	New Place	2023	64.7	12.0	4214.2	2100.1	37.1
15	Screven	50 Acres	2023	15.3	2948.2	5522.5	4156.9	12.7
16	Screven	Gross Place	2023	11.5	142.5	3095.0	1732.1	32.7
17	Screven	Long Leaf	2023	3.5	457.9	3690.9	2359.5	35.0
18	Screven	Pear Tree	2023	5.7	2672.9	3537.1	3155.2	4.4
19	Screven	Pivot 3	2023	11.1	2220.1	5325.3	3968.8	14.1
20	Screven	50 Acres	2024	18.5	1635.8	4587.2	3490.8	18.5
21	Screven	Bog Down	2024	7.4	1926.3	4482.4	3308.9	15.3
22	Screven	H.Hole	2024	2.7	2405.3	4221.8	3423.3	11.1
23	Screven	Hey 17	2024	4.8	1670.9	4803.7	3492.8	17.2

24	Screven	J.M.	2024	4.9	1960.7	3664.0	2977.5	10.5
25	Screven	Long Leaf	2024	3.5	796.4	5193.7	3298.9	30.4
26	Screven	Mule Barn	2024	15.4	1670.3	5096.1	3456.8	20.6
27	Screven	New Place	2024	64.7	243.6	4256.3	2134.5	30.9
28	Screven	Pivot 1	2024	36.4	1762.0	4811.7	3615.6	15.8
29	Screven	Pivot 3	2024	11.1	1778.2	4134.5	3132.2	16.3
30	Screven	Shop	2024	3.3	123.9	3648.3	2173.8	33.0
31	Screven	Terrace	2024	22.3	1390.8	4782.1	3185.0	18.0
32	Screven	Willies	2024	6.5	1835.0	4146.6	3294.0	13.8

Static and Dynamic Data Collection

For each field, a set of static and dynamic predictor variables were generated:

Static predictors: Soil properties, including total clay, silt and sand content (%), were extracted from the 10-m resolution gridded SSURGO (gSSURGO) database using the soilDB package in R (Beaudette D. et al., 2026). Terrain variables, including mean elevation and slope, were derived from 1-m Digital Elevation Models (US Geological Survey, 2019) using the terra package (Hijmans et al., 2022).

Dynamic predictors: Sentinel-2 optical imagery was acquired for the entire growing season for each field via the openEO package in Rstudio (Lahn, 2026). All remote sensing images were obtained at a 10 m spatial resolution with a five-day revisit time. A cloud-masking filter was applied to all images to remove pixels affected by cloud cover, and images affected by cloud cover were removed from the analysis. After cloud-cover removal there were at least 35 images per field to calculate the different VIs. Four VIs were considered: NDVI, GNDVI, NDRE, and RNDVI. These four VIs were selected to provide complementary information on canopy health suitable for tracking cotton phenology: NDVI serves as a widely used standard for green biomass (Yengoh et al., 2015), GNDVI offers greater sensitivity to chlorophyll concentration in dense canopies where NDVI may saturate (Wijayanto et al., 2024), NDRE is particularly effective for detecting variations in chlorophyll content and nitrogen status in later crop growth (Thompson et al., 2020) and RDVI helps to minimize soil background effects in sparser canopy conditions (Sun et al., 2021). In addition, these VIs have been reported for several research in cotton (Thompson et al., 2020; Tilse et al., 2024; Lacerda et al., 2025) and others field crops.

Vegetation indices were aggregated into three phenological stages based on cumulative growing degree days (GDD) and typical cotton growth in the region (Raper et al., 2023), dividing the growing season into: early (0–666 GDD), middle (666-2184 GDD), and late season (2184 GDD to harvest). GDD was calculated daily using the formula GDD = ((Tmax+Tmin))/2-Tbase, with upper and base temperature thresholds of 34°C and 15.5°C, respectively, commonly used for cotton (Raper et al., 2023).

For every grid size and phenological stage, metrics representing the central tendency of each predictor were calculated. The median (Med) was used for all predictors to minimize the influence of outliers, whereas the area under the curve (AUC) was calculated as a measure of cumulative seasonal vigor.

