



OPEN

An integrated approach of field, weather, and satellite data for monitoring maize phenology

Luciana Nieto^{1✉}, Raí Schwalbert¹, P. V. Vara Prasad^{1,2}, Bradley J. S. C. Olson³ & Ignacio A. Ciampitti^{1✉}

Efficient, more accurate reporting of maize (*Zea mays* L.) phenology, crop condition, and progress is crucial for agronomists and policy makers. Integration of satellite imagery with machine learning models has shown great potential to improve crop classification and facilitate in-season phenological reports. However, crop phenology classification precision must be substantially improved to transform data into actionable management decisions for farmers and agronomists. An integrated approach utilizing ground truth field data for maize crop phenology (2013–2018 seasons), satellite imagery (Landsat 8), and weather data was explored with the following objectives: (i) model training and validation—identify the best combination of spectral bands, vegetation indices (VIs), weather parameters, geolocation, and ground truth data, resulting in a model with the highest accuracy across years at each season segment (step one) and (ii) model testing—post-selection model performance evaluation for each phenology class with unseen data (hold-out cross-validation) (step two). The best model performance for classifying maize phenology was documented when VIs (NDVI, EVI, GCVI, NDWI, GVM) and vapor pressure deficit (VPD) were used as input variables. This study supports the integration of field ground truth, satellite imagery, and weather data to classify maize crop phenology, thereby facilitating foundational decision making and agricultural interventions for the different members of the agricultural chain.

Maize (*Zea mays* L.) is one of the leading grain crops, with 193 M ha harvested globally and more than 33 M ha in the United States (US) during the 2018 growing season. US maize production is largely concentrated in the central region, accounting for roughly 85% of total US production¹. The US Department of Agriculture (USDA) via the National Agricultural Statistics Service (NASS) releases a weekly report for in-season crop progress, termed Crop Progress and Report Conditions (CPRC), which provides a subjective estimate of crop phenology and condition in major US producing states¹. Estimates from the CPRC are based on survey data gathered on a weekly basis from an extensive network of regional agricultural agents based on their field observations². Although the data is informative, the collection process is labor intensive, time consuming, financially ineffective, and subject to bias^{3,4}. With challenges eminent due to funding restrictions and travel limitations, new approaches should be pursued, tested, and quickly implemented.

An extensive record of scientific literature portrays the relevance of implementing well-timed management practices to improve yields and input use efficiency, such as timely irrigation, fertilization, crop protection, and harvest^{5–7}. Remotely sensed satellite data presents a significant opportunity to improve timely agricultural interventions and monitoring of crop vegetation. With various spectral, radiometric, temporal and spatial resolutions, satellites function as a critical source of data to aid tracking of field crop phenology progress^{8–10}. Combining satellite data features permit the generation of different vegetation indices (VIs), such as the Normalized Difference Vegetation Index (NDVI)¹¹, Enhanced Vegetation Index (EVI)¹², Green Chlorophyll Vegetation Index (GCVI)¹³, Global Vegetation Moisture Index (GVM)¹⁴ and Normalized Difference Water Index (NDWI)¹⁵, among many others. These VIs describe changes in vegetation dynamics and are correlated with plant traits such as leaf area index, leaf chlorophyll concentration, and canopy water content^{13–18}. Integrating remote sensing with ground truth and weather data has great potential to advance science and improve overall prediction of crop yield, in-season progress, and crop phenology^{2,19,20}.

¹Department of Agronomy, 2004 Throckmorton Plant Science Center, Kansas State University, 1712 Claflin Road, Manhattan, KS 66506, USA. ²Sustainable Intensification Innovation Lab, Kansas State University, 108 Waters Hall, 1603 Old Claflin Place, Manhattan, KS 66506, USA. ³Department of Biology, Kansas State University, Chalmers Hall, 1711 Claflin Road, Manhattan, KS 66506, USA. ✉email: lnieto@ksu.edu; ciampitti@ksu.edu

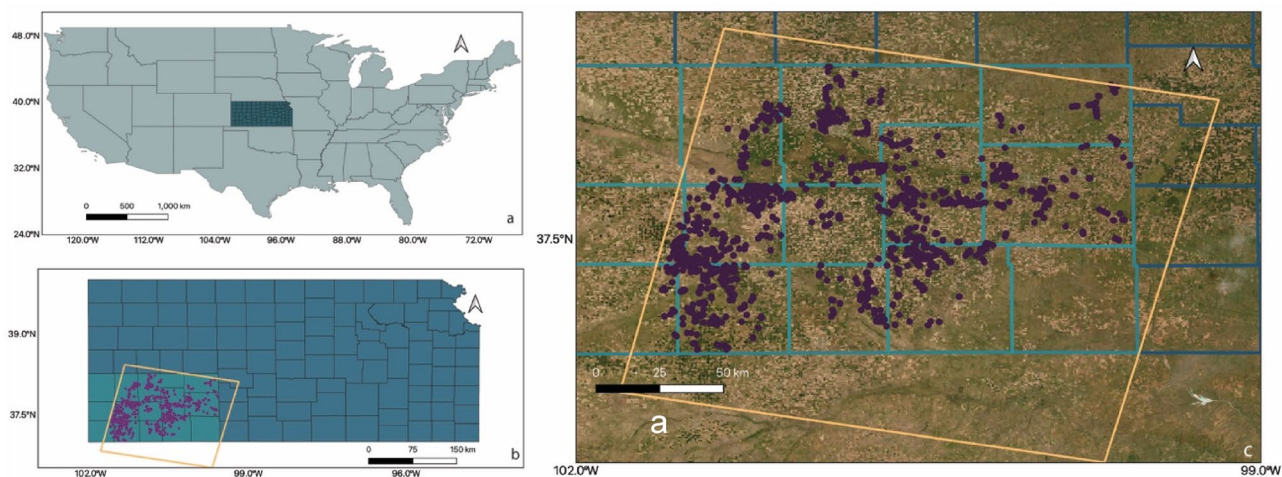


Figure 1. (a) area of study highlighting, the USA, in green and Kansas in blue (b) Area of study. The purple points correspond to phenology measurements in Kansas's Southwest Agricultural District; the orange area corresponds to Landsat 8 tile 30/34, which covers the area under study. (c) Zoomed-in highlight of the Southwest Agricultural District (light blue area) and the location of each phenology measurement. Maps were created using ArcGIS software version 10.7 by Esri. ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license. Copyright Esri. All rights reserved. For more information about Esri software, please visit www.esri.com.

