

Article

High-Resolution Soybean Yield Mapping Across the US Midwest Using Subfield Harvester Data

Walter T. Dado ^{1,*}, Jillian M. Deines ¹, Rinkal Patel ², Sang-Zi Liang ³ and David B. Lobell ¹

¹ Department of Earth System Science and Center on Food Security and the Environment, Stanford University, Stanford, CA 94305, USA; jdeines@stanford.edu (J.M.D.); dlobell@stanford.edu (D.B.L.)

² Granular, A Corteva Agriscience™ Company, San Francisco, CA 94103, USA; rinkalpatel@granular.ag

³ Corteva Agriscience™, Wilmington, DE 19805, USA; sangzi.liang@corteva.com

* Correspondence: wdado@alumni.stanford.edu or tekedado@icloud.com; Tel.: +1-310-889-4220

Received: 12 September 2020; Accepted: 19 October 2020; Published: 22 October 2020



Abstract: Cloud computing and freely available, high-resolution satellite data have enabled recent progress in crop yield mapping at fine scales. However, extensive validation data at a matching resolution remain uncommon or infeasible due to data availability. This has limited the ability to evaluate different yield estimation models and improve understanding of key features useful for yield estimation in both data-rich and data-poor contexts. Here, we assess machine learning models' capacity for soybean yield prediction using a unique ground-truth dataset of high-resolution (5 m) yield maps generated from combine harvester yield monitor data for over a million field-year observations across the Midwestern United States from 2008 to 2018. First, we compare random forest (RF) implementations, testing a range of feature engineering approaches using Sentinel-2 and Landsat spectral data for 20- and 30-m scale yield prediction. We find that Sentinel-2-based models can explain up to 45% of out-of-sample yield variability from 2017 to 2018 ($r^2 = 0.45$), while Landsat models explain up to 43% across the longer 2008–2018 period. Using discrete Fourier transforms, or harmonic regressions, to capture soybean phenology improved the Landsat-based model considerably. Second, we compare RF models trained using this ground-truth data to models trained on available county-level statistics. We find that county-level models rely more heavily on just a few predictors, namely August weather covariates (vapor pressure deficit, rainfall, temperature) and July and August near-infrared observations. As a result, county-scale models perform relatively poorly on field-scale validation ($r^2 = 0.32$), especially for high-yielding fields, but perform similarly to field-scale models when evaluated at the county scale ($r^2 = 0.82$). Finally, we test whether our findings on variable importance can inform a simple, generalizable framework for regions or time periods beyond ground data availability. To do so, we test improvements to a Scalable Crop Yield Mapper (SCYM) approach that uses crop simulations to train statistical models for yield estimation. Based on findings from our RF models, we employ harmonic regressions to estimate peak vegetation index (VI) and a VI observation 30 days later, with August rainfall as the sole weather covariate in our new SCYM model. Modifications improved SCYM's explained variance ($r^2 = 0.27$ at the 30 m scale) and provide a new, parsimonious model.

Keywords: crop yields; yield mapping; US Corn Belt; Landsat; Sentinel; agricultural monitoring; machine learning

1. Introduction

Agricultural yield data provide insights to heterogeneity across space and time, which can help identify yield gaps, inform farm management strategies, and guide sustainable intensification [1,2]. More specifically, modern precision agriculture can use high-resolution yield maps to vary management

at a subfield scale, with potential for closing yield gaps while improving input efficiency and environmental outcomes [3,4]. While some studies use satellite data to provide accurate, fine-scale yield estimation for limited spatial extents (e.g., [5,6]), many satellite-based yield mapping efforts result in county- or state-level products, limiting the usefulness for farm-level management (e.g., [7]).

Furthermore, studies that do produce higher resolution yield maps rarely validate their predictions with extensive, accurate ground truth data. In general, such widespread ground truth data are private, costly, or otherwise infeasible to obtain. As a result, the accuracy of field- or subfield-scale predictions either are not explicitly tested [8,9], tested only in a small contiguous region [5], or tested over a limited sample size [6,10]. However, farmers increasingly utilize combine harvesters with yield monitors that record yields at very fine scales (i.e., 1–5 m), particularly in large commercial systems. These data can inform management directly and have also driven research on yield response to various physical and soil characteristics [11,12]. Incorporating these data into satellite yield estimation studies provides opportunities for extensive, high-resolution validation, comparison between modeling approaches, and improved understanding of key satellite features.

One common approach to yield estimation is training machine learning models on yield data and satellite imagery. A significant body of work highlights the strength of random forests for crop type mapping [13–17] and yield estimation using remotely sensed data [5,6,18,19]. Random forests can capture highly nonlinear relationships by repeatedly splitting the parameter space yet remain robust to overfitting by taking an averaged prediction from many individual trees, each fit with a random subset of predictors and bootstrapped training data [20]. Yield estimation efforts across agricultural systems have employed random forests, including maize in the US Midwest and Middle East [6,19,21], wheat in the UK, China, India, and Australia [5,18,22,23], and soybean in Brazil [24]. Compared to other machine learning approaches, random forests have achieved more robust results in this yield estimation context with high-dimensional, often collinear inputs [21–23,25].

Building these empirically based yield relationships requires extensive, high quality ground data and may not generalize well to alternate settings. Empirically trained models are useful for regions and time periods with available training data but are often not feasible where reliable data are scarce. Crop simulations present an alternative strategy to overcome this ground-data obstacle. These simulations capture key relationships between climate, management, and crop cultivars; when parameterized for a target region, they can be used to generalize the relationship between crop leaf area, weather, and yield. For example, the satellite-based Scalable Crop Yield Mapper (SCYM) approach trains a parsimonious (and thus scalable) model on simulated crop yield data, then applies that statistical model to observed satellite and gridded weather data [26]. SCYM has been implemented across agricultural systems in North America, Africa, and India [9,17,27,28].

Among global agricultural staples, soybean is a crop for which high-resolution yield mapping has not been fully explored. Several recent soybean yield prediction efforts at the scale of aggregated political units (e.g., counties in the United States) have applied deep learning frameworks to multi-spectral imagery for soybean yield mapping [7,29], while others compared a series of machine learning models and feature engineering approaches [8,24]. At the pixel scale, Lobell et al. (2015) considered regional soybean yield estimation at Landsat resolution using multiple linear regression models trained on simulations from crop models rather than ground data [26]. Initial efforts for soybeans in the US Midwest explained about one-third (32%) of soybean yield variability, leaving substantial room for improvement [26].

Of the many possible remote sensing approaches to yield estimation, we focus here on two widely used and publicly available sources of imagery (Landsat and Sentinel-2) and two relatively common approaches: random forests and simulation-based models. Landsat has captured 30-m-resolution multispectral data for more than three decades and is widely used, especially for studies that aim to understand historical changes [30,31]. Sentinel-2 is a newer set of sensors that offers higher spatial, temporal, or spectral resolution than Landsat since 2015 [32]. In particular, Sentinel-2's Multi-Spectral Instrument (MSI) has shown promise for improved estimation of leaf area index (LAI) in crops [33],

a key predictor of crop yield. Sentinel-2 MSI's inclusion of bands in the “red edge” region, at 705, 740 and 783 nm, enables formulation of robust vegetation indices (VIs) that linearly relate to LAI even at high canopy density for which common VIs, like NDVI, saturate [34–36]. Past work has suggested the importance of these red-edge bands for remote-sensing of soybean crops, which have dense canopies that produce high LAI values [35,37].

Here, we utilize a unique harvester yield monitor dataset of over a million field-year soybean yield maps between 2008 and 2018 in the central United States to investigate the capacity of machine learning models to explain subfield yield heterogeneity, using both Sentinel-2 and Landsat satellite imagery (Table 1). Although Landsat provides a much longer record with which to evaluate models, we also include Sentinel-2 in order to assess the potential contribution of Sentinel-2's spectral precision, more frequent return time, and the availability of new red-edge VIs. First, we employ a flexible machine learning algorithm with and without pre-formulated vegetation indices to establish a baseline accuracy for this task and to infer the most relevant predictors. Second, we assess harmonic regressions' ability to capture soybean phenology and test a series of harmonic-based metrics. Third, we build models trained with publicly available county-level data and compare to models trained on the harvester dataset, examining the differences in performance and learned parameters. Finally, we draw on these results to update the SCYM simulation-based approach for soybean, leveraging our extensive ground-truth data to quantitatively compare multiple alternative implementations and produce an improved annual time series of high-resolution yield maps from 2008 to 2018.