Yield Data Processing and Spatial Integration

Raw yield monitor data were cleaned using a dynamic filtering method based on the 10th and 99th percentiles extended by 25% of the inter-percentile range, following (Sudduth et al., 2012). A 10-m negative buffer was applied to the field boundary to remove edge-effect points. Cleaned SCY data were spatially interpolated to a continuous surface using ordinary kriging via the gstat package in R (Gräler et al., 2016), fitted to the best empirical variogram among Spherical, Exponential, and Gaussian models using weighted least squares. The interpolation grid was aligned with Sentinel-2 pixels. All spatial data layers were integrated through spatial joins and tabular left_join operations using the sf package (Pebesma and Bivand, 2023), yielding a unified dataset of 73,426 observations.

Machine Learning Models

Two ML algorithms were selected to represent a performance-efficiency narrative. The Ensemble model combines predictions from multiple base learners (XGBoost, Random Forest, Linear Regression, k-Nearest Neighbors, Support Vector Machine, and Multilayer Perceptron) into a final prediction (Wolpert, 1992; Kuhn and Wickham, 2020), representing the highest achievable accuracy. Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) was selected as the most practical algorithm based on its high accuracy and low computational cost (Scarpin et al., 2026).

All models were implemented in the *tidymodels* framework in RStudio (Kuhn and Wickham, 2020). The dataset was partitioned 70%/30% for training and testing. A 10-fold cross-validation with a search of over 50 hyperparameter combinations using the finetune package was applied to minimize RMSE. Models were trained on four progressive predictor sets: (S) static only; (S+E) static plus early-season VIs; (S+E+M) adding mid-season VIs; and (S+E+M+L) the full dataset including late-season VIs. A LOFO cross-validation was conducted for XGBoost, training on 31 fields and testing on the held-out 32nd, repeated for all fields. Due to its extreme computational cost (>122 hours per run), the Ensemble model was excluded from LOFO validation. All computations ran on a high-performance computing cluster (UGA), with 16 CPU cores and 32 GB RAM per job, using R version 4.4.2.

Model Performance Evaluation

Model performance was assessed using the coefficient of determination (R^2) and Root Mean Square Error (RMSE, kg ha^{-1}) calculated on the independent test set. Computational cost was measured as total CPU time for training and tuning, serving as a benchmark for operational feasibility.

RESULTS AND DISCUSSION

Seed cotton yield showed substantial variation across all 32 fields, ranging from near zero to approximately 6,500 kg ha^{-1} (Figure 1). The overall dataset median was 2,976 kg ha^{-1}, with an interquartile range from 2,250 (Q1) to 3,730 kg ha-1 (Q3). Fields such as 28_2024_Pivot1 and 19_2023_Pivot3 exhibited high productivity, while 16_2023_GrossPlace and 7_2021_WeaverNorth showed lower yields with substantial areas below the global median. Some fields displayed narrow, symmetric distributions indicating uniform growing conditions, while others showed pronounced bimodal or skewed distributions. This pronounced spatial and temporal variability underscores the complexity of yield formation in cotton and the need for robust predictive models.