The development of programs such as Google Earth Engine²¹ facilitates dataset manipulation and analysis by integrating all assets in one place. Directly working in the cloud without downloading large sets of data and using a parallel processing approach allows computation across a large number of machines²². This new tool accelerates analysis of remote sensing projects over large areas and facilitates real-time crop behavior and progress exploration²¹. Of the many tools used to perform classifications, the random forest (RF) algorithm²³ presents steady performance with large²⁴ and unbalanced datasets²⁵ in addition to modeling non-linearity correlations among the feature space and the dependent variable.

Recent studies focused on predicting field crop phenology, particularly in maize and other field crops^{2,26}, concentrated on providing estimates for dormancy, green-up, mid-season growth and development, day of senescence, and end of the season. Although these metrics are convenient from a remote sensing analytical viewpoint, more detailed phenology descriptions^{27,28} are typically desirable to provide timely actionable agricultural decisions and interventions. Due to field data availability, this study is focused on the Southwest (SW) Agricultural District in Kansas, US. Ground truth data was provided by an industry partner, comprising a large dataset on crop phenology for maize fields during the 2013–2018 growing seasons. Therefore, the aim of this research study was to evaluate a classification of satellite-derived maize crop phenology and integrate in-season weather information to develop a classification model benchmarked with field survey data. To achieve this overarching goal, we established the following objectives: (i) model training and validation—understand how different variables affect the model performance and identify the best combination of spectral features, weather parameters, geolocation and ground truth data, resulting in a model with the highest accuracy across years at each season segment (step one); and (ii) model testing—post-selection model performance evaluation for (a) each phenology class with unseen data (hold-out cross-validation); (b) temporal transferability; (c) spatial transferability (step two).

Materials and methods

Study area and environmental conditions. This research was conducted in the SW Agricultural District, comprising a total of 10 counties, Kansas, US. Annual normal precipitation in this area ranges from 381 to 635 mm from west to east, with the southwest corner (bordering the states of Oklahoma and Colorado) presenting the lowest precipitation (Fig. 1). Average mean daily temperature fluctuates from 12 to 14 °C, with the average minimum ranging from 7.5 to 15 °C and the average maximum between 19 and 22 °C²⁹. The maize growing season spans from late April (sowing) to late October (harvest), with yields usually enhanced by groundwater irrigation from the Ogallala Aquifer³⁰, which currently has more than half million of hectares under irrigation³¹.

Data gathering and feature engineering. *Surface reflectance data.* Surface reflectance data was accessed via Google Earth Engine (GEE). Landsat 8, 30 m spatial resolution, Surface Reflectance (SR) collection, tier 1, was retrieved from the GEE data repository.

Images from 2013 to 2018 from May to September (season segment, May = 1, June = 2, July = 3, August = 4, September = 5) each year were collected for the path/row 30/34. Starting the season in May and extending it until September were supported by the Start of Season Time (SOST) and End of Season Time (EOST) analyses of the NDVI each year³². Images with $\leq 30\%$ land cloud cover were selected, and remaining clouds and shadows were masked by taking advantage of the pixel quality band in the SR product. Five indices were calculated (NDVI, EVI, GCVI, NDWI and GVMi) over each of these images, and the mean value per index at each field (“[Field data collection and feature engineering](#)”) was stored by season segment.

Class	Stages on the dataset
IV1	VE–V1
IV2	V2–V4
EV	V5–V8
LV	V9–Vnth
ER	VT–R3
LR1	R4
LR2	R5
LR3	R6
H	Harvest

Table 1. Phenology classes (IV1–H) after re-grouping the phenological stages (VE–R6). *IV1* initial vegetative 1, *IV2* initial vegetative 2, *EV* early vegetative, *LV* late vegetative, *ER* early reproductive, *LR1* late reproductive 1, *LR2* late reproductive 2, *LR3* late reproductive 3, *H* harvest.

Meteorological data. Gridded surface meteorological data from GEE archives was used to extract weather variables³³. This dataset integrates high-resolution spatial data (2.5 arc minute, ~ 4 km) from PRISM³⁴ with daily temporal resolution data from the North American Land Data Assimilation System (NLDAS). The main weather variables gathered were precipitation, mean minimum and maximum temperature, and vapor pressure deficit (VPD). The Growing Degree Units (GDU) were calculated using the following equation: $GDU = [(Max. temperature (C) - Min. temperature (C))/2] - 10 \text{ } ^\circ\text{C (Base temperature)}$ ³⁵.

Field data collection and feature engineering. The initial dataset provided by Crop Quest Inc. included more than 70,000 fields distributed across the states of Kansas, Oklahoma, Colorado, and Nebraska (USA). The first step was to select only the maize data points within the SW District in Kansas, presenting more than 20,000 fields (Fig. 1). This dataset contained geolocated maize phenology measurements (latitude and longitude), associated with the day of the year (DOY), for crop data collected from the 2013 to 2018 growing seasons. To not only avoid potential classifier issues due to the specificity level of the crop growth stages, but also provide informative data more readily translated into actionable decisions for potential agricultural interventions, the stages were re-grouped into nine classes (IV1, IV2, EV, LV, ER, LR1, LR2, LR3 and H), as presented in Table 1.