Table 1. Model definition and overview. The four main modeling approaches tested in the paper, presented in terms of their input satellite data, their training and testing response variables (pixel, county, or simulated yields) and the motivating research question for their exploration. “Harvester” in the first column refers to combine harvesters, from which our pixel yield data comes (Section 2.2). SCYM = Scalable Crop Yield Mapper (Section 2.5.4).

Model	Satellite Data	Training Response Variable	Testing Response Variable	Machine Learning Algorithm	Research Question
Landsat harvester trained	30 m Landsat	30 m harvester yields	30 m harvester yields AND county yields	Random Forest	How well can a machine learning model, trained using pixel-scale harvester yields, perform both at the pixel- and county-scale? Does the additional spectral precision of Sentinel-2 help compared to the same years/test sites using Landsat? Do red-edge vegetation indices add signal?
Sentinel-2 harvester trained	20 m Sentinel-2	20 m harvester yields	20 m harvester yields	Random Forest	How does a model trained with aggregated, freely available data compare to a model trained with pixel-scale data in performance at both pixel and county scales?
County-trained model	Landsat, sampled and aggregated to county scale	County Yields	County yields AND 30 m harvester yields	Random Forest	How does a model trained with simulated data perform on pixel-scale test data? Do insights from the 30 m harvest-trained model improve SCYM methodology?
Simulations-based SCYM Model	30 m Landsat	Simulated crop yields	30 m harvester yields AND county yields	Multiple Linear Regression	

2. Methods

2.1. Study Area

In the United States, farmers planted over 90 million acres of soybean in 2017; in 2018, soybean planted acreage outpaced corn for the first time in decades [38]. Soybean is the United States' most lucrative agricultural export [38], driven by increasing demand for animal feeds associated with rising meat consumption around the world. The vast majority, nearly 80%, of United States' soybean acreage lies in the Midwest, making it an ideal study site for soybean yield mapping efforts [39]. According to USDA Agricultural Census data, three-quarters or more of US soybean production comes from large farms that plant over 250 hectares [40]. Over the 11-year study period, 2008–2018, average yields were 46.5 bushels/acre or 3.1 metric tons per hectare [41]. Here, we focus on a 9-state region in the Midwest (Figure 1). We gathered 4-km-resolution Gridmet weather data from June through September [42]. Over this four-month period, total precipitation averaged 419.25 mm per year. Average minimum daily temperatures were 16 °C in June, 17.4 °C in July, 15.7 °C in August, and 11.9 °C in September; average maximum daily temperatures were 28.8 °C in June, 29.7 °C in July, 27.8 °C in August, and 25.4 °C in September.

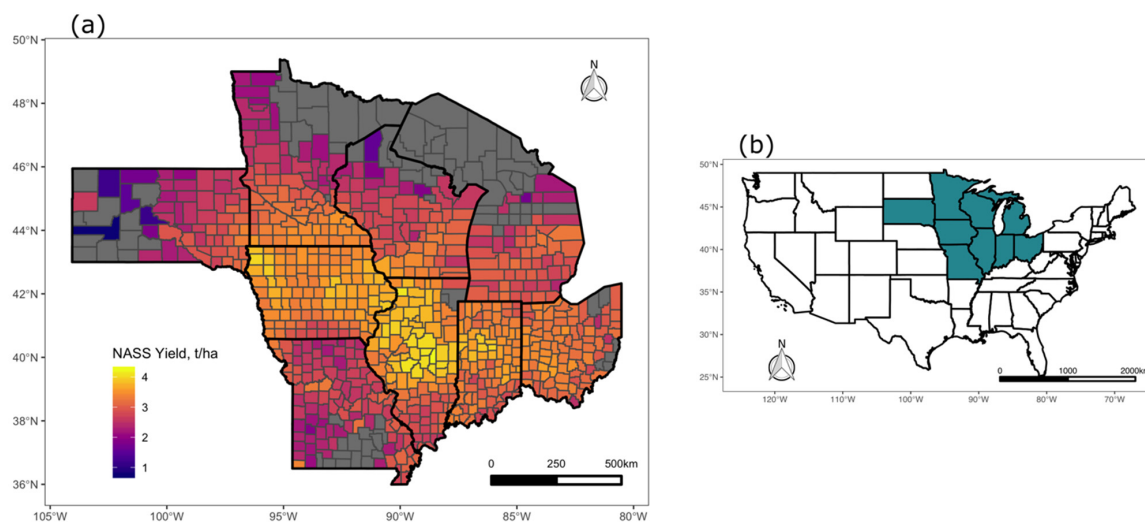


Figure 1. Study Area. (a) Average county yields across the period 2008–2018 from the National Agricultural Statistics Service (NASS) [41]. (b) Study area states (blue) in context of the continental United States.

2.2. Yield Data

We employed two types of yield data: harvester yield monitor data and county-level data. The yield monitor data came via collaboration with Corteva Agriscience. Combine harvesters with yield monitors collect point yield data as the farmer drives throughout the field by integrating measurements of combine width and speed with grain weight and moisture levels [43]. We processed raw point measurements into standardized 5-m yield maps for each field by filtering to remove field edges and faulty values, adjusting yields to a standard grain moisture content, rasterizing to a 5-m grid, and smoothing using a 15-m moving window average. This resulted in over a million field yield maps spanning the study area and time period; some fields had yield maps for multiple years, while other fields had yield data for only one year. Yield maps were unevenly distributed in space and time. To create a balanced dataset, we discretized the region into 50-km² grid cells and randomly sampled up to 150 unique fields in each of the resulting 318 grid cells, each year. If a grid cell contained less than 150 field maps in a year, we used all available fields. This resulted in 402,840 observations over the 11-year period, ranging from 32,343 to 39,029 fields per year (Table 1). From each field-year observation, we randomly sampled one point and extracted the yield value at both 20-m and 30-m

resolution for Sentinel-2 and Landsat analyses, respectively. Finally, outlier measurements less than 0.1 or above 7 metric tons per hectare (t/ha) were removed, leaving approximately 380,000 fields. Hereafter, we refer to these datasets as 20 m and 30 m harvester yield data, and more generally as “pixel scale” data.

We obtained county-level yield data from the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS), which reports average yield values for each county and year with sufficient data [41]. County means from the full harvester yield monitor dataset agreed reasonably well with NASS county yields for 2008–2018 ($r^2 = 0.72$). The mean difference was 0.5 tons per hectare, indicating that the yield monitor data came from fields that systematically achieve above average yields.

2.3. Satellite Data

We collected monthly time series data for Landsat and biweekly time series data for Sentinel-2 due to each sensor’s typical return interval. Using Google Earth Engine [44], we retrieved all available Landsat Tier 1 Surface Reflectance observations overlying each point-year in our sample, drawing from Landsat 5 Thematic Mapper, 7 Enhanced Thematic Mapper Plus, and 8 Optical Land Imager. The nominal resolution is 30 m, and the 3 sensors produce compatible observations for our purposes [45]. To produce a cleaned dataset with a consistent number of observations, we then filtered for clear pixels using the cloud mask from the “pixel_qa” band and extracted one monthly observation per point for May–September (Figure 2), based on the maximum green chlorophyll vegetation index value, or GCVI ($\text{NIR}/\text{green}-1$) [46]. Points which lacked a clear observation in any month between May and September were discarded, resulting in 186,160 points (Table 2).