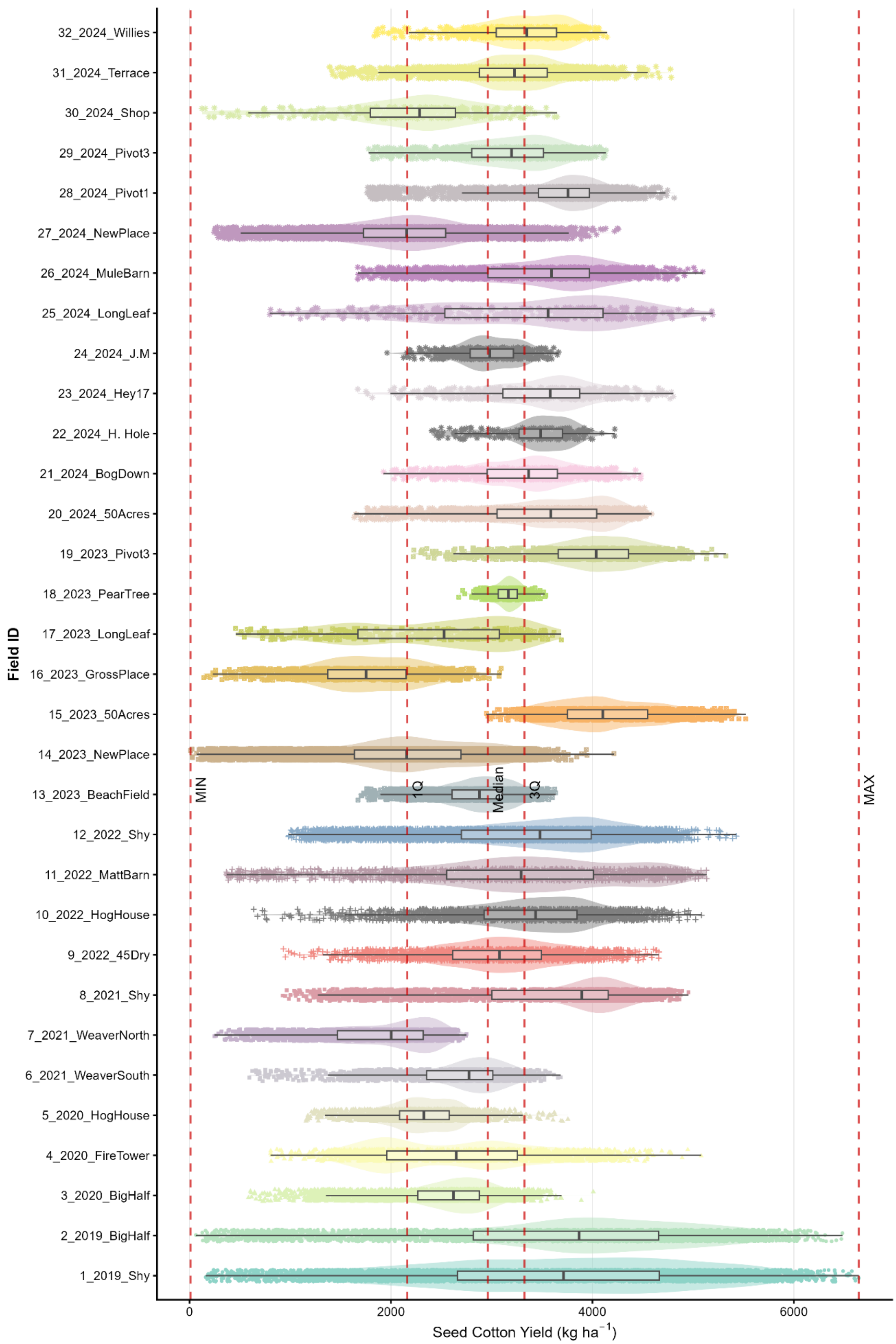

Figure 1. Distribution of seed cotton yield (SCY) across 32 fields in Miller and Screven counties, Georgia, US. Each violin plot represents the yield distribution within a field, with overlaid boxplots indicating the median and interquartile range (IQR). Vertical dashed red lines denote the global minimum, Q1, median, Q3, and maximum yield values across the aggregated dataset.

The comparative performance of Ensemble and XGBoost models across the four progressive predictor datasets under the standard 70/30 split and, for XGBoost, under LOFO cross-validation are presented in Table 2. The inclusion of temporal data was critical for predictive accuracy. When trained solely on static variables, both models performed poorly: XGBoost achieved $R^2 = 0.325$ (RMSE = 859 kg ha^{-1}), while Ensemble reached $R^2 = 0.343$ (RMSE = 847 kg ha^{-1}). This result confirms that pre-season variables alone are insufficient to robustly model the complex determinants of SCY, consistent with Franz et al. (2020), who reported that in-season vegetation condition was a far more important predictor than soil texture.

The most significant improvement occurred with the addition of early-season VIs. XGBoost's R^2 increased from 0.325 to 0.850 and RMSE was reduced from 859 to 406 kg ha^{-1}, while Ensemble improved from 0.343 to 0.883 (RMSE = 358 kg ha^{-1}). This result suggests that early-season remote sensing data captures the initial yield potential established during stand establishment and early vegetative growth, which then determines much of the final productivity. Subsequent inclusion of mid-season and late-season data yielded additional, though progressively smaller, improvements, suggesting diminishing returns as phenological stages advance.

With the full S+E+M+L dataset, the Ensemble model achieved the highest predictive accuracy ($R^2 = 0.949$, RMSE = 237 kg ha^{-1}), closely followed by XGBoost ($R^2 = 0.935$, RMSE = 268 kg ha^{-1}). These results are consistent with Kashyap et al. (2024) and Mitra et al. (2024), who reported R^2 values above 0.95 for ensemble and Random Forest models, respectively. The near-equivalent performance of XGBoost to the more complex Ensemble model—at a fraction of the computational cost—is one of the most practically significant findings of this study.

