

**Enhanced temporal resolution of satellite imagery can assist in improving maize
crop phenology classification**

Luciana Nieto^{1*}, names, P. V. Vara Prasad ^{1,3}, Bradley J.S.C. Olson⁴, Ignacio A.
Ciampitti^{1*}

^{1*} Department of Agronomy, 2004 Throckmorton Plant Science Center, 1712 Claflin Road,
Kansas State University, Manhattan, Kansas 66506, US.

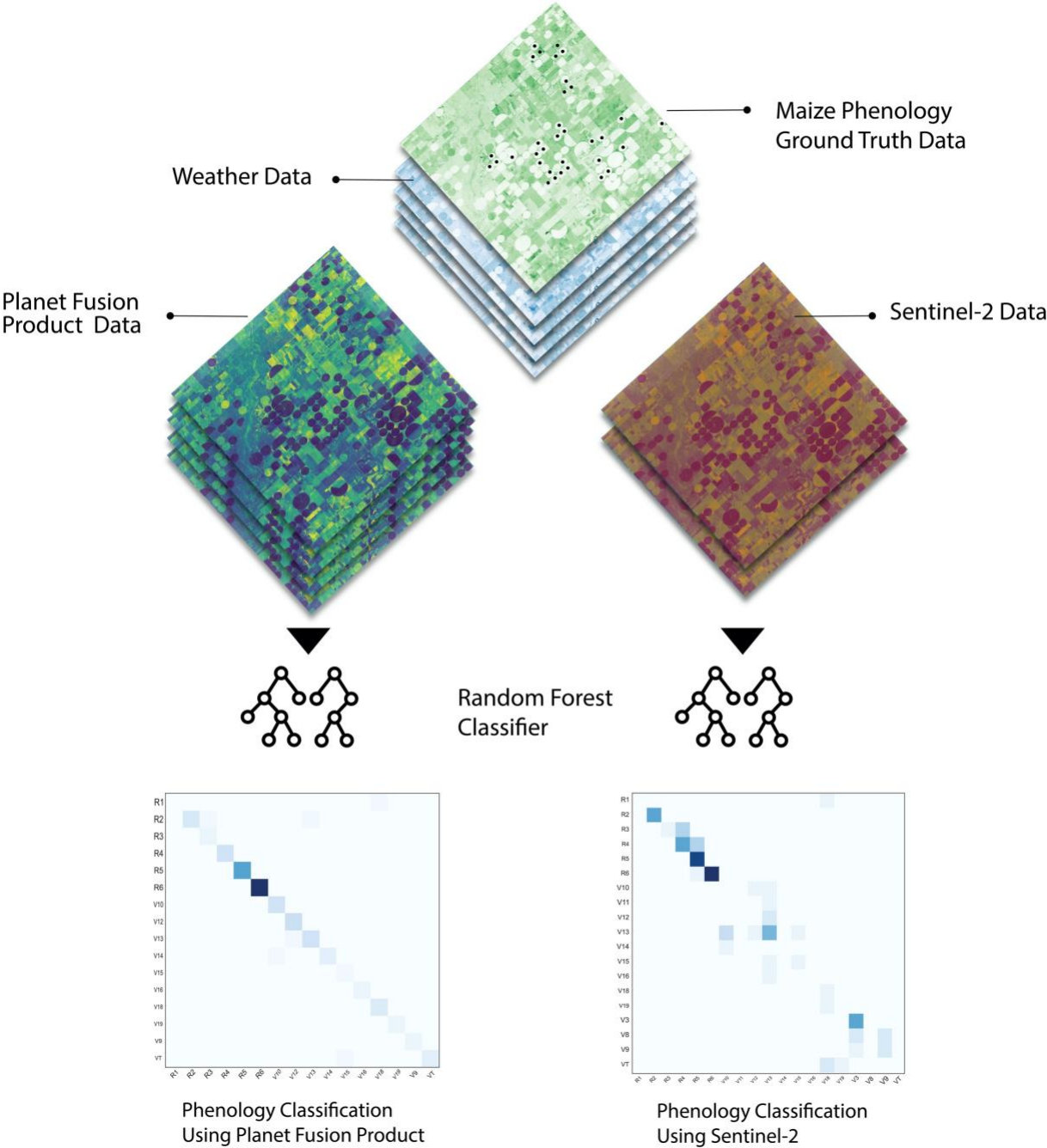
² Planet, --

³ Sustainable Intensification Innovation Lab, 108 Waters Hall, 1603 Old Claflin Place,
Kansas State University, Manhattan, Kansas 66506, US.

⁴ Department of Biology, Chalmers Hall, 1711 Claflin Road, Kansas State University,
Manhattan, Kansas, 66506, US.

Target Journal: SPRS Journal of Photogrammetry and Remote Sensing

19 **Graphical Abstract**



20
21
22
23
24
25

Abstract

Improving accuracy to assess crop phenology and their space-time fluctuations are essential for farmers, policy makers, and government agencies. Weekly field survey data from extension agricultural agents is currently utilized to generate reports on progress estimates of crop phenology. Although this data is critical, it is labor-intensive, time-consuming, and prone to human error. Remote sensing can improve the accuracy of these estimates via reducing bias and missing data, but this process requires gathering high observational frequency and spatial resolution, combining multi-sensor data streams. Robust classifiers (e.g., random forest, RF) are also a key component for handling large datasets with acceptable trade-off and stability. The aim of this manuscript focuses on comparing the output of RF classifier model (dealing with imbalanced datasets) testing two different satellite sources (Planet Fusion, PF and Sentinel-2, S-2) in combination with weather data to improve maize (*Zea mays* L.) crop phenology classification evaluated in two regions (southwest and central) in Kansas (US) during the 2017 season. Our findings showed that the use of very high temporal resolution resulted on higher classification metrics (f1-score = 0.94) in both regions when compared to Sentinel-2 (S-2) (f1-score = 0.86). In addition, a temporal sensitivity analysis over S-2, showed a drop in f1-score with values of 0.74 and 0.60. This research establishes the value of having access to very high temporal resolution (daily basis) for crop monitoring and its effect on actionable insights in the agricultural decision-making process.

Key words: maize phenology; Planet Fusion; Sentinel-2; Random Forest classifier.