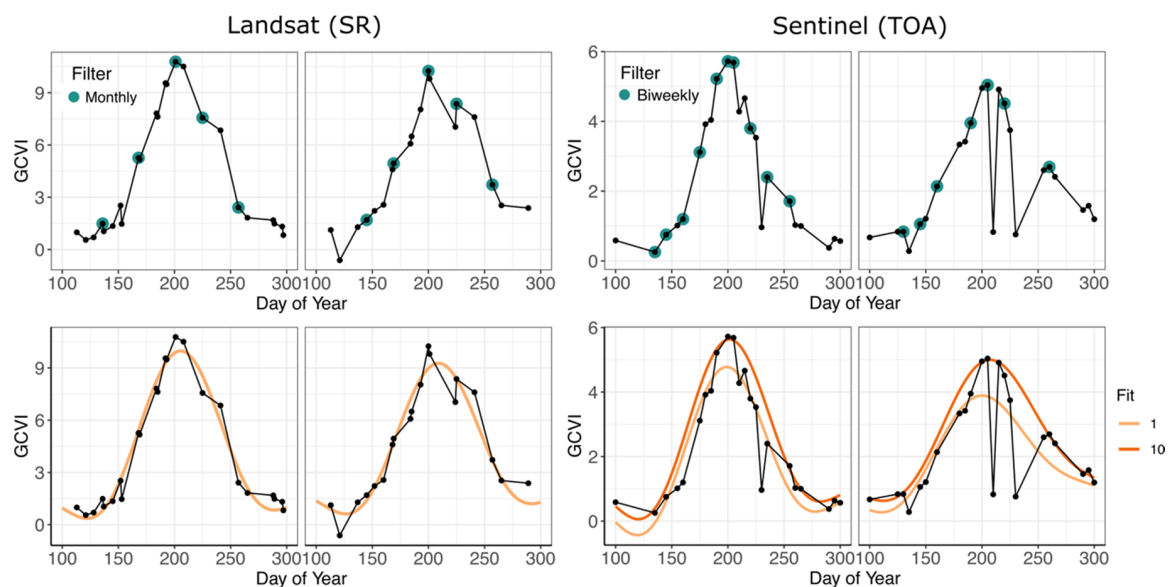


Figure 2. Satellite Processing Methods. Illustration of data processing for two pixels in 2018. Top Row: to create a standardized data set, we selected the satellite image with the highest vegetation index (VI) in each monthly (Landsat) or biweekly (Sentinel-2) period during the growing season. Bottom Row: harmonic fits applied to the cloud-masked time series. The black points show all observations before filtering. We used the first fit for Landsat and 10th recursive fit for Sentinel-2, due to residual noisiness from its less reliable cloud mask (see Methods). SR = surface reflectance; TOA = top of atmosphere; GCVI = Green Chlorophyll Vegetation Index.

Table 2. Distribution of sampled points. The sampling design created a balanced sample across space and time. The “Point Sample Distribution” column provides the number of points by year. The “Points after Landsat Filter” column provides the final number available for the Landsat analysis after filtering on the condition that each point must have one clear observation each month from May to September. The “Points after Sentinel Filter” column provides the final number available for Sentinel-2 analysis after filtering on the condition that each point must have one clear observation in each biweekly period from May to September.

Year	Point Sample Distribution	Points after Landsat Filter	Points after Sentinel Filter
2008	32,343	15,745	
2009	28,385	13,653	
2010	37,163	17,946	
2011	29,761	18,086	
2012	35,772	1792	
2013	34,884	16,057	
2014	37,823	18,442	
2015	32,940	15,766	
2016	39,029	25,134	
2017	38,647	24,412	15,436
2018	35,108	19,127	24,142

Sentinel-2 MSI data were obtained at the 20-m scale. We analyzed Sentinel-2 over just the period 2017–2018 as they are the first two years in which both sensors were operating. We resampled bands for which the native resolution is 10 m to the 20-m scale in order to match the resolution of the three red-edge bands. Because surface reflectance data or processing routines were unavailable for data prior to December 2018 on Google Earth Engine at the time of this study, we used Level 1C (top of atmosphere) data. Although methods exist for manually performing atmospheric correction, deploying them at the scale of our study presents a significant obstacle. Past work has shown that vegetation indices computed with top of atmosphere reflectance perform adequately in capturing crop phenology and land cover classification [17,47,48]. In particular, for relevant land cover classes—highly vegetated or cropland—and wavelengths (>650 nm for examining red, red-edge, and NIR values), atmospherically-corrected reflectance values and Sentinel 2 Level 1C reflectance values appear quite similar (see Figure 2 in [47] and Figure 3 in [48]). Additionally, studies have demonstrated that the Level 1C product possesses high r^2 (>0.9) with field spectrometry results across bands as well as good agreement for vegetation indices (NDVI) on vegetated land covers [49]. Though top of atmosphere absolute band values will not match atmospherically-corrected ones, we chose not to correct using a simple linear correction since our model of choice, the random forest, is not sensitive to monotonic transformations of predictor variables [50]. To improve data quality given Sentinel-2’s less reliable cloud mask [51], we leveraged the high temporal resolution (five-day return time) and selected observations based on maximum VI in a series of 2-week periods (Figure 2). This approach standardized the number of observations and improved their quality, since clouds tend to be highly reflective in the visible spectrum and suppress vegetation indices [51]. In total, we began with ~72,000 sampled 20-m pixels of yield data (Table 2). Of those, ~40,000 possessed at least one observation in each 2-week period (Table 2). In conjunction with the Landsat monthly filter described in the previous paragraph, these two data processing methods will be referred to as “monthly/biweekly” data henceforth.

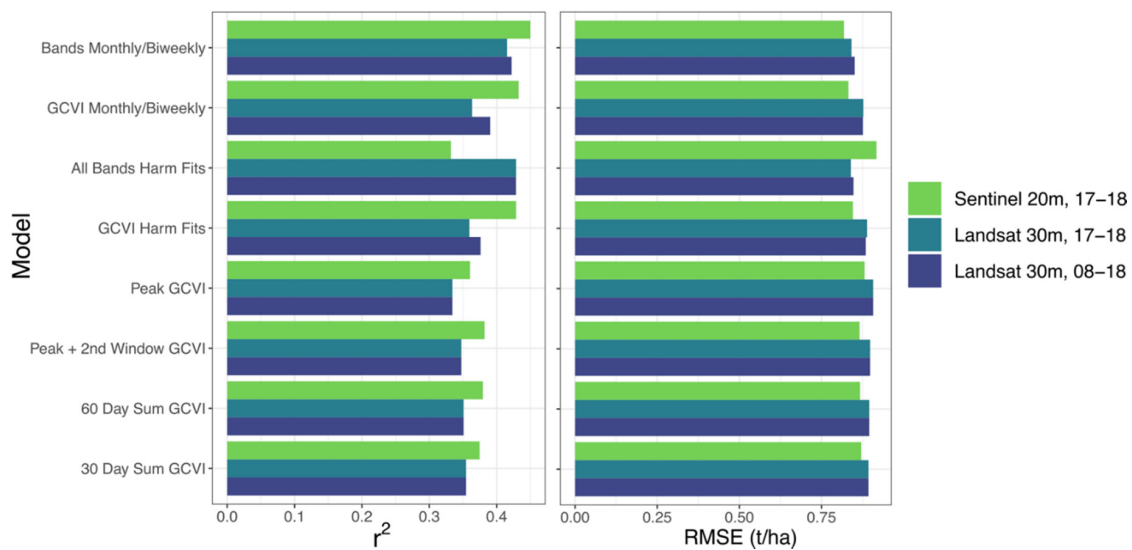


Figure 3. Pixel Scale Random Forests Comparison. Validation statistics for models trained on pixel yield data on held-out pixel-level validation data. Legend refers to the input satellite data, spatial resolution of yield and satellite data, and the temporal extent (Table 1, Section 2.5). The implementations here all employed random forests based on our three main feature engineering approaches: monthly/bi-weekly time series (for Landsat and Sentinel-2, respectively), harmonic fits, and VI-metrics (Section 2.4). For Sentinel, the preferred model is based on biweekly time series of all bands, while for Landsat the harmonic coefficients perform the best.