The computational trade-off was remarkable. XGBoost completed its most complex training run in approximately 1 hour and 42 minutes, while the Ensemble required over 122 hours (approximately 5.1 days) for the same dataset. This 60-fold difference in runtime, for only a marginal gain in accuracy (R^2 0.935 vs. 0.949), represents a critical operational consideration. Consultants, cooperatives, and extension agents often operate without access to high-performance computing clusters, making the Ensemble model impractical for routine deployment. These findings reinforce conclusions by Dhaliwal et al. (2022) and İrik et al. (2024), who highlighted tree-based ensemble methods as the most practical choice for agricultural prediction tasks.

Table 2. Performance of Ensemble and XGBoost models across four cumulative predictor datasets under standard 70/30 split validation and, for XGBoost, LOFO cross-validation. Performance metrics include R^2, RMSE (kg ha^{-1}), and CPU training time (HH:MM:SS). LOFO was not computationally feasible for the Ensemble model.

Model	Dataset	R^2 (Split)	RMSE (Split)	CPU Time (Split)	Mean R^2 (LOFO)	Mean RMSE (LOFO)
XGBoost	Static (S)	0.325	859 kg ha^{-1}	0:26:01	0.06	857 kg ha^{-1}
XGBoost	S + Early (E)	0.850	406 kg ha^{-1}	2:05:06	0.07	901 kg ha^{-1}
XGBoost	S + E + Middle (M)	0.911	312 kg ha^{-1}	1:30:07	0.41	697 kg ha^{-1}
XGBoost	S + E + M + Late (L)	0.935	268 kg ha^{-1}	1:42:10	0.47	619 kg ha^{-1}
Ensemble	Static (S)	0.343	847 kg ha^{-1}	48:48:24	—	—
Ensemble	S + Early (E)	0.883	358 kg ha^{-1}	93:47:42	—	—
Ensemble	S + E + Middle (M)	0.935	267 kg ha^{-1}	96:24:27	—	—
Ensemble	S + E + M + Late (L)	0.949	237 kg ha^{-1}	122:45:08	—	—

The LOFO validation results for XGBoost reveal a substantial reduction in predictive performance compared to the standard 70/30 split—a pattern consistent with recent work by Croci et al. (2025), who highlighted the challenge of temporal transferability in ML-based crop yield models. Using only static data, XGBoost achieved a LOFO median R^2 of only 0.06, indicating that soil texture and terrain alone capture virtually none of the variation across unseen fields. This suggests that static variables encode site-specific relationships that do not transfer well across environments. Inclusion of early-season VIs provided only marginal improvement (median R^2 = 0.07), indicating that early spectral signals do not yet differentiate between fields in ways that generalize. The most notable improvement occurred with the addition of mid-season data (S+E+M), where XGBoost's median LOFO R^2 increased to 0.41 (RMSE = 697 kg ha^{-1}). With the full dataset, median LOFO R^2 reached 0.47 (RMSE = 619 kg ha^{-1}), confirming that late-season signals provide additional, generalizable information on yield outcomes. Figure 2 illustrates the field-level distribution of these results, revealing substantial variability in model performance across fields and datasets—particularly for R^2, which ranged from near zero to approximately 0.79 depending on field characteristics and predictor set.