1. Introduction

Crop phenology is the study of biological processes such as emergence, flowering, and senescence, linked and in response to environmental growing conditions (Liang et al., 2011). Understanding changes in crop phenology and their response to climatic conditions are critical for production-beneficial management practices measures (Ruml and Vulic, 2005). Thus, understanding crop phenology is a critical step in ensuring our ecosystems' long-term sustainability and viability.

Current official phenology crop progress estimates in the United States are based on weekly survey data obtained from a large network of regional extension agricultural agents based on their periodic field observations (Gao and Zhang, 2021). Although this is a valuable source of knowledge, the data collection process is a time-consuming and labor-intensive task that may not accurately represent, with a high level of fidelity, a county or district (Gao et al., 2017).

In addition to ground-based field observations, implementing remote sensing technologies for Earth observation has been proved to be a useful tool, giving spatial and temporal information for vegetation monitoring across vast regions of land (White et al., 1997; Zhang et al., 2003; Vina et al., 2004). Current sensors and future missions result in a large amount of available data with countless temporal, spectral, and spatial resolutions (Houborg et al., 2015; Rast and Painter, 2019; Wulder et al., 2019). This observational frequency is one of the most important factors to address crop phenology progress and monitor changes in vegetation dynamics at both spatial and temporal level (Gao and Zhang, 2021).

Although a large progress for monitoring crop phenology has been achieved, sensors with daily or close to daily coverage are not available at the spatial resolution needed for field analysis, while those sensors with high spatial resolution present a low revisit period. Several efforts are being conducted to merge those characteristics, scattered through different sensors, into one product capable to retrieve high temporal and spatial resolution (Gao et al., 2017; Claverie et al., 2018; Liao et al., 2019). Data fusion is currently the best path to achieve within season analysis capable of characterizing vegetation dynamics in both large- and small-scale environments or landscapes (Gao and Zhang, 2021).

Robust classifiers are needed to manage large datasets with acceptable trade-offs and stability while remaining computationally feasible (Gislason et al., 2006). For decades, Random Forest (RF, Breiman, 2001) has been proved to be useful for remote sensing applications (Pal, 2005; Rodriguez-Galiano et al., 2012; Belgiu and Drăguț, 2016) producing superior classification metrics relative to other classifiers, mainly when dealing with large and unbalanced datasets (Fernández-Delgado et al., 2014). These characteristics, along with their simplicity and processing speed, make RF the classifier of choice in a wide variety of scenarios.

This manuscript focuses on comparing the output of RF classifier model testing two different sources of satellite varying in their temporal and spatial resolution in combination with weather data to improve maize (*Zea mays* L.) crop phenology classification. The specific objectives for this study were: i) to utilize surface reflectance data from Planet's data fusion product (PF-SR) and ESA's Sentinel-2 sensors to quantify performance for crop phenology classification; and ii) to evaluate the universality of the model by testing its performance in two distinct maize production regions, southwest: mainly irrigated with

average yields $>11 \text{ Mg ha}^{-1}$; and central: rainfed, yields $<9 \text{ Mg ha}^{-1}$, within the state of Kansas (US) during the 2017 growing season.

2. Materials and methods

2.1. Study area

The present study was conducted in the State of Kansas (US), with the southwest (SW) region comprising the counties of Morton, Stanton, Grant, and Stevens and the central (CK) region constituted by the counties of Stafford and Pratt (Figure 1, a)

The precipitation pattern in Kansas is highly uneven in both quantities and distribution, ranging from over 1000 millimeters from the southeast to less than 300 millimeters in the west region of the State. For the two regions selected in this study, these levels ranged from 500 to 800 mm for the CK (rainfed region) and from 300 to 500 mm for the SW region (Lin et al., 2017). The size of the fields, as well as standard management practices such ground water irrigation, differ in these two regions. The SW region is mostly dominated by larger fields relative to the CK, and more concentration of irrigation pivots (fed from the Ogallala Aquifer), resulting in higher maize yields, average 11 Mg ha^{-1} for year 2019 (USDA-NASS, n.d.). In contrast, CK region is characterized for smaller fields and mainly dominated by rainfed agriculture, with average maize yield of 9 Mg ha^{-1} for the year 2019 (USDA-NASS,).