2.4. Harmonic Regressions and Feature Engineering

While the monthly or biweekly processing described above provides a consistent data structure, it reduces temporal resolution and discards potentially helpful data. As an alternative approach, we applied a harmonic regression (discrete Fourier transform) to the observed satellite data, for both Landsat and Sentinel (Figure 2). Fitting a harmonic regression in this context provides a smoothed, continuous function that can help capture the magnitude and timing of crop development and is key to agricultural outcomes (Ghazaryan et al., 2018; Wang et al., 2019). In particular, we fit a two-term harmonic regression:

$$f(t) = c + a_1 \cos(2\pi\omega t) + b_1 \sin(2\pi\omega t) + a_2 \cos(4\pi\omega t) + b_2 \sin(4\pi\omega t) \quad (1)$$

Here, t is the time step in days and ω is the frequency, which we set to 1.5, based on improved fit to the phenology of corn and soybeans in the US [14]. Additionally, we employed a recursive fitting approach for Sentinel to help capture the peak, due to the noisiness of the time series [52] (Figure 2). Of the fitted parameters, c represents the intercept, a_1 and a_2 are cosine coefficients, and b_1 , b_2 are sine coefficients. We used these fitted parameters for each individual band and, separately, GCVI as inputs to our machine learning models.

To quantify the fit of harmonic regressions on the satellite data in this study, we report r^2 for GCVI observations. For Landsat, the median r^2 of the harmonic regressions was 0.86 for GCVI. For Sentinel-2, the median r^2 of harmonic regressions and observed GCVI (after using the biweekly filter to remove cloudy observations missed by the cloud mask) was 0.80. Notably, though, r^2 for Sentinel-2 was considerably higher in 2018 ($r^2 = 0.84$) compared to 2017 ($r^2 = 0.70$).

Based on the harmonic predictions for GCVI (a daily time series), we constructed additional predictors, hereafter termed “VI metrics”, based on recent work demonstrating their potential [8,53]. Specifically, we compared two basic metrics—peak GCVI (“Peak”) and the peak GCVI along with GCVI 30 days later (“Peak + 2nd Window”)—alongside two cumulative VI-metrics—the sum of the

30 days following peak GCVI (“30-day sum”) and the sum of the thirty days on either side of the peak (“60-day sum”).

2.5. Modeling Approach

Our modeling approach encompassed three broad categories: (1) empirical models using pixel ground data; (2) empirical models using county level data from both ground data and agricultural statistics; and (3) models based on crop simulations (Table 1). At a high level, we trained a series of models on their respective training sets, selected the best model using evaluation metrics from performance on unseen validation data, then output an estimate of generalization error using the held-out test set for this preferred model.

We report both root mean squared error (RMSE) (2) and mean absolute error (MAE) (3), the latter being less sensitive to outlier values. RMSE and MAE are defined, respectively, as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3)$$

where n is the number of observations, \hat{y} is the predicted value, and y is the observed value. The error values for RMSE and MAE are in t/ha, the same units as measured yields. Additionally, we report the squared Pearson’s correlation coefficient (r^2) for our validation data and predictions as an estimate of explained variance.

2.5.1. Training, Validation, and Test Data

Broadly, we applied 80/10/10 splits to our data to assign observations to training, validation, and test sets, respectively. In order to apply a consistent split to both pixel- and county-level data, we considered each county-year as an observation for random assignment and assigned all pixel-scale observations within that county-year accordingly. As a result, our data splits were non-overlapping for a given year and county. We used two main data splits. One set, on which we evaluated Landsat harvester-trained models and compared to county-trained models, encompassed 11 years (2008–2018). The other set, on which we evaluated and compared Sentinel-2 and Landsat harvester-trained models, encompassed just the period 2017–2018. For the Landsat 11-year data, 186,160 soybean points (i.e., single pixel yield observations from unique field-years, see Section 2.2) possessed at least one clear Landsat observation each month (Table 2). Of those, 149,168 points went into a training set based on the county-year splits above. Of the remaining points, 18,333 went into the validation set and 18,659 went into the test set. For our analogous split on the Sentinel-2 data from 2017 to 2018, the training set had 31,791 observations, with 3629 in validation and 4140 in test. The Landsat data for 2017–2018 had the same split as for Sentinel.

2.5.2. Pixel Scale Random Forest Models

For our pixel-scale models, we established baseline model performance by passing all satellite bands and weather covariates to a random forest algorithm, using default parameters in the *sklearn* package in Python [54]. As in Section 2.1, we utilized 4-km gridded weather data from Gridmet [42]. For each month, we gathered average minimum and maximum daily temperatures, total precipitation, total solar radiation, and average vapor pressure deficit (VPD). For Landsat, predictors for red, green, blue, near-infrared (NIR), and short-wave infrared bands from each monthly period were used, in addition to a variable representing year. We also tested models based on the monthly time series of a single VI, along with weather covariates. Here we tested the performance of GCVI, based on demonstrated robustness to saturation at soybean’s high leaf area [37]. This type of model will be

referred to as “Landsat Harvester Trained” (Table 1). For the Landsat-based random forest models, we trained two versions: one using all years (2008–2018), and one using only data in the period 2017–2018 to enable comparison with Sentinel-2.

Analogously, we established baseline performance for Sentinel-2 using the biweekly time series for all bands and monthly weather covariates, including the three additional red-edge bands (Table 1). Then, we explicitly tested a series of VIs based on prior work, either for soybean LAI estimation or wheat and maize yield prediction elsewhere [46,55–62]. Table 3 provides a list of VIs tested, which include several red-edge VIs in order to assess the added value of these bands in yield prediction. In these comparisons, we still used the biweekly time series approach and weather covariates. Models trained using Sentinel-2 data in this way will be referred to as “Sentinel-2 Harvester Trained” (Table 1).

Table 3. Vegetation Indices. All vegetation indices tested, source for their derivation, and equation (RDED = red edge). Sentinel-2 analyses tested all vegetation indices. Landsat analyses tested GCVI.

Vegetation Index	Citation	Equation
Simple Ratio (SR)	Jordan, 1969 [55]	$\frac{NIR}{RED}$
Normalized Difference Vegetation Index (NDVI)	Rouse et al., 1973 [63]	$\frac{NIR-RED}{NIR+RED}$
Green Chlorophyll Vegetation Index (GCVI)	Gitelson et al., 1996 [56]	$\frac{NIR}{GREEN} - 1$
Near Infrared Reflectance of vegetation (NIRv)	Badgley et al., 2017 [57]	$\frac{NIR-RED}{NIR+RED} \times NIR$
Optimized Soil Adjusted Vegetation Index (OSAVI)	Rondeaux et al., 1996 [58]	$1.16 \frac{NIR-RED}{NIR+RED+0.16}$
Sentinel-2 LAI-Green Index (SeLI)	Pasqualotto et al., 2019 [59]	$\frac{NIR-RDED1}{NIR+RDED1}$
MERIS Terrestrial Chlorophyll Index (MTCI)	Dash and Curran, 2004 [60]	$\frac{NIR-RDED1}{RDED1-RED}$
Modified Chlorophyll Absorption in Reflectance Index (MCARI)	Daughtry et al., 2000 [61]	$\frac{(RDED1-RED)-0.2(RDED1-GREEN) \times \frac{RDED1}{GREEN}}{1.16(NIR-RDED1)(NIR+RDED1+0.16)}$
Chlorophyll Index, Red-Edge (Cir)	Gitelson et al., 2003 [46]	$\frac{RDED3}{RDED1} - 1$
Normalized Difference Red Edge Index, 1 (NDRE1)	Gitelson and Merzlyak, 1994 [62]	$\frac{RDED2-RDED1}{RDED2+RDED1}$
Normalized Difference Red Edge Index, 2 (NDRE2)	Gitelson and Merzlyak, 1994 [62]	$\frac{RDED3-RDED1}{RDED3+RDED1}$

Additionally, we tested a series of feature engineering approaches based on the harmonic regressions for both Landsat and Sentinel-2 data (Section 2.4). First, we used the harmonic regression coefficients for all bands (“All Bands Harmonic Fits”), and then for GCVI only (“GCVI Harmonic Fit”). These decompositions encode the magnitude and timing of crop spectral signatures, with demonstrated effectiveness for crop classification and yield prediction [14,53]. Second, we used the VI-based metrics described in Section 2.4 (Peak GCVI, Peak GCVI + 2nd Window, 30-day sum GCVI, 60-day sum GCVI). To compare these approaches, we tested individual random forests given one of the six VI-based metrics along with monthly weather covariates.