Remaining data points were geographically aligned and merged with polygon shapefiles containing the field boundaries. This new file was uploaded as an asset into GEE and used to summarize the satellite and weather information within each field. Additionally, as an approximation, a filter was applied to separate the center-pivot irrigation fields from rainfed agricultural farms. This filter was built using the Canny Edge Detector³⁶. Although this technique returned nearly all fields under irrigation, some omission errors were present due to the different types of irrigation systems; only circular pivots of 500 m radius were extracted.

The final dataset included 21 independent variables and Growth Class as the dependent variable used to train and validate the models (full description in Supplementary Appendix A, Supplementary Table A).

Model training and validation, random forest classification model. Random forest (RF) has been extensively used in the remote sensing field to solve classification problems. This method has been proven to be easy to train with a little or non-computational cost while also dealing with complex interactions and highly correlated variables^{37,38}. Furthermore, RF is extensively documented in the scientific literature to outperform other methods. In the research conducted by Ref.³⁹, testing 178 classifiers, RF resulted in the classifier with the best performance. Analysis conducted by Refs.^{37,40,41} among others, presented the same outcome.

Following the abovementioned rationale, all models were trained using a RF classification algorithm from *scikit learn* software⁴² in a Python 3.8.0 environment⁴³. This algorithm was adjusted by fine-tuning hyperparameters, executing a grid search to find the best data combinations, once the best features were obtained (more details about feature selection can be found later on “Effects of feature selection on model performance”) The main hyperparameters explored in our analysis were the following: bootstrap, maximum depth of the tree, maximum number of features for the best split, minimum samples per leaf, minimum samples to perform the split, and the number of estimators²³. The best model considered the following: bootstrap = true, maximum depth of the tree = 50, maximum number of features for the best split = 9, minimum samples per leaf = 2, minimum samples to perform the split = 4, and the number of estimators = 50.

To evaluate the classifier, datasets were split into training, validation, and test sections. The proportion retained for training and validation was 80% of the total (of this 80%, 80% were used as training and 20% as validation), and 20% was used as a test for model evaluation, the numbers for each class each year can be found in Supplementary Appendix C, Supplementary Table A. An example of the geographical distribution for field observations is presented as supplementary material (Supplementary Appendix B, Fig. 1). The selected proportions resulted from a sensitivity analysis using the entire dataset spanning from 2013 to 2018 period with all variables (later referred to as the Full model). In this analysis, the overall accuracy was used as a metric to measure model performance, with data proportions adjusted at increments of 10% (from 10 to 90%, to find the optimal combination for training and validation (80/20%).

Model testing. *Performance measurements.* For this study, model classification performance was evaluated using five different metrics calculated using the test portion of the dataset. These metrics were as follows: precision, the number of true positives divided by the sum of true and false positives; recall, the number of true positives divided by the sum of true positives and false negatives; overall accuracy metric, the sum of all correctly classified elements divided by total elements; kappa coefficient of agreement, a measure describing how well the model is performing compared to the randomness (taking into account the possibility of agreement occurring by chance); and finally, F-score, the harmonic mean of the precision and recall⁴⁴. The harmonic mean is biased towards observations with lower values. When all values have the same magnitude, the harmonic and arithmetic means are equivalent. The F-score has a maximum value of 1, representing perfect precision. Therefore, the F-score is maximized by simultaneously maximizing precision and recall.

The out of bag score (OOB) was computed in addition to evaluating the performance of the model. This measure computes the average misclassification ratio of non-training samples. The importance of this metric relies on offering an unbiased assessment of classification performance, with values closer to 1 indicating good performance³⁷.

Effects of feature selection on model performance. The openly exploration of different combinations of variables to address model performance was motivated by two main reasons. First to understand how different features affect the performance of the model⁴⁵; second to overcome any potential bias resulting from a feature importance analysis. This technique has been proven to favor variables presenting high cardinality^{46,47}, condition present on the dataset for geographical features as well as growth DOY. Nonetheless a feature importance was calculated for one of the models to observe the behavior.

To test this abovementioned point, the following variable combinations were used for each year and for each season segment: all bands (B) including spectral bands 2, 3, 4, 5, 6, 7, 10, and 11 for Landsat 8; all bands and all weather parameters (BW), including the mentioned spectral bands, precipitation, maximum and minimum temperature, VPD, and GDU; all indices and all bands (VIB), including the mentioned spectral bands and NDVI, EVI, GCVI, GVMI, and NDWI as vegetation indices; all indices and weather parameters (VIW), including the five above mentioned weather parameters and five vegetation indices; only weather parameters (W), testing five weather parameters; only indices (VI), including only five vegetation indices; indices, weather parameters, and bands together (FULL), and finally a model only containing latitude, longitude and day of the year (LLD). The models most stable across years and within season (VIW; W; VI) were trained again analyzing all possible interactions between the variables presented. Thus, we explored 31 combinations for the W models, 31 combinations for the VI models, and 1023 combinations for the VIW models. The number of models actually fitted was greater than the previous sum (31 + 31 + 1023) due to the grid search approach for fine-tuning the hyperparameters. In order to account for spatio-temporal correlation in the response variable induced by other features that may not be captured by remote sensing (surface reflectance) and weather data, all model combinations included field geolocation and DOY⁴⁸.

Because not all crop phenology classes were present between season segments due to the characteristics of this variable, the accuracy values for each model during each season segment for each year were retrieved to better understand the performance within and between years.

Model performance under temporal and spatial transferability. Following the completion of the first stage, the subsequent analyses were established to better understand the performance and stability of the final model. To learn about the stability of the model across years, the dataset was divided into training and validation based on year separation. For this evaluation, we grouped the 6 years (2013–2018) according to the weather trend combining average years (2014, 2017) with dry (2013, 2018) and wet ones (2015, 2016). The same metrics (Overall accuracy, OOB, precision, recall and F1-score) were utilized to test this analysis.