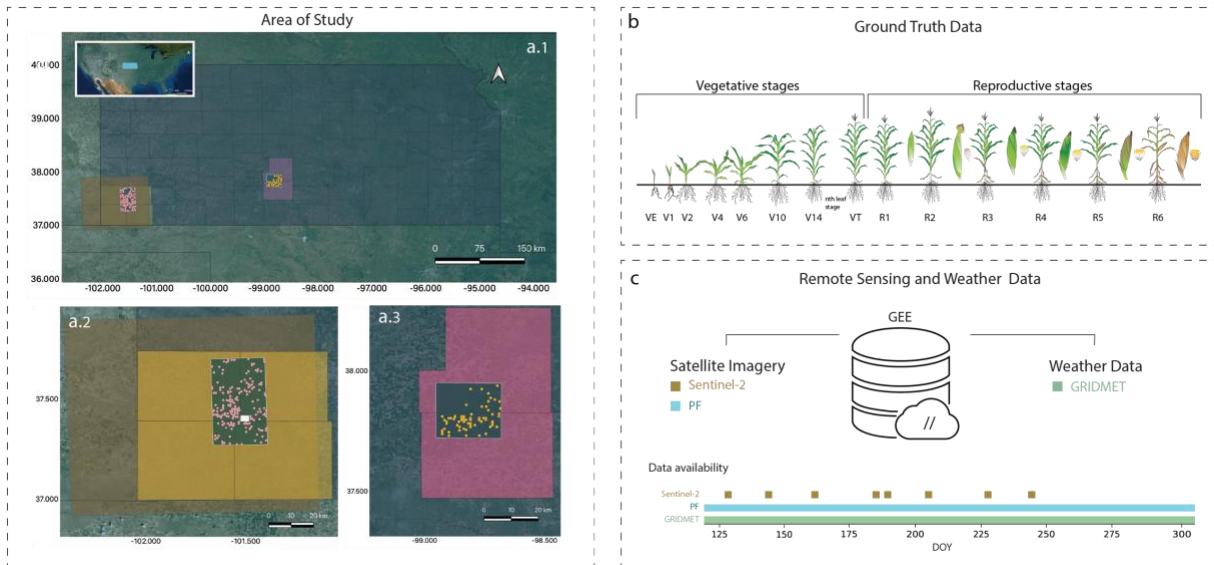


Figure 1. a) Area of study (a.1) state of Kansas (US), a.2) southwest (SW) region; large orange rectangle representing Sentinel-2 tile covering the area; green rectangle corresponds with the Planet fusion (PF) tiles over the area; pink dots correspond to the ground truth field data collection, a.3) central KS (CK) region; green square corresponds with the PF tile over the area; yellow dots correspond to the ground truth field data collection. b) Maize phenology stages, scale used during ground truth data collection, following Ciampitti et al. (2011). c) Sentinel-2 and PF data imagery and GRIDMET data from Google Earth Engine (GEE) repository, data availability differs temporally with 8 images collected for Sentinel-2, and daily temporal data for both PF and weather data during the 2017 growing season, varying with the day of the year (DOY).

2.2. Data characteristics

2.2.1. Reference ground field data

Acquiring quality environmental data may be challenging and expensive, even more so when dealing with extensive geographic areas (Hanks et al., 2011; Ruiz-Gutierrez et

al., 2016). This requires not just data collection and subsequent storage, but also the training of a person capable of delivering the information with minimal to no bias or errors. Despite these substantial obstacles, ground-truth field data continues to be the most reliable and closest to the true condition of the system(Hooten et al., 2007; Dickinson et al., 2012). For this study, we were able to acquire a valuable source of ground-truth data from Crop Quest Inc., which included geolocated fields spanning many States in the central region of the United States. This dataset was developed during a five-year period, from 2013 to 2017, by conducting repeated trips to each farmer field to document maize crop phenology changes during the growing season.

To demonstrate contrasting case studies for maize crop production, two regions were selected for presenting geographical proximity but different productivity, SW and CK regions, during the 2017 growing season (Figure 1a). On average, each field was visited 5 times during the season, although this varies greatly within all field observation collected and summarized in this large database. The crop phenology stages present in the dataset followed the scale described on Figure 1, b, based on Ciampitti et al. (2011). The final database included geolocated fields for both regions, crop phenology measurements per field, and the date for data collection (time stamp).

2.2.2. Remote sensing data and weather variables

Google Earth Engine (GEE) was used to retrieve and analyze spectral remote sensing data, as well as for the weather information (Figure 1c). This cloud-based platform enabled the processing of large datasets and provided access to a wide library of pre-processed satellite images, climate, and land cover data, as well as built-in algorithms and vector data (Gorelick, 2017; Kumar and Mutanga, 2018).

2.2.2.1. Planet fusion product

For the 2017 season, our industry partner provided daily Planet Fusion (PF) harmonized product. This product combines public and commercial sensors (data from multi-constellation Planet Scope, Sentinel-2, and Landsat 8) to provide a daily, gap-filled, temporally consistent, radiometrically robust surface reflectance with 3-meter spatial resolution and four spectral bands: blue (0.45 – 0.51 μm), green (0.53 – 0.59 μm), red (0.64 – 0.67 μm) and near infrared (0.85 – 0.88 μm). Due to the high temporality, spatial resolution, and quality of data, this product can be an excellent resource for dealing with detailed crop phenology and its temporal variation. For the SW region, tiles 11E-172 N and 11E-173 N were utilized, as well as tile 21E-174 N for CK region, with each tile covering an area of 24 by 24 km. Both spectral and metadata from each imagery were uploaded to GEE via a Google Cloud Repository and organized into a collection for convenient access during feature extraction.

2.2.2.2. Sentinel-2

Sentinel-2 (S-2) is an optical satellite developed by the European Space Agency (ESA) that is composed of two complementary spacecraft that work together to collect data (S-2A and S-2B). This satellite was selected to create a comparison between the PF product and a publicly available and extensibility-enabled satellite with high temporal (5 to 10 days) and spatial resolution (10, 20 and 60 m, Drusch et al., 2012). Level-1C orthorectified reflectance product, from GEE archives was used, particularly the tile T14SKG covering SW Kansas. Even though the 2017 growing season had 12 images, only eight of them presented less than 20% cloud cover, and only five of them matched in time to those observed in the ground truth field data collection.

2.2.2.3. Spectral bands and Vegetation indices

Vegetation indices (VI) such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Green Chlorophyll Vegetation Index (GCVI), Chlorophyll Vegetation Index (CVI) and Normalized Green Index (NGI) were calculated on GEE based on the spectral bands from both products (PF and S-2). Extensive literature is associated to these indices and its capacity to monitor vegetation dynamics (Tucker 1977; 1979; Turner et al., 1999; Gitelson et al., 2003; Viña et al., 2011).

Briefly, the NDVI is determined because healthy plants often have a higher reflectance in the near infrared (NIR) than in the visible bands (red) (Tucker, 1979). While extensively utilized, it is widely documented that this index saturates at large biomass levels due mainly to its sensitivity to chlorophyll concentration. The EVI is an index that incorporates the blue band to mitigate the impact of certain atmospheric effects, more accurately representing changes in plant canopy, including high productivity regions (Liu and Huete, 1995; Huete et al., 2002). The GCVI has been proven to present a greater linear connection with the leaf area index of maize than other indices and, unlike NDVI, does not saturate at high biomass levels (Gitelson et al., 2003). Lastly, both CVI and NGI were suggested as adequate vegetation indices due to their responsiveness to changes in the chlorophyll content of the crop (Vincini and Frazzi, 2011; Sripada et al., 2006).

2.2.2.4. Weather variables

Weather variables were derived from the GEE archives based on gridded surface meteorological data (GRIDMET, Abatzoglou, 2013). This dataset integrates high-resolution spatial data from PRISM (2.5 arc minute, 4 km) with regular temporal resolution data from the North American Land Data Assimilation System (NLDAS) (Abatzoglou et

al., 2014). The main meteorological variables retrieved for the analysis were mean, minimum, and maximum temperatures, precipitation, and vapor pressure deficit (VPD), all on a daily temporal scale.