2.5.3. County Scale Random Forest Models

In order to better understand the differences between training on our pixel yield data and more commonly used county-level yields, we trained a random forests model using county-level data (“County trained”, Table 1). To do so, we sampled 50 random points classified as soybean based on USDA’s Cropland Data Layer (CDL) in each county [64]. As above (Section 2.5.2), we derived a monthly Landsat and Gridmet weather time series for each of these points. We then aggregated to a single training example per county-year by averaging each of our spectral and weather covariates. We employed the same baseline model implementation described in Section 2.5.2, with monthly Landsat observations for all bands and monthly weather covariates. The response variable was USDA’s NASS county-level average yield.

For this analysis, we split the available county-year observations into only a training set and a test set (80/20) by combining the county-years in the validation and test sets defined in Section 2.5.1. We do so since we only examined the baseline model and thus did not carry out model selection for both harvester-trained and county-trained models. This relatively simple random partitioning of county-years into training and test sets enabled consistency with the full Landsat harvester-trained random forest analysis and thus was elected over alternative out-of-sample error estimation approaches (i.e., cross validation).

Most often, models are trained with county statistics since those data are freely available. In order to understand how a county-trained model might perform on fine-scale yield data, we tested our county-trained model on the 30-m harvester yield data, and compared performance with the baseline Landsat harvester-trained model (Section 2.5.2) across both scales and on the period 2008–2018. First, we took the model trained with 30-m harvester data, applied it to the county-aggregated data, and evaluated with NASS county yield data in the combined validation/test set defined in the previous paragraph. Then, we applied the county-trained model to 30-m harvester yields in these same county-years (predictions from the Landsat harvester-trained model were also on the same validation set).

To put our county-level results in context, we compared performance to that of a null model in which county predictions were made based on that county's yield trend over the 11-year study period. In other words, for each county in our sample, we fit a linear regression predicting yield with year, and used this simple model to make predictions for all counties (2008–2018). Furthermore, we used this null model to help differentiate spatial and temporal variability. To do so, we calculated the distribution of r^2 by both county and year for the null model and the harvester- and county-trained models. In order to have a robust enough sample to do so, we calculated these metrics on the full sample of county-years to have as many counties with the full 11 years of data as possible. This means that the by-county and by-year r^2 estimates are positively biased because they include predictions on data from the training set. Thus, while the absolute values should not be interpreted as an estimate of generalization error, the relative distributions help describe the capacity of the random forest models to capture temporal variability.

Finally, we leveraged one of the strengths of random forests, their ability to output relative variable importance, as a way to explore differing results when modeling with empirical data of different scales. Specifically, we compared variable importance measures in *sklearn* based on predictors that have the largest effect on decreasing variance within a “node” or split in the decision tree. Feature importance can offer information as to which covariates are most predictive of soybean yields.

2.5.4. Pixel Scale Simulations-Based Model

In order to build simulations-based models, we employed the satellite-based Scalable Crop Yield Mapper, or SCYM [26]. Specifically, this process takes LAI, biomass, and yield output from the Agricultural Production Systems sIMulator (APSIM) [65] and fits a multiple linear regression using a subset of weather covariates and VI (estimated from simulated LAI) to predict end-season yield. The simulations vary management parameters—e.g., sowing date, sowing density, and fertilizer application—across realistic ranges to produce a distribution of potential outcomes. We converted LAI from APSIM to GCVI via a linear regression based on field experiments conducted in Nebraska [37]:

$$\text{GCVI} = 1.1 + 1.4 \times \text{LAI}^{1.3} \quad (4)$$

The statistical relationships to predict yield from VI and weather are then deployed at scale to observed satellite and weather data using Google Earth Engine. SCYM has been used to create 30-m resolution soybean yield maps across 3 of the 9 states in our study area (Indiana, Illinois, and Iowa) [26,66]. The initial implementation used maximum observed GCVI in two time windows, early and late season, and employed a regression model trained specifically for the pair of observation

dates available at each pixel. To do so, regressions were derived for all possible combinations of early and late season observation dates. Here we updated the SCYM methodology to better capture yield variability across the nine-state region, informed by our validation data and variable importance from the random forests models.

To begin updating the SCYM methodology, we returned to the APSIM simulations. Since we were evaluating over a larger area than the initial implementation, we added two new training sites in the more northern latitudes, one in South Dakota and one in Minnesota (Table 4). We retained the six original training sites, spread across Iowa, Illinois, and Indiana. For each of the eight sites, we gathered Gridmet weather data at a daily time step exported using Google Earth Engine [42]. We largely maintained management parameters as in the original implementation (Table 4). We employed two cultivars in the simulations, Pioneer93M42 3.4 and Pioneer_94B01 4.0, with their default parameters in APSIM. We chose similar cultivars to past work using APSIM-Soybean [67], and maturity groups appropriate for a large swath of our geographic extent. For soil, we used the same basic soil profile as in previous implementations based on a Johnston, Iowa study site (coordinates 41.73N, 93.72W) [9,26] and varied nitrogen application and soil water at sowing (Table 4). Crop specific water and nitrogen responses were controlled by the default soybean parameters from APSIM.

Table 4. Summary of Agricultural Production Systems sIMulator (APSIM) settings for crop simulations. Comments note changes from Lobell (2015) baseline parameters [26].

Factor	Values Used	Units	Comments
Year	2007–13		
Site	Newton, IA (−93.1°E, 41.7°N) Marshalltown, IA (−92.9°E, 42.1°N) Clinton, IL (−89.0°E, 40.1°N) Chenoa, IL (−88.7°E, 40.7°N) Marion, IN (−85.7°E, 40.6°N) Munice, IN (−85.3°E, 40.2°N) Benson, MN (−95.6°E, 45.3°N) Aberdeen, SD (−98.5°E, 45.5°N)		Latter two sites added from baseline
Fertilizer Rate	0, 25, 50	kg of urea N per ha	
Sowing Density	3, 5, 7	Plants per m ²	
Row Spacing	380	mm	Reduced from baseline as per [67]
Cultivar Choice	Pioneer93M42 3.4, Pioneer_94B01 4.0		Similar cultivars as [67]
Soil Water At Sowing	0.8, 1.0	% of total water holding capacity	
Sow Date	April 25, May 5, May 20, June 14		Added April 25th date for additional variability

Based on successful implementation in maize SCYM [9], we tested whether predicting plant biomass, scaled by a constant harvest index to output yield, improves model performance versus directly estimating yield. We expect that complex grain-filling mechanics are challenging to capture in crop simulations; employing a more reliable estimation of biomass as the dependent variable in our regressions, then, could help yield estimation when applied to observed imagery. We held the harvest index constant across years, selected to minimize RMSE on validation data.

Drawing on the modeling work above, we tested a series of SCYM regressions using both the suite of VI metrics defined in Section 2.4 and subsets of weather covariates. Specifically, we tested each of the six VI metrics without weather, with the 4 baseline weather covariates, and with new sets of weather covariates based on random forest variable importance. The baseline approach's weather predictors were based on seasonal radiation and rainfall, July VPD, and August max temperature.

Since SCYM models train only on simulated data, we allocated the 30-m harvester training set for model selection. We report error at the pixel (30 m) scale using the combined harvester data set of county-years from unseen validation and test sets, as with county-scale analysis described in Section 2.5.3. Additionally, we report validation at the county scale to compare with past work in this area using these same held-out county-years. This is the same validation set of county-years used to compare error for the baseline Landsat harvester-trained and county-trained (Section 2.5.3). To do so, we averaged the models 30-m outputs in a given county-year.