To explore the ability of the model while leading with spatial transferability³⁷, the dataset was divided into training and validation based on spatial attributes, with a total of 40,519 elements (sum of all the phenological measurements across the 6 years). This resulted on two datasets, a larger one comprised by the fields located in the western part of the area under study (training) where fields are more aggregated. The validation dataset was smaller and composed by the fields in the east side of the geographic area.

Results

Best combination of features and model performance. All models explored in this study classified crop phenology classes with an accuracy above 70%, except for the model combining all 21 variables (accuracy below 60%) (Fig. 2). For this model, we also conducted a feature importance analysis (Supplementary Appendix B). Although most of the variables shared a similar importance, some were more relevant, such as latitude, longitude and growth doy, associated to each phenology measurement (variables with high cardinality), while others such as precipitation were deleterious. For remaining models, accuracy ranged from 70 to 100% within a season and between years. The models with the best performance between years and within season were (I) combination between vegetation indices and weather, (II) weather (III) vegetation indices, and (IV) latitude, longitude and doy. For these four models, all possible combinations between variables were tested. This resulted in two models with the best performances, one composed by 5 vegetation indices and only one weather parameter, VPD, where the accuracy assessment ranged from 86 to 98%; and the other model composed only by latitude, longitude and doy, with an accuracy of 100% for almost all the season segments—years (Fig. 2). Further discussion on this model is presented in the discussion, later on this manuscript.

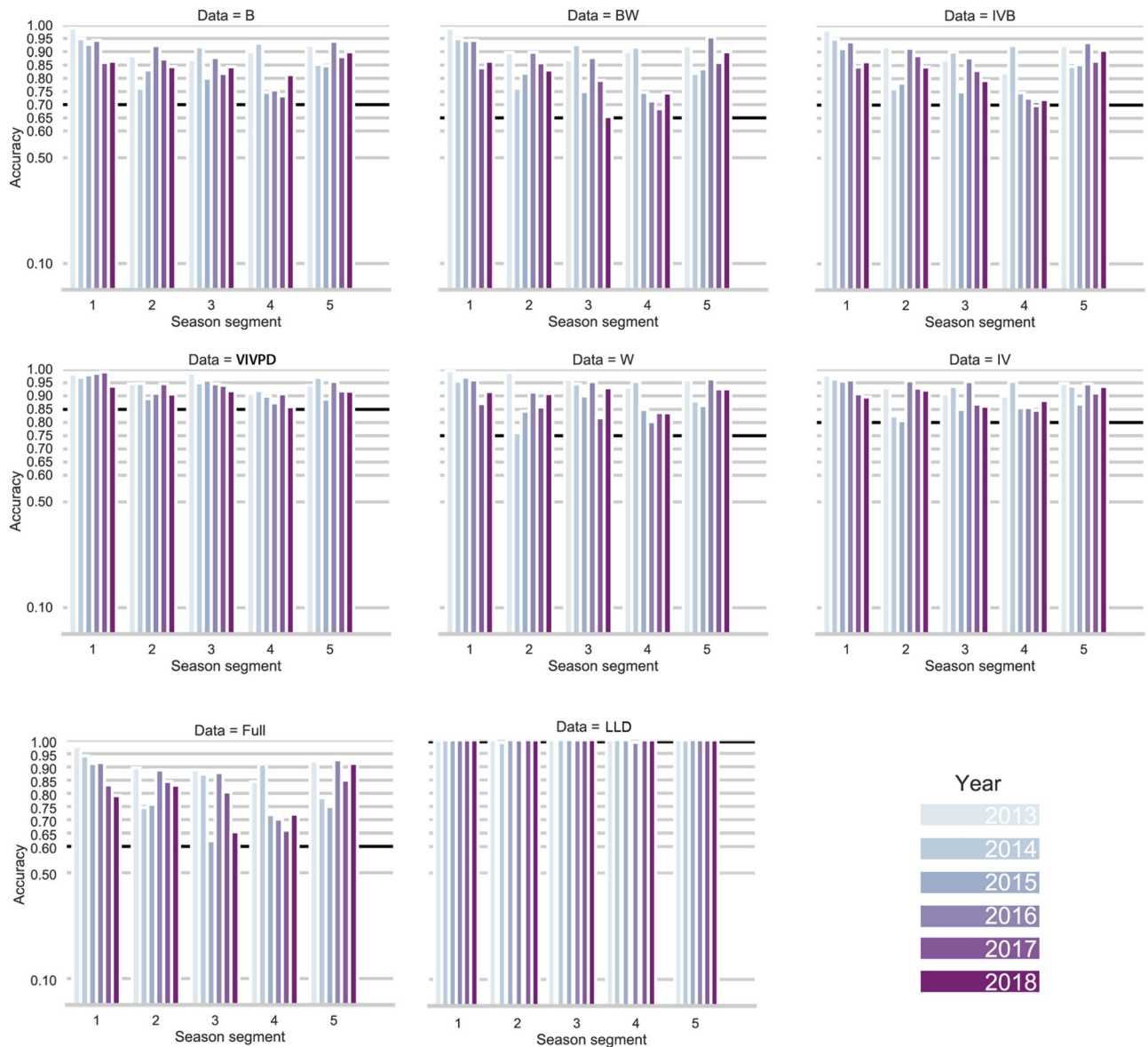


Figure 2. Accuracy values for each model during the season from 2013 to 2018. The black horizontal line shows the lowest value recorded. *B* spectral bands 2, 3, 4, 5, 6, 7, 10, and 11 for Landsat 8, *BW* spectral bands in addition to precipitation, maximum and minimum temperature, VPD, and GDU (weather parameters), *VIB* spectral bands in addition to NDVI, EVI, GCVI, GVMI, and NDWI (vegetation indices), *VIVPD* vegetation indices and vapor pressure deficit (VPD), *W* weather parameters, *VI* vegetation indices, *FULL* spectral bands, weather parameters, and vegetation indices, *LLD* latitude, longitude and doy.