2.3. Data preparation

To extract circular fields linked to irrigated pivot operations, the Canny edge detector method was implemented on the GEE platform. Using this mask, in conjunction with the fields under study, we were able to construct a layer that included information about the geolocation and phenology of each irrigation field in the dataset. As a result, the mean values of the spectral bands (red, green, blue, NIR), VIs (NDVI, EVI, GCVI, CVI, NGI), and the weather parameters (precipitation, minimum temperature, maximum temperature, VPD) associated with each field for each given date were all gathered to construct the main database.

Even though PF is available daily, only images matching to the exact same day in the phenology collection (a total of 91 dates) were utilized (Figure 2). When the same method was applied for S-2 data, cloud masking and filtering images based on cloud cover were included. Images with less than 20% cloud cover remained in the dataset. Following this procedure, only five images (dates: June 15, June 30, July 5, July 25, and August 14) were found to match the crop collection (same day) (Figure 2). Due to this temporal limitation on availability of S-2 imagery, we decided to explore more images including three and ten days before and after the crop phenology data was collected. Therefore, three datasets were created for this study: i) containing the spectral band values, VI, and weather parameters for the exact phenology date, ii) spanning three days before and after

the crop phenology collection date (7 days total), and iii) ranging ten days before and after the crop phenology collection date (21 days total).

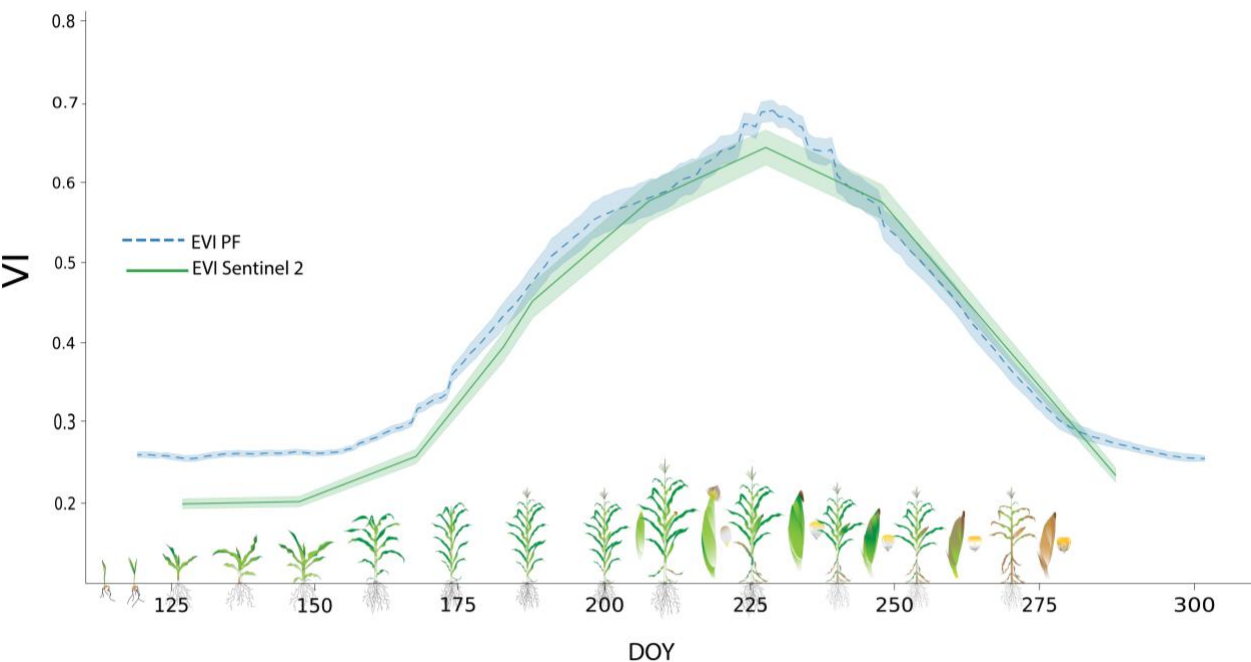


Figure 2. EVI temporal series for the 2017 maize growing season for SW fields using Sentinel-2 (solid green line) and PF (dash blue line). Shady area round both lines represents the standard deviation for the EVI values corresponding to each field under study (pink dots Figure a.2). Phenology progress is illustrated at the bottom of the image showing the progression of crop growth and the relation with EVI values.

2.4. Random Forest

The Random Forest (RF) classifier was implemented with the Scikit-Learn implementation (Pedregosa et al., 2011). This is a supervised classifier that learns from training samples, thus the quality of training instances is a key component in the classification process. The RF classifier, in addition, is an ensemble classifier, in which

multiple decision trees are utilized to improve generalizability and, therefore, resilience by using a randomly selected subset of training samples and variables (Dietterich, 2000). All these aspects contribute to the creation of accurate classifications, as well as the generation of high metrics and the processing of large amounts of data rapidly. It has been utilized in remote sensing for more than two decades, helping in the handling of high-dimensional data, unbalanced datasets, and multi-target problems, while staying simple and flexible using multiple hyperparameters to accommodate the scenario (Pal, 2005; Rodriguez-Galiano et al., 2012; Belgiu and Drăguț, 2016; Gomez et al., 2016; Pelletier et al., 2016; Tatsumi et al., 2015).

To produce a more robust model in this study, the hyperparameters linked with it were chosen using an exhaustive grid search checking for various values. The dataset for this test was created by combining the variables in their final configuration (further details presented in the section below). This resulted in the following hyperparameter values: bootstrap: True; maximum tree depth: 10; maximum number of features required to do the best split: 'auto'; minimum sample size per leaf: 2; minimum sample size required to execute the split: 2; number of estimators: 1000. To evaluate performance, the datasets (PF and S-2) were split into training, validation, and test sections for each experiment. 70% of the total was reserved for training and validation (70% for training, 30% for validation), and 30% for model assessment.

2.4.1. Performance metrics

When evaluating the performance of RF, the most commonly used metrics are overall accuracy, precision, recall, and F-score (Congalton, 1991). Most of the literature studying the performance of these metrics, centers in balance binary classifications, a not

commonly scenario find in remote sensing with real-world data. In this paper the ground-truth dataset presents more than 18 classes, with considerable unbalance frequency of members. According to Branco et al. (2016), f1-score, defined as the harmonic mean of precision and recall, results in a useful metric to deal with unbalanced dataset. In contrast, the use of overall accuracy, total of correctly identified elements divided by total, can be misleading since it will favor the most represented class; precision summarizes the fraction of samples assigned to the class that in fact are part of the class; and recall summarizes how well the class was predicted.