3. Results

3.1. Pixel Scale Random Forest Models

The pixel-scale random forest model applied to Landsat imagery from 2008–2018 was able to explain 44% of the yield variability against the test dataset ($r^2 = 0.44$, RMSE = 0.85 t/ha). When focusing just on 2017–2018, the 20-m Sentinel-2 model performed similarly ($r^2 = 0.45$, RMSE = 0.82 t/ha) and slightly better than the Landsat model compared against data from the same 2017–2018 period ($r^2 = 0.41$, RMSE = 0.84 t/ha). The preferred Sentinel-2 model included all available bands using biweekly time series; for Landsat, a model with all harmonic regression coefficients for each band, except for blue, outperformed a model with monthly observations (Figure 3). This improvement in performance suggests that Landsat, with its lower temporal resolution, can benefit from the harmonic's ability to capture magnitude and timing of phenology. On the other hand, despite residual noisiness in the biweekly Sentinel-2 time series (Figure 2), the models did not improve with harmonic fits, whether for individual bands or GCVI. This may be due to the relatively worse fit of harmonic regressions on Sentinel-2 data compared to Landsat, particularly in 2017 (Section 2.4). The all-bands harmonic fit performed the worst for Sentinel-2 (Figure 3), which may have resulted from higher dimensionality (15 additional predictors from the three red-edge bands compared to Landsat implementation) in addition to the lower harmonic performance.

Variable importance measures for the baseline Landsat harvester-trained model indicated that the models relied more on spectral observations near the peak for VIs, in late July and August (Figure 4). Placing relatively little weight on early or late spectral observations, for which cloudy observations are much more likely, minimizes the interference from those noisy periods. It may also be that these periods are simply less informative for yield prediction, since our data should have relatively little interference from clouds after filtering. Overall, the random forest models performed better with individual bands than pre-calculated VIs or VI-related metrics. This finding reflects the ability of machine learning models to discover interactions between bands that are useful for predictions, even in a high-dimensional setting.

The VI-based metrics derived from harmonic regressions (peak, peak+ 2nd-window, cumulative) performed similarly for Landsat and Sentinel (Figure 3). In both cases, VI-based metrics explained between 75 and 85% of the variability captured by the preferred models (Figure 3). Corresponding increases in RMSE between 5 and 10% (moderately increased bias) suggest that compressing VI-related signal into one or two predictors can still result in similar error, when also given weather data as our models were. Between the four VI-based metrics, cumulative VI metrics did not capture substantially more variability than simpler metrics like peak VI for soybean (Figure 3).

In testing a series of Sentinel-derived VIs at the 20-m scale, the red-edge VIs performed similarly to the most robust VIs using wavelengths available on other sensors (e.g., NIRv and GCVI; Table 5). Overall, NIRv performed the best, but many red-edge VIs performed nearly as well (Table 5). The canopy/chlorophyll-related VIs generally outperformed “structural” VIs such as MCARI or MTCI (Table 5). With the caveat that we employed top-of-atmosphere data, red-edge VIs did not add a significant amount to the traditional bands. Moreover, we found that including the three red-edge bands' biweekly time series in a random forest model did not improve or hurt RMSE or r^2 compared

to a model without them (but with the other bands); it seems that any information gained was largely cancelled out by the cost of increased dimensionality.

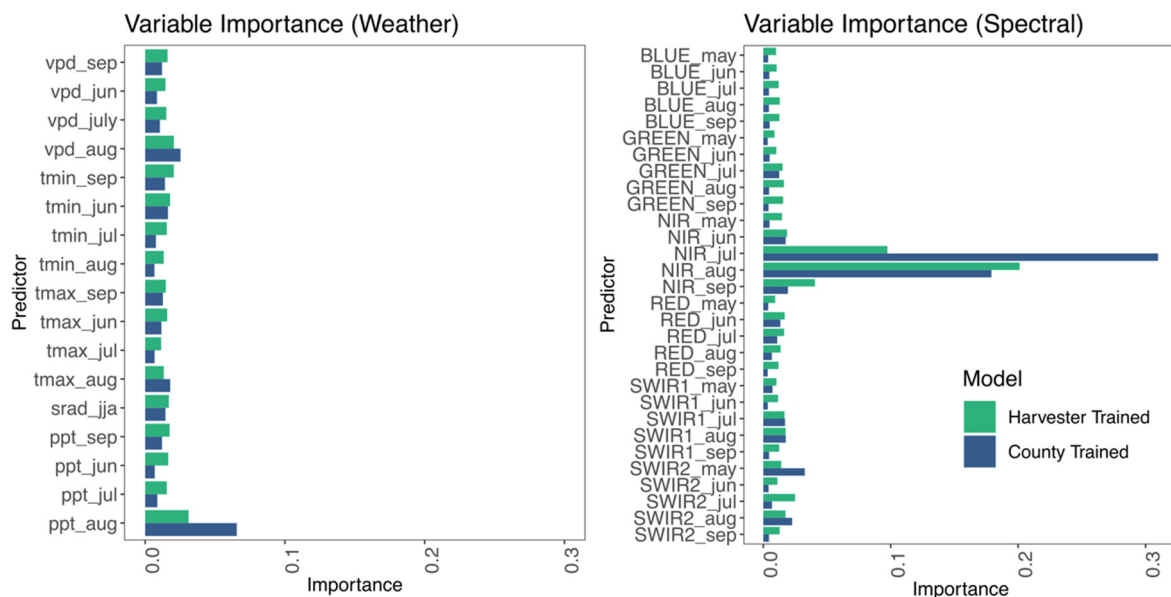


Figure 4. Variable Importance from County- and Harvester-Trained random forest (RF). The figure displays variable importance, measured by increase in node purity (decrease in variance), from random forest models trained at the pixel-scale using harvester yield data and the county-scale using NASS data. Both models were trained on data from 2008 to 2018 using monthly Landsat and weather data.

Table 5. Sentinel-2 Vegetation Index Performance. Validation statistics reported from performance on held-out validation set for a suite of vegetation indices.

Vegetation Index	r ²	RMSE (t/ha)	MAE (t/ha)
SR	0.42	0.841	0.648
NDVI	0.42	0.840	0.648
GCVI	0.43	0.832	0.641
NIRv	0.45	0.822	0.634
OSAVI	0.42	0.839	0.648
SeLI	0.44	0.828	0.638
MTCI	0.39	0.863	0.665
MCARI	0.29	0.932	0.720
Cir	0.44	0.829	0.638
NDRE1	0.43	0.835	0.643
NDRE2	0.44	0.829	0.637

3.2. County-Scale Random Forest Models

The county-trained random forests explained over 80% of variance in unseen county-years ($r^2 = 0.82$, RMSE = 0.24 t/ha, Table S1). Since our interest here was to assess the difference in feature importance and generalization between models trained at fine and coarse scales (see Methods Section 2.5.3), we also compared validation across scales. Using the Landsat harvester-trained model to predict average yields in new county-years explained nearly as much variability in NASS county yield statistics ($r^2 = 0.79$, RMSE = 0.65 t/ha, Figure 5, Table S1). Although variation explained (r^2) in held-out NASS county yield data was quite high, the harvester-trained model did display a strong positive bias. This might be expected based on the higher average yields in the yield monitor dataset, discussed in Section 2.2.