Classification using VIVPD model. The combination of features optimizing model performance was the following: DOY, EVI, NDVI, NDWI, GCVI, GVMI, VPD, latitude, and longitude. A more detailed view of model performance, spanning all growing seasons from 2013 to 2018 period, is presented as a confusion matrix (Fig. 3). The overall accuracy for this classifier was 94%, (OOB 0.94), and the Kappa coefficient was 0.93. Further details for each year and class are presented in Supplementary Appendix C, Supplementary Table A. Overall, model behavior was similar across classes and years except when the number of elements (supports) was small (<10). The latter scenario can be visualized with class H for the 2018 year, where only 6 elements were present, and the metrics were considerably smaller compared to other classes with more elements.

Classification using VIVPD model across years. Table 2 presents the results when the model was transferred across years. In this analysis the years 2014, 2015, 2017 and 2018 were part of the training dataset, and 2013 and 2016 (dry and wet years respectively) used as validation dataset. The overall accuracy was 93 percent, which was close to the hold-out cross-validation accuracy (Fig. 3). Furthermore, all of the parameters for each class were similar, if not better, with precision values ranging from 0.85 to 1, recall values ranging from 0.85 to 0.99, and f1 scores ranging from 0.86 to 0.97.

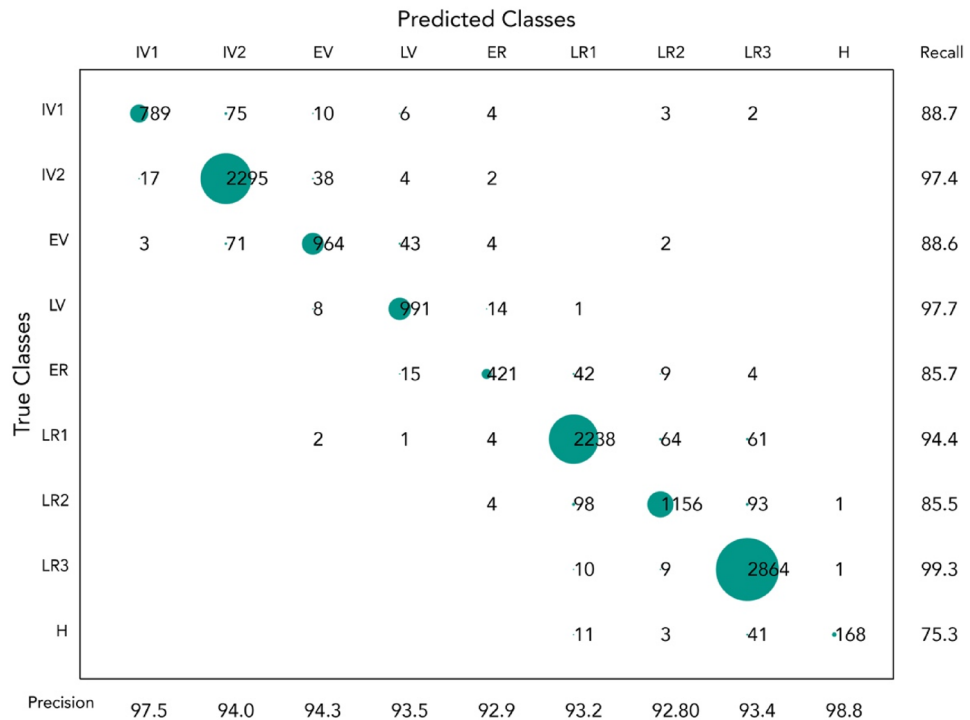


Figure 3. For the years 2013 to 2018, a classification matrix was created that included all classified elements, their magnitude, as well as precision and recall scores. The true classes are on the Y axis, while the model’s predictions are on the X axis. The number of fields classified is represented visually by the size of the bubbles; the numbers along the main diagonal represent those successfully classified.

Class	f1-score	Precision	Recall
IV1	0.91	0.98	0.85
IV2	0.94	0.94	0.94
EV	0.86	0.85	0.86
LV	0.91	0.85	0.96
ER	0.92	0.94	0.91
LR1	0.96	0.98	0.94
LR2	0.88	0.88	0.88
LR3	0.97	0.95	0.99
H	0.95	1.0	0.90
Accuracy	0.93		
OOB	0.94		

Table 2. Classification metrics when 2014, 2015, 2017, and 2018 years were used to train the model, and 2013 and 2016 as validation dataset. A total of 14,431 elements (corresponding to the phenological measurements for the years 2013 and 2016) were used for validation and a total of 26,088 elements (corresponding to the phenological measurements for the years 2013, 2015, 2017 and 2018) were used for training.

In addition to this split of years, other combinations were tested to observe the behavior when using only wet years for the validation (2015 and 2016) and not present in the training set, and the opposite, allowing only dry years for the validation (2013 and 2018). In both scenarios the overall accuracy was 0.94 for the first approach and 0.95 for the second, and OOB score of 0.95 for both.

Classification using VIVPD model across space. The model tested for the spatial transferability of RF resulted on lower values of overall accuracy and low metrics for some classes, such as H, LR2, ER, intermediate for IV1, EV, LR1 and better for IV2, LV and LH3. In all the cases the values were lower when compared against the models previously mentioned (Table 3).

Class	f1-score	Precision	Recall
IV1	0.60	0.97	0.43
IV2	0.83	0.77	0.90
EV	0.67	0.73	0.62
LV	0.85	0.80	0.91
ER	0.40	0.60	0.30
LR1	0.65	0.55	0.78
LR2	0.33	0.68	0.22
LR3	0.85	0.76	0.96
H	0.23	0.40	0.15
Accuracy	0.72		
OOB	0.75		

Table 3. Classification metrics when testing for spatial transferability using all the years.