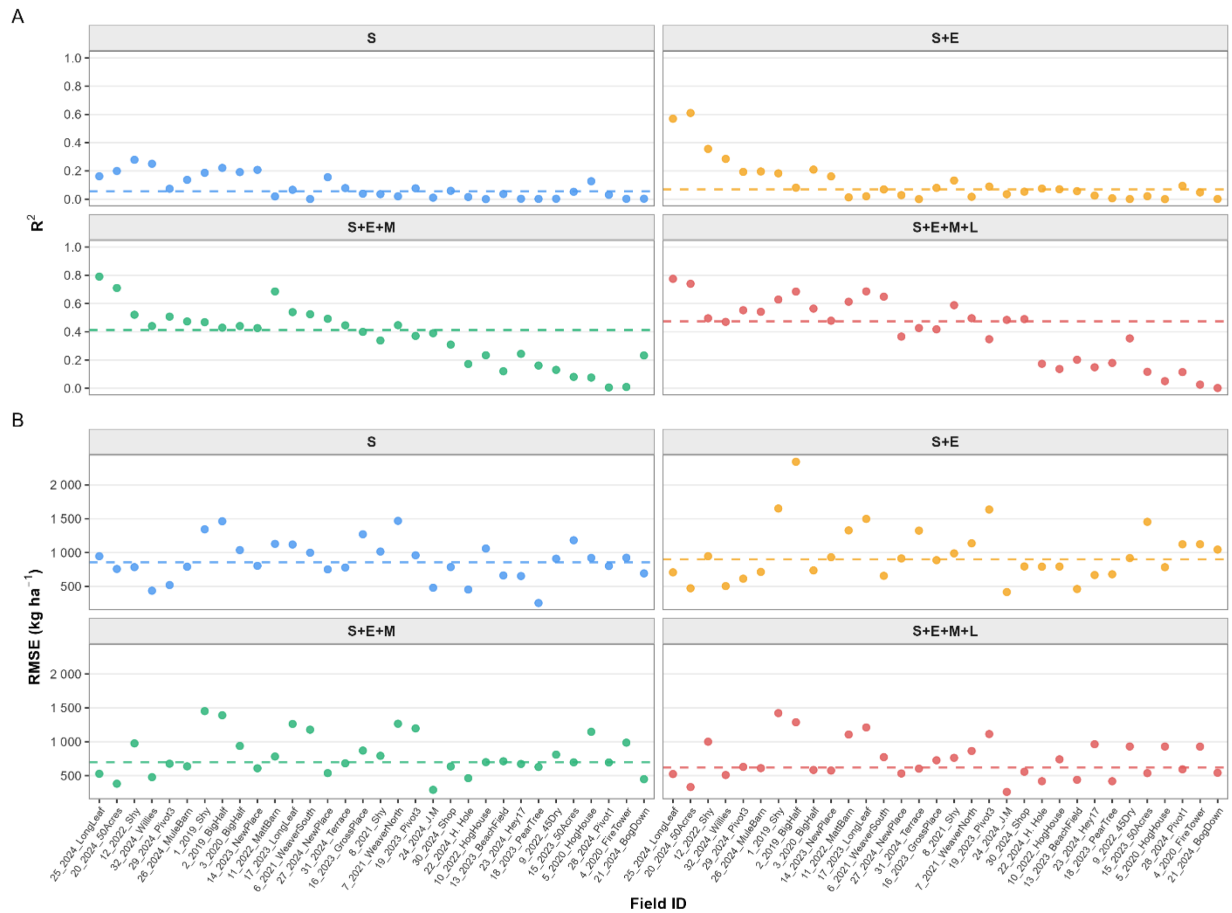

Figure 2. Field-level XGBoost LOFO cross-validation performance across four cumulative predictor datasets (S, S+E, S+E+M, S+E+M+L). Panel A shows the coefficient of determination (R^2) and Panel B shows the root mean square error (RMSE, kg ha^{-1}) for each of the 32 fields, ordered from highest to lowest median R^2 across datasets. Dashed lines indicate the median value for each dataset.

Despite the generalizability gap, XGBoost's LOFO performance reinforces its practical value. The 60-fold computational advantage over the Ensemble model—approximately 13 hours for the full LOFO analysis versus an estimated >500 hours for an equivalent Ensemble run—makes XGBoost a scalable and tractable algorithm for repeated retraining as new field data become available, and its median R^2 of 0.47 under LOFO remains competitive among published ML-based cotton yield prediction studies.

The dual-model framework evaluated here—Ensemble for maximum accuracy versus XGBoost for operational practicality—provides a clear decision framework for experts. When the goal is to develop a research-grade model or optimize predictions within a well-characterized set of fields, the Ensemble model's additional accuracy (RMSE 237 vs. 268 kg ha^{-1}) may justify the computational investment. However, for operational deployment by agronomists and cooperatives requiring frequent retraining, ingestion of new seasonal data, or rapid in-season predictions, XGBoost is unambiguously the preferred choice.