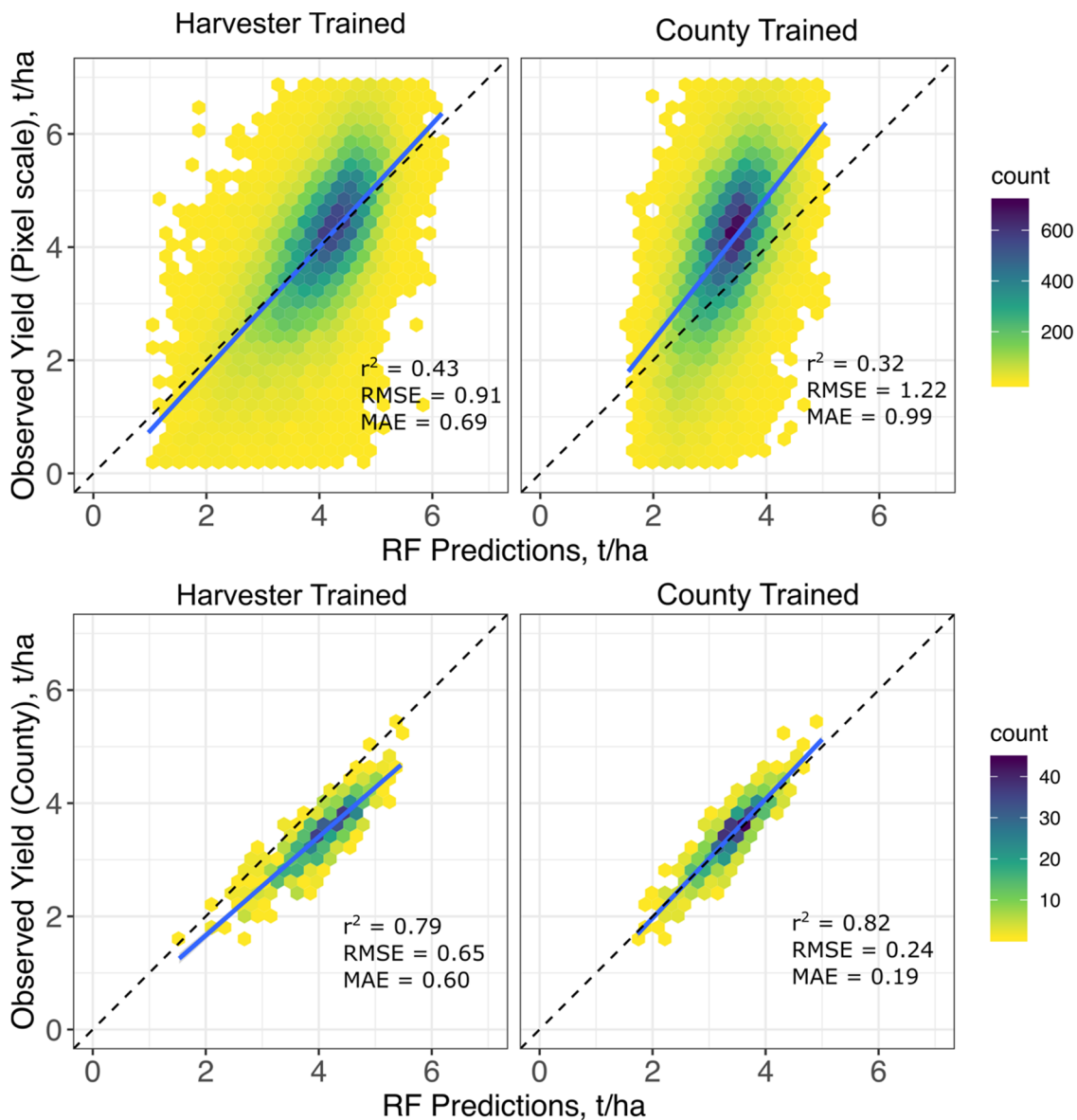


Figure 5. County- and Pixel-Scale Random Forest Validation. The figure shows validation statistics on held-out data for random forests trained at the pixel-scale (Harvester Trained) and county-scale (County Trained). Both models employ a monthly time series of Landsat and weather data. RMSE and MAE are reported in t/ha.

The null model (county trend) described in Section 2.5.3 achieved an overall r^2 of 0.70 (RMSE = 0.33 t/ha, MAE = 0.26 t/ha) across all county-years. This null model displayed very low bias, likely due to the removal of spatial variability by training within county. Compared to this null model, both the harvester- and county-trained random forests explained more variation in NASS county yields across the study period (r^2 of 0.79 and 0.82, respectively; Figure 5).

To examine temporal versus spatial variability captured by the models, we examined by-county and by-year r^2 on the full data set (Section 2.5.3). The median r^2 by county for the null model was 0.43; the median r^2 by county was 0.74 for the Landsat harvester-trained model and 0.96 for the county-trained model. The median r^2 by year was 0.70 for the null model; because the model predicts using by-county trend lines, it quite effectively captures spatial variability. Comparatively, the r^2 by year was 0.69 for the Landsat harvester-trained model and 0.95 for the county-trained model. Together, these results indicate that the random forest approaches do effectively capture temporal and spatial variability, particularly so for the county-trained model.

Furthermore, we tested the county-trained model at the pixel scale, using the 30-m harvester data. Applying the county-trained model to fine scale data resulted in a large decrease in performance relative to the Landsat harvester-trained model ($r^2 = 0.32$ vs. 0.43 , Figure 5). Although at very different levels of aggregation, applying Landsat harvester-trained model to the county data decreased r^2 by 3 percentage points while applying county-trained to the harvester data decreased r^2 by 11 percentage points relative to models trained and tested at those scales.

County-level models relied more heavily on a small subset of variables, learning a simpler function that did not generalize to the pixel scale (Figures 4 and 5). In particular, the county-trained model placed a larger emphasis on August weather variables (VPD, precipitation, maximum temperature), while using almost exclusively the NIR in July and August out of the spectral bands. The Landsat harvester-trained model also emphasized the July and August NIR, but otherwise distributed importance more widely across inputs. To test whether a county-trained model was indeed learning a simpler function, we trained and tested models at both scales given only the six covariates with greatest importance scores from the county-trained model (NIR in July and August, August precipitation and VPD, and SWIR2 in May and August). On the pixel-scale validation data, variance explained decreased by 11 percentage points (r^2 : $0.43 \rightarrow 0.32$, RMSE: 0.92 t/ha \rightarrow 0.98 t/ha) for the harvester-trained model but only by 4 points (r^2 : $0.32 \rightarrow 0.28$, RMSE: 1.22 t/ha \rightarrow 1.26 t/ha) for the county-trained model. At the county scale, the harvester-trained model explained far less variance (r^2 : $0.79 \rightarrow 0.50$, RMSE: 0.65 t/ha \rightarrow 0.92 t/ha), while the county-trained model actually improved (r^2 : $0.82 \rightarrow 0.93$, RMSE: 0.24 t/ha \rightarrow 0.16 t/ha). The county-trained random forest seemed to benefit from reduced dimensionality at its native scale, while the harvester-trained model performed considerably worse at both scales.

Based on these results, it appears county-trained models depended highly on a small subset of weather covariates that lacked signal at the pixel scale. This cross-scale experiment highlights the importance of fine-scale validation data for understanding the performance of yield mapping algorithms. As demonstrated, algorithms that rely on county-scale yield labels for training may not be able to capture the underlying population model at a subfield, pixel scale.

3.3. Simulations-Based Models (SCYM)

Validated on 30-m harvester data, the biomass SCYM model using harmonic “Peak + 2nd Window” performed the best (Table 6, Figure 6). At this scale, the preferred SCYM model explained 27% of yield variability on a held-out test set ($r^2 = 0.27$). Applying this method to the full set of available soy CDL points for our validation county-years resulted in r^2 of 0.63 and RMSE = 0.4 t/ha at the county scale (Table S1). The median r^2 by county was 0.63 as well; compared to the null model presented in Section 3.2 above, the SCYM model outperformed in terms of capturing temporal variability (median r^2 by county based on the county trend was 0.43). Additionally, the median r^2 by year was 0.53 .

Though ultimately a somewhat naïve approach, predicting biomass and scaling by a constant harvest index led to a more robust model than predicting yields directly. Despite ignoring some variability from late-season (post-August) weather effects, eliminating other sources of error improved performance. Relative to work in maize [9], implementing the biomass \times harvest index approach did not improve performance as much in soybean. This might be due to a weaker relationship in soybeans between high LAI or biomass and end-season yields. Using APSIM simulations in maize and soybean over a suite of management scenarios, maize yield and biomass exhibited a much stronger linear relationship ($r^2 = 0.95$, $r^2 = 0.50$ for soybean). The looser yield-biomass relationship in soybean may occur for a variety of reasons, including increasing water stress and decreasing light use efficiency at high LAIs, making for a less predictable response [67,68]. In addition to netting only a small improvement in r^2 , switching to the biomass-based approach seemed to reduce the variance of predictions. As a result, SCYM tended to overpredict in low-yielding counties and underpredict in high-yielding ones. This phenomenon likely stemmed from a combination of the limited weather effect and the constant biomass scaling, meaning that SCYM was unable to differentiate fields which

achieved high LAIs and average yields from fields with high LAIs and outstanding yields. In this area, in particular, there remains opportunity to improve the soybean SCYM methodology in future work.