Discussion

The proposed approach to characterize maize phenology presented the following main advantages: (i) use of publicly available data as model input, (ii) prediction of phenology using high-spatial resolution (30 m), and (iii) use of random forest for classification, allowing the trained model to be easily deployed in GEE for efficient computing processing. These aspects, associated with the stable model performance of phenology prediction when transferred across years, ensures high generalization power. Future uses can be easily adapted for using higher spatial resolution satellite data (e.g., Sentinel 2, 10 m), providing better opportunities for small-scale farming in developing countries.

In exploring a wide feature space for model training and validation (5 weather variables, 5 VIs and 8 surface reflectance bands, latitude, longitude, and DOY for image collection), we found that the full model (considering all 21 predictors) did not perform well. The inferior results for the full model were potentially related to the high correlation among the surface reflectance bands. Predictive models suffer from multicollinearity issues when the independent variables are correlated, causing unpredictable variance in the model outputs (overfitting)^{49,50}. In the opposite scenario, a model based solely on geographical features and dates created a potentially spurious ideal classification. When dealing with classification problems, this situation is often avoided by dropping or transforming the features with high cardinality⁴⁸. Even yet, this aspect requires further consideration, particularly when it comes to classification studies integrating agronomical data.

Alternatively, we documented superior performance when VIs and VPD were combined. In addition to VPD^{51–53}, VIs NDVI, EVI, and GCVI have been reported as useful in predicting crop development during the growing season, especially for maize and soybean (*Glycine max* L.)^{2,17,30}. Furthermore, latitude, longitude, and DOY are strong indicators of how weather and solar radiation patterns influence crop development^{54,55} and dropping them for the rest of the models could result on losing relevant information.

From a testing standpoint, the model comprised of VI and VPD was stable across all phenology classes, resulting in high classification metrics for the vast majority of them, even in years when weather dynamics substantially impacted crop progress^{56,57} and when tested explicitly for temporal transferability. This latter becomes even more relevant if we consider that the classes correctly classified in all the scenarios correspond with the key developmental stages of the crop, such as emergence, effective density, and yield definition.

These findings align with what was reported by Refs.^{45,58}. It is worth mentioning that at least one quality image (with < 30% cloud cover) was retrieved each month between May and September in every year. Although images were available, data retrieved from certain fields was not sufficient to balance the dataset with regard to phenology classes due in part to the characteristics of the ground truth dataset. This resulted in good precision for classes with more observations but poor performance for the underrepresented classes with a smaller number of data points^{59,60}. The underperformance of the model transferability across space is also congruent with findings from Refs.^{37,45,61,62}. According to Refs.^{58,61} this can be overcome when the territory remains similar as the training dataset, or they share similar weather parameters. Although this could be true for land classification problems, our findings suggest that phenology classification can be more susceptible to these changes.

The lack of quality satellite imagery data, the resulting decrease in the number of classified fields, and the changing environment with the consequent data shift could be remediated with at least three possible solutions not currently addressed by this study. First, thanks to the ability to obtain regular intervals of data for time series analysis, a fusion of satellites^{27,63} would facilitate retrieval of higher-quality data, with improved spatio-temporal resolution preventing data loss, and a better scenario for smoothing techniques such as Savitzky–Golay fitting, locally weighed regression, spline smoothing, and others. Second, because not all classes contained a balanced number of elements, different methods can be applied to managed unbalanced data^{64,65} and ultimately result in a more robust classifier. Lastly, integration of remote sensing data and crop model outputs (i.e., mechanistic- or process-based models) could aid in maximizing predictability power and spatio-temporal limits^{66–71}.

Conclusions

Crop phenology monitoring is crucial for agricultural management since it enables growers, stakeholders, policymakers, and government agencies to determine when the most critical stages are occurring particularly during non-ideal conditions, due to biotic or abiotic stress, where the theory departs from the reality. Is particularly in those cases where models like the one described on this paper can facilitate the report of phenology, and to take more inform decisions. Improving crop phenology classification is becoming more prevalent as current and future research focuses on satellite data fusion and the use of mechanistic or process-based models to enhance spatio-temporal resolution. The findings presented on this paper showed strong classification metrics across years and proven the ability of spectral and weather features to assist in phenology classification. The proposed model also has been tested for spatial transferability and although promising future work should be put on understand the mechanism behind the model behavior.