Even though all of these measures were calculated in this analysis, the majority of the findings are focused on the f1-score, precision, and recall for the reasons that were previously stated about the nature of the dataset.

2.4.2. Feature importance

In classification problems, feature selection is commonly used to reduce the dimensionality of the dataset (Gislason et al., 2004b; Rodriguez-Galeano et al., 2012) informing on the influence of each variable in the model. To narrow down the possible combinations of variables to address the best classifier, a feature importance analysis was conducted over the PF dataset, for both the SW and CK regions, composed by all the variables (VI, bands, weather, geolocation of the field and day of the year).

3. Results:

3.1. Best combination of variables

The analysis of the feature importance from the SW region dataset revealed that the most distinguishing variables, NIR band (B4) and EVI, as well as the minimum and

maximum temperatures (Tmin, Tmax), VPD, and DOY, were associated with the highest scores. There were fewer disparities in the variables in the dataset from the CK region, but the components stated above were still the most important ones. When all the variables were combined, we found that the best f1-scores were 0.94 for the SW region and 0.93 for the CK region. In addition, other combinations of these factors were examined (only spectral data; only weather parameters; EVI and weather; B4 and weather, etc.). The lowest f1-scores were derived when the model was only driven by weather variables, with 0.79 for SW and 0.76 for CK regions. Lastly, the final model consisted of B4, EVI, Tmin, Tmax, VPD, and DOY was selected to study the performance across crop phenological stages, in both geographical regions.

3.2. Model performance over the space using PF data

The final model including satellite, weather, and date of data collection was utilized to classify regions, yielding f1 values of 0.94 for the SW and 0.93 for the CK regions (Figure 3 a, b). The classification results for each crop phenological stage are presented as a confusion matrix, where all the components that were correctly identified are along the main diagonal, as well as the resulting precision and recall along the X and Y axes. When examining the SW area in detail, most of stages produced values between 1.0 and 0.75 (recall) and 1.0 to 0.80 (precision), whereas certain stages, such as R1, R3, and V15, showed poor metrics ranging from 0 to 0.66. In a similar manner, the findings for the CK region showed that most stages had recall values ranging from 1.0 to 0.87, but some (R2, V10, V11, V12, and V13) exhibited recall values ranging below 0.67. When the precision is considered, all stages ranged from 0.76 to 1.0.

3.3. Model performance using Sentinel-2 data

When the same final model was applied to S-2 data, the findings were inferior to those previously discussed in this paper. To begin, the dataset included just three crop phenological stages, matching satellite data with field collection. The categorization resulted in an overall high f1-score (0.86), but inconsistent metrics for each stage, for example for VT stage both recall and precision were zero (Figure 4, a). In addition, we tested the performance of this S-2 data for 3 days before and after the time for field data collection, overall f1-score was reduced to 0.74 but included more phenological stages. However, this f1-score is considerably lower when we compare this metric with the one obtained when using the PF product (Figure 4, b). Lastly, this test was repeated by expanding our search for satellite imagery data from 3 to 10 days before and after the field data collection time. Overall, the final f1-score was 0.58 but several crop phenology stages were not classified (null metrics) and of those stages presenting f1-scores never outperform the value obtained when using the PF data product (Figure 4c).

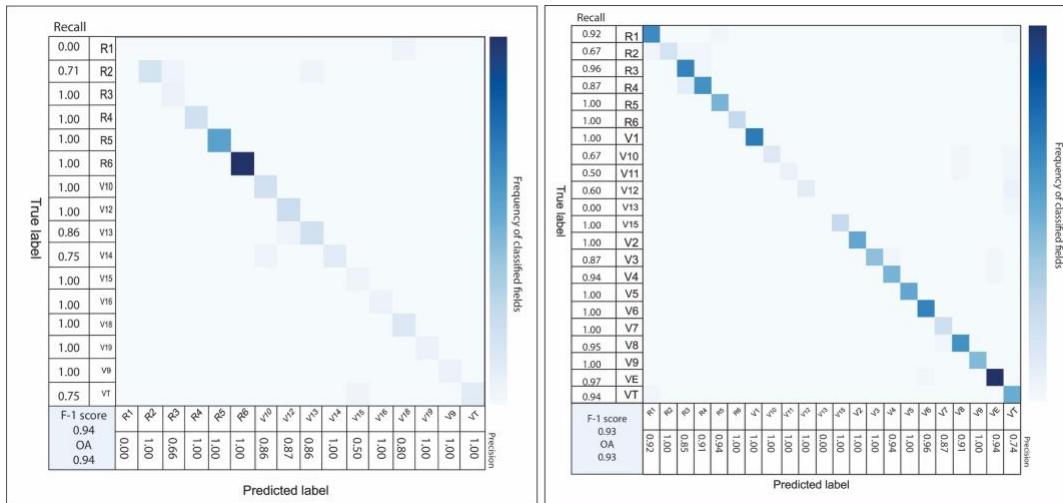


Figure 3: (a) Confusion matrices with classification results for the southwest region (SW) and for (b) central KS region (CK), both using PF product, including recall and precision for each crop phenology stage, and f1-score and overall accuracy (OA).