Several limitations should be acknowledged. Dynamic weather data (rainfall, temperature, solar radiation) were not included as predictors; incorporation of these variables would likely improve temporal transferability across seasons and regions. Furthermore, cotton genetics and cultivar information were not accounted for, despite being a critical determinant of yield potential (Scarpin et al., 2025). Critical soil variables beyond texture—such as nutrient levels—and in-season management records (pest pressure, growth regulators, among others) were unavailable, and their inclusion would likely explain additional yield variance. Future work should address these gaps and explore data fusion strategies combining Sentinel-2 imagery with high-resolution UAV or thermal data.

CONCLUSIONS

This study provides a comprehensive, operationally-grounded evaluation of Ensemble and XGBoost models for within-field cotton yield prediction using satellite-derived vegetation indices and static environmental data across 32 commercial fields in Georgia, USA (2019–2024). The main conclusions of this study are:

In-season temporal data is essential for accurate yield prediction. Static pre-season variables (soil texture, topography) alone yielded poor performance ($R^2 \leq 0.343$), while the addition of early-season vegetation indices produced the largest single improvement across both models, with further, diminishing gains from mid- and late-season data.

XGBoost represents the most practical model for operational precision agriculture. With the full dataset, XGBoost achieved $R^2 = 0.935$ and RMSE = 268 kg ha^{-1} in approximately 1.7 hours of CPU time—matching the Ensemble model's performance ($R^2 = 0.949$, RMSE = 237 kg ha^{-1}) that required over 122 hours, representing a ~60-fold improvement in efficiency for a minimal accuracy gain of 1.4% in R^2.

Model generalizability to unseen fields remains a key challenge. Under LOFO cross-validation, XGBoost's median R^2 dropped to 0.47—a substantial but gap. These results argue for more diverse training datasets and spatial validation strategies as standard practice in ML-based yield prediction research.

Together, these findings provide actionable benchmarks for practitioners and researchers seeking to implement scalable, ML-driven yield prediction systems in cotton production.

REFERENCES

Beaudette D., Skovlin J., Roecker S., A., B., 2026. soilDB: Soil Database Interface.

Celik, M.F., Isik, M.S., Taskin, G., Erten, E., Camps-Valls, G., 2023. Explainable Artificial Intelligence for Cotton Yield Prediction With Multisource Data. *IEEE Geoscience and Remote Sensing Letters* 20, 1–5.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, San Francisco, California, USA, pp. 785–794.

Croci, M., Ragazzi, M., Grassi, A., Impollonia, G., Amaducci, S., 2025. Assessing the temporal transferability of machine learning models for predicting processing pea yield and quality using Sentinel-2 and ERA5-land data. *Smart Agricultural Technology* 12, 101207.

Dhaliwal, J., Panday, D., Saha, D., Lee, J., Jagadamma, S., Schaeffer, S., Mengistu, A., 2022. Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. *Computers and Electronics in Agriculture* 199.

Filippi, P., Han, S.Y., Bishop, T.F.A., 2024. On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. *Precision Agriculture* 26.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-Temporal Interpolation using gstat. *The R Journal* 8, 204–218.

Hijmans, R.J., Bivand, R., Forner, K., Ooms, J., Pebesma, E., 2022. Package 'terra'.

İrik, H.A., Ropelewska, E., Çetin, N., 2024. Using spectral vegetation indices and machine learning models for predicting the yield of sugar beet (*Beta vulgaris* L.) under different irrigation treatments. *Computers and Electronics in Agriculture* 221.

Kashyap, G.R., Sridhara, S., Manoj, K.N., Gopakkali, P., Das, B., Jha, P.K., Prasad, P.V.V., 2024. Machine learning ensembles, neural network, hybrid and sparse regression approaches for weather based rainfed cotton yield forecast. *International Journal of Biometeorology* 68, 1179–1197.

Kuhn, M., Wickham, H., 2020. Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles.

Lacerda, L.N., Ardigueri, M., C. Barboza, T.O., Snider, J., Chalise, D.P., Gobbo, S., Vellidis, G., 2025. Using High-Resolution Multispectral Data to Evaluate In-Season Cotton Growth Parameters and End-of-the-Season Cotton Fiber Yield and Quality. *Agronomy* 15.

Lahn, F., 2026. openeo: Client Interface for 'openEO' Servers.

Leo, S., De Antoni Migliorati, M., Grace, P.R., 2021. Predicting within-field cotton yields using publicly available datasets and machine learning. *Agronomy Journal* 113, 1150–1163.