Received: 27 April 2021; Accepted: 21 July 2021

Published online: 03 August 2021

References

1. USDA-NASS., 2019. Crop Production 2018 Summary. (February), 1–132. https://www.nass.usda.gov/Publications/Todays_Reports/reports/cropan19.pdf.
2. Zhong, L., Hu, L., Yu, L., Gong, P. & Biging, G. S. Automated mapping of soybean and corn using phenology. *ISPRS J. Photogramm. Remote Sens.* **119**, 151–164. <https://doi.org/10.1016/j.isprsjprs.2016.05.014> (2016).
3. Carletto, C., Gourlay, S. & Winters, P. *From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis* (The World Bank, 2013).
4. Gourlay, S., Kilic, T. & Lobell, D. *Could the Debate be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda* (The World Bank, 2017).
5. Cai, Y. *et al.* Detecting in-season crop nitrogen stress of corn for field trials using UAV-and CubeSat-based multispectral sensing. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **12**, 5153–5166. <https://doi.org/10.1109/JSTARS.2019.2953489> (2019).
6. Morris, T. F. *et al.* Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement. *Agron. J.* **110**(1), 1–37. <https://doi.org/10.2134/agronj2017.02.0112> (2018).
7. Ruml, M. & Vucic, T. Importance of phenological observations and predictions in agriculture. *J. Agric. Sci. Belgrade* **50**(2), 217–225. <https://doi.org/10.2298/JAS0502217R> (2005).
8. Duncan, J., Dash, J. & Atkinson, P. M. The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. *Front. Environ. Sci.* **3**, 56. <https://doi.org/10.3389/fevns.2015.00056> (2015).
9. Bolton, D. K. & Friedl, M. A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* **173**, 74–84. <https://doi.org/10.1016/j.agrformet.2013.01.007> (2013).
10. Sakamoto, T. *et al.* A crop phenology detection method using time-series MODIS data. *Remote Sens. Environ.* **96**(3–4), 366–374. <https://doi.org/10.1016/j.rse.2005.03.008> (2005).
11. Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **8**(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0) (1979).
12. Huete, A. *et al.* Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **83**(1), 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2) (2002).
13. Gitelson, A. A. *et al.* Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* **30**(5), 1248. <https://doi.org/10.1029/2002GL016450> (2003).
14. Ceccato, P., Gobron, N., Flasse, S., Pinty, B. & Tarantola, S. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. *Remote Sens. Environ.* **82**, 188–197. [https://doi.org/10.1016/S0034-4257\(02\)00037-8](https://doi.org/10.1016/S0034-4257(02)00037-8) (2002).
15. Gao, B. C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **58**(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3) (1996).
16. Huete, A., Justice, C. & Liu, H. Development of vegetation and soil indices for MODIS-EOS. *Remote Sens. Environ.* **49**, 224–234. [https://doi.org/10.1016/0034-4257\(94\)90018-3](https://doi.org/10.1016/0034-4257(94)90018-3) (1994).
17. Nguy-Robertson, A. *et al.* Green leaf area index estimation in maize and soybean: Combining vegetation indices to achieve maximal sensitivity. *Agron. J.* **104**, 1336–1347. <https://doi.org/10.2134/agronj2012.0065> (2012).
18. Schlemmer, M. *et al.* Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *Int. J. Appl. Earth Observ. Geoinf.* **25**, 47–54. <https://doi.org/10.1016/j.jag.2013.04.003> (2013).
19. Cai, Y. *et al.* Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* **274**, 144–159. <https://doi.org/10.1016/j.agrformet.2019.03.010> (2019).
20. Sakamoto, T., Gitelson, A. A. & Arkebauer, T. J. Near real-time prediction of US corn yields based on time-series MODIS data. *Remote Sens. Environ.* **147**, 219–231. <https://doi.org/10.1016/j.rse.2014.03.008> (2014).
21. Gorelick, N. *et al.* Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031> (2017).
22. Kennedy, R. E. *et al.* Implementation of the LandTrendr algorithm on google earth engine. *Remote Sens.* **10**(5), 691. <https://doi.org/10.3390/rs10050691> (2018).
23. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. <https://doi.org/10.1023/A:1010933404324> (2001).
24. Zhong, L., Yu, L., Li, X., Hu, L. & Gong, P. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. *Sci. Rep.* **6**(1), 1–14. <https://doi.org/10.1038/srep36240> (2016).
25. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **26**(1), 217–222. <https://doi.org/10.1080/01431160412331269698> (2005).
26. Viña, A. *et al.* Monitoring maize (*Zea mays* L.) phenology with remote sensing. *Agron. J.* **96**, 1139–1147. <https://doi.org/10.2134/agronj2004.1139> (2004).
27. Liao, C. *et al.* Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass and yield estimates for corn and soybean. *Sci. Total Environ.* **650**, 1707–1721. <https://doi.org/10.1016/j.scitotenv.2018.09.308> (2019).
28. Ghamghami, M., Ghahreman, N., Irannejad, P. & Pezeshk, H. A parametric empirical Bayes (PEB) approach for estimating maize progress percentage at field scale. *Agric. For. Meteorol.* **281**, 107829. <https://doi.org/10.1016/j.agrformet.2019.107829> (2020).
29. Goodin, D. G., Mitchell, J. E., Knapp, M. C. & Bivens, R. E. Climate and weather atlas of Kansas—An introduction: Kansas Geological Survey. *Educ. Ser.* **12**, 24 (1995).
30. Masiale, I., Egbert, S. & Wardlow, B. D. A comparative analysis of phenological curves for major crops in Kansas. *GISci. Remote Sens.* **47**, 241–259. <https://doi.org/10.2747/1548-1603.47.2.241> (2010).