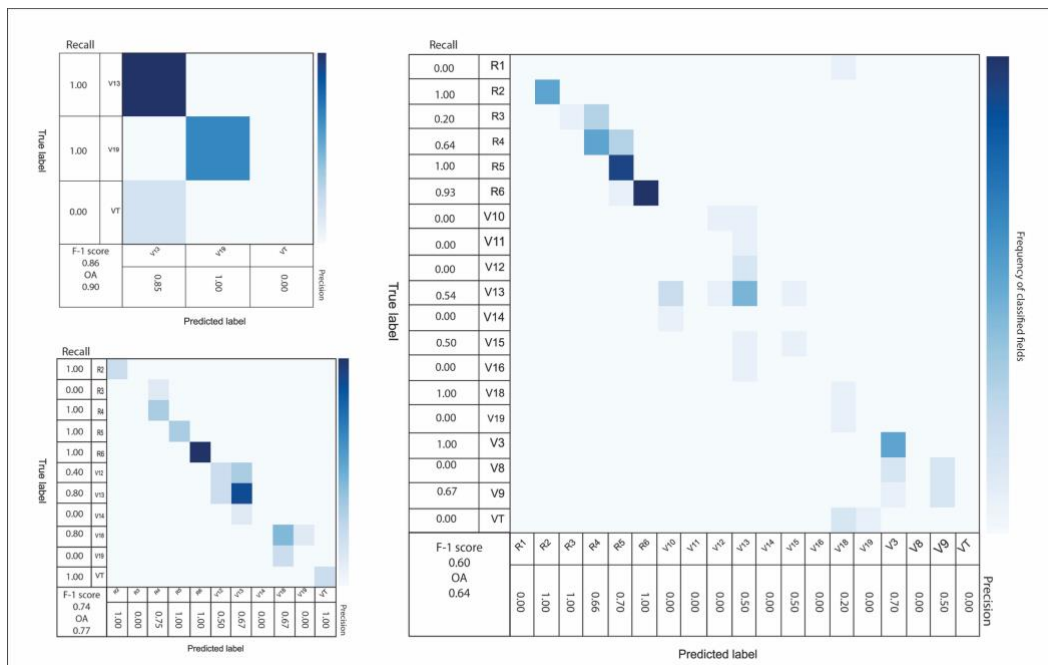


Figure 4: Confusion matrices with classification results for the southwest (SW) region, using Sentinel 2, (a) when matching image with the data of data collection, (b) 3 days before and after, (c) 10 days before and after. All matrices include recall and precision for each crop phenology stage, and f1-score and overall accuracy (OA).

4. Discussion

This study portrays the benefits of accessing high-resolution in both spatial and temporal scales to achieve high-quality crop phenology classification to develop a simple model integrating key (easy-to-access) weather variables. In addition, it is worth noticing the importance of collecting and having available relevant ground truthing data (with adequate labels) to train more complex models. The findings presented in this study are in accordance with those observed when tracking crop phenology (White et al., 1997; Liang et al., 2011; Peng et al., 2018; Bandaru et al., 2020; Nieto et al., 2021), most notably the importance of enhanced temporal resolution (Gao et al., 2017; Gao et al., 2020; Gao and Zhang, 2021).

One of the first lessons learned from this study relates to the variables selected in the model, reflecting the complexity for describing biological processes, integrating satellite imagery data (NIR band, EVI) with weather (minimum and maximum temperature, VPD), and with proper field labels (DOY). Previous studies highlighted the importance of VIs when monitoring changes in crop phenology, EVI has been found to outperform a broad variety of VIs in maize fields (Tatsumi et al., 2015; Zhong et al., 2016a; Zhong et al., 2016b, Cai et al., 2019). In addition, promising results were achieved by Cai et al. (2018) regarding NIR band performance. On the other side, inclusion of relevant weather variable is key to describe changes in crop phenology or development, particularly demonstrated in modeling scenarios (Bai et al., 2010; Joshi et al., 2021), with air temperature among the most relevant (Azzari et al., 2017). In addition, VPD has been strongly associated with water stress and the consequent impact in crop development (Zhang et al., 2017; Hsiao et al., 2019).

A second lesson was linked to the relevancy of daily satellite data from the same day of field data collection to classify crop phenology more accurately. One disadvantage of our study is that just a few images for S-2 were available during the 2017 growing season. This scenario reduced the likelihood of finding an image acquired on the same day as the timing for field data collection, and the consequent impact on the ability to classify more crop phenology stages. To emphasize the importance of temporal resolution, we introduce here a new sensitivity test using PF data only, evaluating a departure for the timing of satellite data in 1-, 2-, and 3-days before and after the field data collection. The ability of the classifier to achieve high metrics was substantially reduced by 12% with the 2-day and by roughly more than 40% with the 3-day departure, especially during crop developmental phases around flowering time (e.g., V12, V13, V19). These results emphasized the importance of acquiring “timely” daily spectral data, particularly during periods of rapid crop growth.

Other caveats from the analysis are related to problems associated with multiclass classification and imbalance dataset. Most of the methods to treat imbalance datasets rely on binary problems (Hoens et al., 2012) and solutions are not extensively proven. Obtaining access to daily spectral data as well as increasing the number of ground truth samples are key factors to build more reliable predictions. In this study, this was partially overcome using the PF product, demonstrating the impact that daily availability has in the classification. More efforts are being conducted towards the development and availability of fusion products, such as the Harmonized Landsat-Sentinel (HLS), although it was not implemented on this analysis since, at the time, it was suggested to not be yet used for scientific purposes (LP DAAC – HL30, n.d.).

Future research should focus on testing the developed model in different scale of agriculture (e.g., small holder farmers) and under changes in management approaches and expanding the temporal evaluation via adding more years for sampling seasonal weather environments relevant for current climate conditions (e.g., heat and drought conditions). Lastly, not only more quantity of ground truth data is required but the quality (labeled-field data) is equally relevant for future progress on developing satellite, weather, and field-based mechanistic and predictive models relevant for the different scale of farming operations.

5. Conclusions

The proposed model combining NIR spectral band with EVI, weather (maximum and minimum temperatures, VPD), and day of the year permitted to obtain a high-quality crop phenology classification for maize crop under two contrasting production regions. Furthermore, the effect of having a daily satellite data was highlighted when comparing the outcomes of limited temporal availability with the metrics produced by a daily product, particularly when the same image is used to make inferences about the future or the past.

This study was successful at providing evidence using two case study regions that the developed model can be most likely transferred to other relevant regions. Improving both classification and reporting of crop phenology has a large impact on every step of the food production chain, enabling farmers, stakeholders, and policymakers to take a proactive, rather than reactive, approach at both the regional and field level.

401 **Acknowledgements:**

402 This work was supported by the Kansas Corn Commission. The dataset with the field
403 ground-truth data were provided by Crop Quest inc. This is contribution no. 22-xxx-J from
404 Kansas Agricultural Experiment Station.

405

406 **Author contributions:**

407 Conceptualization, L.N., I.A.C.; methodology, L.N.; formal analysis, L.N.; resources
408 xxxxx; writing-original draft preparation, L.N., I.A.C.; writing-review and editing, L.N.,
409 I.A.C., P.V.V.P., xxxxxxxx; funding acquisition, I.A.C.