Li, H., Lascano, R.J., Barnes, E.M., Booker, J., Wilson, L.T., Bronson, K.F., Segarra, E., 2001. Multispectral Reflectance of Cotton Related to Plant Growth, Soil Water and Texture, and Site Elevation. *Agronomy Journal* 93, 1327–1337.

Mitra, A., Beegum, S., Fleisher, D., Reddy, V.R., Sun, W., Ray, C., Timlin, D., Malakar, A., 2024. Cotton Yield Prediction: A Machine Learning Approach With Field and Synthetic Data. IEEE Access 12, 101273–101288.

Oosterhuis, D.M., 1990. Growth and Development of a Cotton Plant. Nitrogen Nutrition of Cotton: Practical Issues, pp. 1–24.

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., Athanasiadis, I.N., 2021. Machine learning for large-scale crop yield forecasting. *Agricultural Systems* 187.

Pebesma, E., Bivand, R., 2023. Spatial Data Science: With Applications in R.

Quddus, M.A., Chowdhury, S., Jasper, W., Das, L., 2025. Enhancing cotton yield prediction with robust deep neural network-based framework. *European Journal of Agronomy* 170.

Raper, T.B., Ward, R., Rushing, C., Brown, S., Sandlin, T., Norton, R., Hutmacher, B., Snider, J.L., Fromme, D., Dodds, D., Jones, M., Edmisten, K., Collins, G., Byrd, S., Griffin, J., Maeda, M., Frame, H., Pieralisi, B., Meeks, C., Whitaker, J., Nichols, R.L., Morgan, G., 2023. Reevaluation of the degree day base 60°F concept in US cotton (*Gossypium hirsutum* L.) production. *Agronomy Journal* 115, 3045–3061.

Scarpin, G.J., Bhattarai, A., Hand, L.C., Snider, J.L., Roberts, P.M., Bastos, L.M., 2025. Cotton lint yield and quality variability in Georgia, USA: Understanding genotypic and environmental interactions. *Field Crops Research* 325.

Scarpin, G.J., Studstill, S.B., Monfort, W.S., Tubbs, R.S., Pilon, C., Jakhar, A., Bhattarai, A., Dhaliwal, A.K., Bastos, L.M., 2026. Peanut yield and grade prediction in Georgia, USA: integrating management, climate, and remote sensing data with explainable AI. *Computers and Electronics in Agriculture* 241.

Snider, J., Collins, G., Whitaker, J., Davis, J., 2013. Quantifying Genotypic and Environmental Contributions to Yield and Fiber Quality in Georgia: Data from Seven Commercial Cultivars and 33 Yield Environments. *Journal of Cotton Science* 17, 285–292.

Sudduth, K.A., Drummond, S.T., Brenton Myers, D., 2012. Yield Editor 2.0: Software for Automated Removal of Yield Map Errors. 2012 Dallas, Texas, July 29 - August 1, 2012. ASABE, St. Joseph, MI.

Sun, Q., Jiao, Q., Qian, X., Liu, L., Liu, X., Dai, H., 2021. Improving the retrieval of crop canopy chlorophyll content using vegetation index combinations. *Remote Sensing* 13, 470.

Thompson, C.N., Mills, C., Pabuayon, I.L.B., Ritchie, G.L., 2020. Time-based remote sensing yield estimates of cotton in water-limiting environments. *Agronomy Journal* 112, 975–984.

Tilse, M.J., Filippi, P., Whelan, B., Bishop, T.F.A., 2024. Downscaling crop production data to fine scale estimates with geostatistics and remote sensing: a case study in mapping cotton fibre quality. *Precision Agriculture* 25, 2921–2957.

van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* 177.

Wijayanto, Y., Safitri, M., Purnamasari, I., Budiman, S.A., Saputra, T.W., Regar, A.F., Ristiyana, S., 2024. Estimating the contents of Chlorophyll, Nitrogen, and Yields on Rice Through Sentinel-2 Vegetation Indices in Heterogeneous Land Management. *The Indonesian Journal of Geography* 56, 357–367.

Wolpert, D.H., 1992. Stacked generalization. *Neural networks* 5, 241–259.

Yengoh, G.T., Dent, D., Olsson, L., Tengberg, A.E., Tucker III, C.J., 2015. Use of the Normalized Difference Vegetation Index (NDVI) to assess Land degradation at multiple scales: current status, future trends, and practical considerations. *Springer*.