31. Lanning-Rush, J. L. *Irrigation Water Use in Kansas, 2013* (No. 981). (US Geological Survey, 2016).
32. Hao, P., Zhan, Y., Wang, L., Niu, Z. & Shakir, M. Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA. *Remote Sens.* **7**(5), 5347–5369 (2015).
33. Abatzoglou, J. T. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* <https://doi.org/10.1002/joc.3413> (2012).
34. Daly, C. *et al.* Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.* **28**, 2031–2064. <https://doi.org/10.1002/joc.1688> (2008).
35. Gilmore, E. C. & Rogers, J. S. Heat units as a method of measuring maturity in corn 1. *Agron. J.* **50**, 611–615. <https://doi.org/10.2134/agronj1958.00021962005000100014x> (1958).
36. Canny, A. Computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698. <https://doi.org/10.1109/TPAMI.1986.4767851> (1986).
37. Belgiu, M. & Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31 (2016).
38. Zhang, L. *et al.* Identification of seed maize fields with high spatial resolution and multiple spectral remote sensing using random forest classifier. *Remote Sens.* **12**(3), 362 (2020).
39. Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. Do we need hundreds of classifiers to solve real world classification problems?. *J. Mach. Learn. Res.* **15**(1), 3133–3181 (2014).
40. Gómez, C., White, J. C. & Wulder, M. A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **116**, 55–72 (2016).
41. Talukdar, S. *et al.* Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sens.* **12**(7), 1135 (2020).
42. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
43. Van Rossum, G. Python programming language. In *USENIX Annual Technical Conference*, vol. 41, 36 (2007).
44. Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **37**(1), 35–46 (1991).
45. Cai, Y. *et al.* A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* **210**, 35–47 (2018).
46. Strobl, C., Boulesteix, A. L., Zeileis, A. & Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **8**(1), 1–21 (2007).
47. Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **9**(1), 1–11 (2008).
48. Moeyersoms, J. & Martens, D. Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector. *Decis. Support Syst.* **72**, 72–81 (2015).
49. Dormann, C. F. *et al.* Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x> (2013).
50. Schwalbert, R. A. *et al.* Forecasting maize yield at field scale based on high-resolution satellite imagery. *Biosyst. Eng.* **171**, 179–192. <https://doi.org/10.1002/csc2.20053> (2018).
51. Basso, B. & Ritchie, J. T. Evapotranspiration in high-yielding maize and under increased vapor pressure deficit in the US Mid-west. *Agric. Environ. Lett.* **3**(1), 170039. <https://doi.org/10.2134/ael2017.11.0039> (2018).
52. Messina, C. D. *et al.* Limited-transpiration trait may increase maize drought tolerance in the US Corn Belt. *Agron. J.* **107**, 1978–1986. <https://doi.org/10.2134/agronj15.0016> (2015).
53. Fletcher, A. L., Sinclair, T. R. & Allen, L. H. Jr. Transpiration responses to vapor pressure deficit in well-watered “slow-wilting” and commercial soybean. *Environ. Exp. Bot.* **61**, 145–151. <https://doi.org/10.1093/jxb/erp350> (2007).
54. Deines, J. M. *et al.* Mapping three decades of annual irrigation across the US High Plains Aquifer using Landsat and Google Earth Engine. *Remote Sens. Environ.* **233**, 111400 (2019).
55. Jin, Z., Azzari, G. & Lobell, D. B. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. *Agric. For. Meteorol.* **247**, 207–220 (2017).
56. White, M. A., Thornton, P. E. & Running, S. W. A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Glob. Biogeochem. Cycles* **11**(2), 217–234. <https://doi.org/10.1029/97GB00330> (1997).
57. Gao, F. *et al.* A within-season approach for detecting early growth stages in corn and soybean using high temporal and spatial resolution imagery. *Remote Sens. Environ.* **242**, 111752. <https://doi.org/10.1016/j.rse.2020.111752> (2020).
58. Wang, S., Azzari, G. & Lobell, D. B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* **222**, 303–317 (2019).
59. Japkowicz, N. & Stephen, S. The class imbalance problem: A systematic study. *Intell. Data Anal.* **6**, 429–449. <https://doi.org/10.3233/IDA-2002-6504> (2002).
60. Shan, Y., Paull, D. & McKay, R. I. Machine learning of poorly predictable ecological data. *Ecol. Model.* **195**(1–2), 129–138. <https://doi.org/10.1016/j.ecolmodel.2005.11.015> (2006).
61. Pelletier, C., Valero, S., Inglada, J., Champion, N. & Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **187**, 156–168 (2016).
62. Jin, S. *et al.* The transferability of Random Forest in canopy height estimation from multi-source remote sensing data. *Remote Sens.* **10**(8), 1183 (2018).
63. Bhowan, U., Zhang, M., & Johnston, M. Genetic programming for image classification with unbalanced data. In *2009 24th International Conference Image and Vision Computing New Zealand* 316–321. (IEEE, 2009). https://doi.org/10.1007/978-3-642-12148-7_1.
64. Gao, F. *et al.* Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* **188**, 9–25. <https://doi.org/10.1016/j.rse.2016.11.004> (2017).
65. Kitchenham, B. A procedure for analyzing unbalanced datasets. *IEEE Trans. Software Eng.* **24**(4), 278–301. <https://doi.org/10.1109/32.677185> (1998).
66. Dal Pozzolo, A., Caelen, O., Johnson, R.A., & Bontempi, G. Calibrating probability with under sampling for unbalanced classification. In *2015 IEEE Symposium Series on Computational Intelligence* 159–166. (IEEE, 2015). <https://doi.org/10.1109/SSCI.2015.33>.
67. Azzari, G., Jain, M. & Lobell, D. B. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sens. Environ.* **202**, 129–141. <https://doi.org/10.3390/rs10091489> (2017).
68. Jin, Z., Azzari, G., Burke, M., Aston, S. & Lobell, D. Mapping smallholder yield heterogeneity at multiple scales in Eastern Africa. *Remote Sens.* **9**, 931. <https://doi.org/10.3390/rs9090931> (2017).
69. Lobell, D. B., Thau, D., Seifert, C., Engle, E. & Little, B. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* **164**, 324–333. <https://doi.org/10.1016/j.rse.2015.04.021> (2015).
70. Sibley, A. M., Grassini, P., Thomas, N. E., Cassman, K. G. & Lobell, D. B. Testing remote sensing approaches for assessing yield variability among maize fields. *Agron. J.* **106**, 24–32. <https://doi.org/10.2134/agronj2013.0314> (2014).
71. Clevers, J. G. P. W. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sens. Environ.* **61**, 221–228. [https://doi.org/10.1016/S0034-4257\(97\)00004-7](https://doi.org/10.1016/S0034-4257(97)00004-7) (1997).

Acknowledgements

This work was supported by the Kansas Corn Commission. The datasets with the field ground truth data were provided by Crop Quest Inc. This is contribution no. 21-689-J from Kansas Agricultural Extension Station. Special thanks to Ms. Rachel Veenstra for her valuable feedback.

Author contributions

Conceptualization, L.N., R.S., and I.A.C.; methodology, L.N. and R.S.; formal analysis, L.N.; resources, L.N., R.S. and I.A.C.; writing—original draft preparation, L.N.; writing—review and editing, R.S., P.V.V.P., B.O., I.A.C.; supervision, R.S. and I.A.C.; funding acquisition, I.A.C. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Kansas Corn Commission and Kansas State University provided funds to complete this work, supporting Mrs. Nieto's stipend and Dr. Ciampitti's research program.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-95253-7>.

Correspondence and requests for materials should be addressed to L.N. or I.A.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021