410

411 **Funding:** this work was supported by the Kansas Corn Commission, and Kansas State
412 University provided found to complete this work, supporting Mrs. Nieto's Ph.D. stipend
413 and Dr. Ciampitti research program.

414

415

416 **Reference:**

- 417 Liang, L., Schwartz, M. D., & Fei, S. (2011). Validating satellite phenology through
418 intensive ground observation and landscape scaling in a mixed seasonal forest.
419 Remote Sens.Environ. 115(1), 143-157.
- 420 Ruml, M., & Vulić, T. (2005). Importance of phenological observations and predictions in
421 agriculture. J.Agric.Sci.(Belgrade), 50(2), 217-225.
- 422 Gao, F., & Zhang, X. (2021). Mapping crop phenology in near real-time using satellite
423 remote sensing: Challenges and opportunities. J.Remote Sens. 2021.
- 424 Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., ... & Prueger,
425 J. H. (2017). Toward mapping crop progress at field scales through fusion of
426 Landsat and MODIS imagery. Remote Sens.Environ. 188, 9-25.
- 427 White, M. A., Thornton, P. E., & Running, S. W. (1997). A continental phenology model
428 for monitoring vegetation responses to interannual climatic variability. Global
429 Biogeochemical Cycles, 11(2), 217-234.
- 430 Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., ... & Huete,
431 A. (2003). Monitoring vegetation phenology using MODIS. Remote Sens.Environ.,
432 84(3), 471-475.
- 433 Vina, A., Gitelson, A. A., Rundquist, D. C., Keydan, G. P., Leavitt, B., & Schepers, J.
434 (2004). Monitoring maize (Zea mays L.) phenology with remote sensing.

435 Houborg, R., Fisher, J. B., & Skidmore, A. K. (2015). Advances in remote sensing of
436 vegetation function and traits.

437 Rast, M., & Painter, T. H. (2019). Earth observation imaging spectroscopy for terrestrial
438 systems: An overview of its history, techniques, and applications of its missions.
439 *Surveys in Geophysics*, 40(3), 303-331.

440 Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C.
441 E., ... & Zhu, Z. (2019). Current status of Landsat program, science, and
442 applications. *Remote Sens. Environ.* 225, 127-147.

443 Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., ... & Justice,
444 C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set.
445 *Remote Sens. Environ.* 219, 145-161.

446 Liao, C., Wang, J., Dong, T., Shang, J., Liu, J., & Song, Y. (2019). Using spatio-temporal
447 fusion of Landsat-8 and MODIS data to derive phenology, biomass and yield
448 estimates for corn and soybean. *Science Total Environ.*, 650, 1707-1721.

449 Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land
450 cover classification. *Pattern Recognition Letters*, 27(4), 294-300.

451 Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

452 Pal, M. (2005). Random forest classifier for remote sensing classification. *International*
453 *J. Remote Sens.* 26(1), 217-222.

454 Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.
455 P. (2012). An assessment of the effectiveness of a random forest classifier for
456 land-cover classification. *ISPRS J. Photogrammetry Remote Sens.* 67, 93-104.

457 Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications
458 and future directions. *ISPRS J. Photogrammetry Remote Sen.* 114, 24-31.

459 Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need
460 hundreds of classifiers to solve real world classification problems? *The J. Machine*
461 *Learning Res.* 15(1), 3133-3181.

462 Lin, X., Harrington, J., Ciampitti, I., Gowda, P., Brown, D., & Kisekka, I. (2017). Kansas
463 trends and changes in temperature, precipitation, drought, and frost-free days from
464 the 1890s to 2015. *J. Contemporary Water Res. Education* 162(1), 18-30.

465 USDA/NASS quickstats ad-hoc query tool. (n.d.). Retrieved July 15, 2021, from
466 <https://quickstats.nass.usda.gov/>

467 Hanks, E. M., Hooten, M. B., & Baker, F. A. (2011). Reconciling multiple data sources to
468 improve accuracy of large-scale prediction of forest disease incidence. *Ecological*
469 *Applications* 21(4), 1173-1188.

470 Ruiz-Gutierrez, V., Hooten, M. B., & Grant, E. H. C. (2016). Uncertainty in biological
471 monitoring: a framework for data collection and analysis to account for multiple
472 sources of sampling bias. *Methods in Ecology and Evolution*, 7(8), 900-909.

473 Hooten, M. B., Wikle, C. K., Dorazio, R. M., & Royle, J. A. (2007). Hierarchical
474 spatiotemporal matrix models for characterizing invasions. *Biometrics*, 63(2), 558-
475 567.

476 Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., ... & Purcell, K.
477 (2012). The current state of citizen science as a tool for ecological research and
478 public engagement. *Front.Ecology Environ.*, 10(6), 291-297.

479 Ciampitti, I. A., Elmore, R. W., & Lauer, J. (2011). Corn growth and development. *Dent*,
480 5(75).

481 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017).
482 Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote*
483 *Sens.of Environ.* 202, 18-27.

484 Kumar, L., & Mutanga, O. (2018). Google Earth Engine applications since inception:
485 Usage, trends, and potential. *Remote Sensing* 10(10), 1509.

486 Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... & Bargellini,
487 P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational
488 services. *Remote Sens.Environ.* 120, 25-36.

489 Tucker, C. J. (1977). Asymptotic nature of grass canopy spectral reflectance. *Applied*
490 *Optics* 16(5), 1151-1156.

491 Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring
492 vegetation. *Remote sensing of Environment*, 8(2), 127-150.

493 Turner, D. P., Cohen, W. B., Kennedy, R. E., Fassnacht, K. S., & Briggs, J. M. (1999).
494 Relationships between leaf area index and Landsat TM spectral vegetation indices
495 across three temperate zone sites. *Remote Sens. Environ.* 70(1), 52-68.

496 Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B.
497 (2003). Remote estimation of leaf area index and green leaf biomass in maize
498 canopies. *Geophysical Res. Letters* 30(5).

499 Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of
500 biophysical characteristics of vegetation. *J. Plant Physiol.* 161(2), 165-173.

501 Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of
502 different vegetation indices for the remote assessment of green leaf area index of
503 crops. *Remote Sens. Environ.* 115(12), 3468-3478.

504 Liu, H. Q., & Huete, A. (1995). A feedback-based modification of the NDVI to minimize
505 canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote*
506 *Sens.* 33(2), 457-465.

507 Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002).
508 Overview of the radiometric and biophysical performance of the MODIS vegetation
509 indices. *Remote Sens. Environ.*, 83(1-2), 195-213.

510 Vincini, M., & Frazzi, E. (2011). Comparing narrow and broad-band vegetation indices to
511 estimate leaf chlorophyll content in planophile crop canopies. *Precision Agric.*
512 12(3), 334-344.

513 Sripada, R. P., Heiniger, R. W., White, J. G., & Meijer, A. D. (2006). Aerial color infrared
514 photography for determining early in-season nitrogen requirements in corn.
515 Agron.J. 98(4), 968-977.

516 Abatzoglou, J.T., 2013. Development of gridded surface meteorological data for
517 ecological applications and modelling. Int. J. Climatol. 33, 121–131.
518 <https://doi.org/10.1002/joc.3413>

519 Abatzoglou, J. T., Rupp, D. E., & Mote, P. W. (2014). Seasonal climate variability and
520 change in the Pacific Northwest of the United States. J.Climatol. 27(5), 2125-2142.

521 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... &
522 Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. J.Machine
523 Learning Res. 12, 2825-2830.

524 Dietterich, T. G. (2000). Ensemble methods in machine learning. In International
525 workshop on multiple classifier systems (pp. 1-15). Springer, Berlin, Heidelberg.

526 Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series
527 data for land cover classification: A review. ISPRS J.Photogrammetry Remote
528 Sens. 116, 55-72.

529 Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the
530 robustness of Random Forests to map land cover with high resolution satellite
531 image time series over large areas. Remote Sens.Environ. 187, 156-168.

532 Tatsumi, K., Yamashiki, Y., Torres, M. A. C., & Taïpe, C. L. R. (2015). Crop classification
533 of upland fields using Random Forest of time-series Landsat 7 ETM+ data.
534 Computers Electronics Agric., 115, 171-179.

535 Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely
536 sensed data. Remote Sens. Environ. 37(1), 35-46.

537 Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on
538 imbalanced domains. ACM Computing Surveys (CSUR), 49(2), 1-50.

539 Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2004b). Random forest
540 classification of multisource remote sensing and geographic data. In IGARSS
541 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium (Vol.
542 2, pp. 1049-1052). IEEE.

543 Nieto, L., Schwalbert, R., Prasad, P. V., Olson, B. J., & Ciampitti, I. A. (2021). An
544 integrated approach of field, weather, and satellite data for monitoring maize
545 phenology. Scientific Rep. 11(1), 1-10.

546 Gao, F., Anderson, M., Daughtry, C., Karnieli, A., Hively, D., & Kustas, W. (2020). A
547 within-season approach for detecting early growth stages in corn and soybean
548 using high temporal and spatial resolution imagery. Remote Sens. Environ., 242,
549 111752.

550 Peng, B., Guan, K., Pan, M., & Li, Y. (2018). Benefits of seasonal climate prediction and
 551 satellite data for forecasting US maize yield. *Geophysical Research Letters*,
 552 45(18), 9662-9671.

553 Bandaru, V., Yaramasu, R., Koutilya, P. N. V. R., He, J., Fernando, S., Sahajpal, R., ... &
 554 Justice, C. (2020). PhenoCrop: An integrated satellite-based framework to
 555 estimate physiological growth stages of corn and soybeans. *Internat.J.Applied*
 556 *Earth Observ.Geoinformation* 92, 102188.

557 Zhong, L., Hu, L., Yu, L., Gong, P., & Biging, G. S. (2016a). Automated mapping of
 558 soybean and corn using phenology. *ISPRS J.Photogrammetry Remote Sens.*,
 559 119, 151-164.

560 Zhong, L., Yu, L., Li, X., Hu, L., & Gong, P. (2016b). Rapid corn and soybean mapping in
 561 US Corn Belt and neighboring areas. *Scientific Rep.* 6(1), 1-14.

562 Cai, Y., Guan, K., Lobell, D., Potgieter, A. B., Wang, S., Peng, J., ... & Peng, B. (2019).
 563 Integrating satellite and climate data to predict wheat yield in Australia using
 564 machine learning approaches. *Agric.Forest Meteorol.*, 274, 144-159.

565 Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2018). A high-
 566 performance and in-season classification system of field-level crop types using
 567 time-series Landsat data and a machine learning approach. *Remote*
 568 *Sens.Environ.* 210, 35-47.

569 Bai, J., Chen, X., Dobermann, A., Yang, H., Cassman, K. G., & Zhang, F. (2010).
570 Evaluation of NASA satellite-and model-derived weather data for simulation of
571 maize yield potential in China. *Agron.J.*, 102(1), 9-16.

572 Joshi, V. R., Kazula, M. J., Coulter, J. A., Naeve, S. L., & y Garcia, A. G. (2021). In-season
573 weather data provide reliable yield estimates of maize and soybean in the US
574 central Corn Belt. *Internat.J.Biometeorology* 65(4), 489-502.

575 Azzari, G., Jain, M., & Lobell, D. B. (2017). Towards fine resolution global maps of crop
576 yields: Testing multiple methods and satellites in three countries. *Remote*
577 *Sens.Environ.* 202, 129-141.

578 Zhang, S., Tao, F., & Zhang, Z. (2017). Spatial and temporal changes in vapor pressure
579 deficit and their impacts on crop yields in China during 1980–2008.
580 *J.Meteorological Res.*, 31(4), 800-808.

581 Hsiao, J., Swann, A. L., & Kim, S. H. (2019). Maize yield under a changing climate: The
582 hidden role of vapor pressure deficit. *Agric.Forest Meteorol.*, 279, 107692.

583 Hoens, T. R., Qian, Q., Chawla, N. V., & Zhou, Z. H. (2012). Building decision trees for
584 the multi-class imbalance problem. In *Pacific-Asia Conference on Knowledge*
585 *Discovery and Data Mining* (pp. 122-134). Springer, Berlin, Heidelberg.

586 LP DAAC - HLSL30. (n.d.). LP DAAC - HLSL30. Retrieved August 5, 2021, from
587 <https://lpdaac.usgs.gov/products/hls130v015/>.

588

589

590

591