Table 6. SCYM Model Performance across Implementations. Baseline 2-Window refers to implementation from Lobell et al. 2015 [26]. “No met” means weather covariates not included. “Met” indicates the original four weather covariates (Methods). “Aug rain” models include only August rainfall as weather covariate.

SCYM Model	r^2	RMSE (t/ha)	MAE (t/ha)
Baseline 2-Window	0.24	1.09	0.86
Peak GCVI, Met	0.25	1.00	0.77
Peak GCVI, Aug Rain	0.26	0.96	0.74
Peak GCVI, No Met	0.24	0.98	0.75
Peak GCVI, DOY, Met	0.18	1.11	0.85
60d Sum, No Met	0.23	1.42	1.15
60d Sum	0.24	1.05	0.83
60d Sum, Aug Rain	0.24	1.12	0.90
30d Sum, No Met	0.24	1.00	0.80
30d Sum, Met	0.25	0.98	0.75
30d Sum, Aug Rain	0.24	1.00	0.76
Peak + 2nd Window, No Met	0.26	0.97	0.76
Peak + 2nd Window, Met	0.27	1.01	0.79
Peak + 2nd Window, Aug Rain	0.27	0.96	0.75

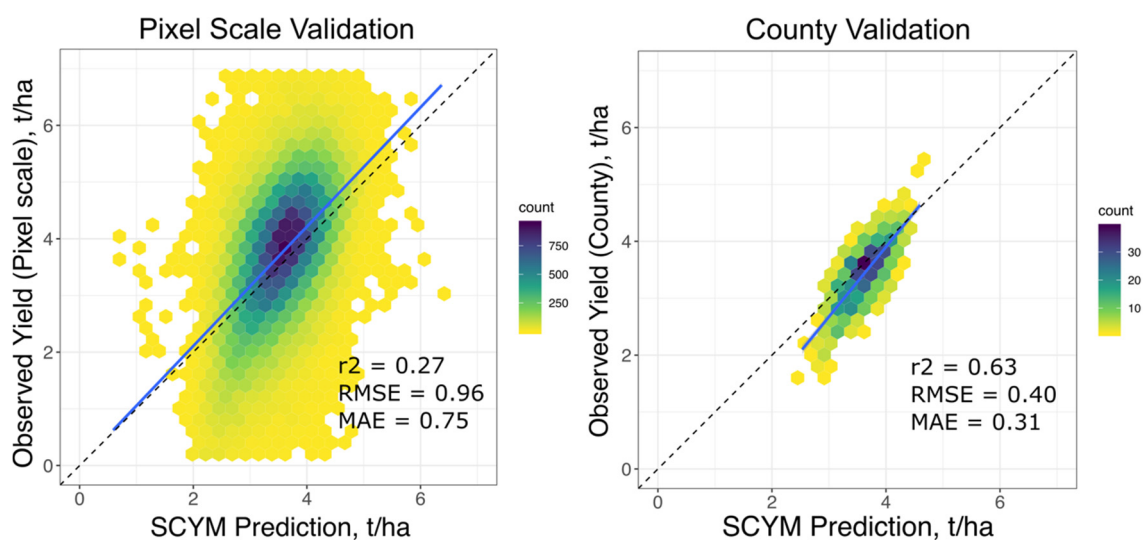


Figure 6. SCYM Performance at the Pixel and County Scale. Figure displays SCYM performance against pixel- and county-scale validation yields, across 2008–2018. Data are shown for the same county-years which were held out from model selection. County validation, at right, is shown based on averaged predictions across all soybean CDL pixels for that year, not just fields for which we had harvester data. RMSE and MAE are reported in t/ha.

In general, the SCYM regressions were unable to effectively incorporate additional meteorological data. Including sets of four weather covariates, both from the baseline implementation [26] and based on random forest variable importance, generally did not improve model performance on our validation set. However, as the random forests and the literature consistently pointed to August precipitation as highly impactful, we tested a model with August precipitation as the only weather input [69,70]. This specification performed the best on our validation set (Table 6). Comparison to the random forests revealed further evidence of SCYM’s limited ability to incorporate additional signals from weather data. The random forests’ partial dependence on peak GCVI showed that the underlying response function was essentially linear, and a random forest with only GCVI data performed quite similarly to SCYM ($r^2 = 0.27$, RMSE = 0.96 t/ha).

Together, these indicate that linear SCYM models captured the effect of GCVI on yield, but struggled to gain additional information from weather—beyond what APSIM built into the simulated LAIs.

4. Discussion

4.1. Pixel-Scale Yield Prediction

Direct comparison with other field-scale yield estimation studies is difficult because of inherent differences in crop type, spatial and temporal extent, and heterogeneity of soil and other factors. Nonetheless, we note that our work here compares favorably with models using spectral and weather inputs but underperforms for more data-intensive approaches drawing on either extremely high-resolution hyperspectral imagery or additional soil/environmental covariates. Our ability to estimate field-scale yield using random forests is similar to work over smaller spatial extents in maize ($r^2 = 0.50$) and soybean in Brazil (MAE = 0.28) [6,24]. On the other hand, our models achieved significantly lower accuracy than work using Sentinel-2 and random forests in British wheat systems, using a data set of 39 fields over a limited spatial extent [5]. Though accessing additional soil moisture data, that work's much higher r^2 (0.91) indicates either that wheat is easier to map using random forests and spectral imagery, or simply that our larger spatio-temporal extent (1000–10,000x as many fields) significantly increased the estimation challenge by increasing heterogeneity. Data assimilation approaches that incorporate remotely sensed data into crop simulations achieved moderately higher mean absolute percentage error (MAPE, 17% vs. 21–23%) predicting subfield-scale yields in Midwestern maize [10]. MAPE is defined as

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

where n represents the number of observations, \hat{y} the predicted value, and y the observed value. Finally, deep learning models with very high-resolution hyperspectral data from an unmanned aerial vehicle outperform our models here ($r^2 = 0.72$) on sub-10-m-scale yields, though they were evaluated on a small series of plots in a single season [71]. Overall, these studies generally have a reduced spatio-temporal extent compared with our study, ranging from 1 to ~700 fields.

4.2. County-Scale Yield Prediction

The county-level random forest, built by aggregating pixel-level Landsat predictors (as opposed to aggregating pixel-level predictions), explained over 80% of variance in unseen county-years ($r^2 = 0.82$, RMSE = 0.24 t/ha). This relatively simple implementation matched the performance of many county-scale approaches proposed in the literature, including complex deep learning models, which have resulted in RMSE between 0.3 and 0.4 t/ha on the same study area, over a smaller subset of years [7,29].

The two random forest models, one trained at the pixel scale and one at the county scale, overall agreed with regard to weather covariate importance (Figure 4). Their variable importance reflected similar patterns as in past work on soybean yield determinants. Analysis of historical yield trends in the Midwest, 1960–2006, emphasized the importance of August rainfall for soybean, with deviation above/below the mean by 1.5 inches changing yield outcomes by 0.1–0.25 t/ha [70]. Analogously, another study analyzing soybean yield responses also concluded that August precipitation had a large positive effect on yield outcomes, with an equivalent change in precipitation associated with between 0.2 and 0.6 t/ha increase in yield [69]. The effect was significant in only three out of the nine Midwestern states analyzed here, however. Both of these historical analyses examined yields aggregated to the state level in the Midwest, using linear regression to infer impacts of weather on soybean yields. These findings align with the random forests models, both of which determined August rainfall to be the weather covariate most predictive of yield. At the same time, placing emphasis on the NIR measurements near peak LAI for soybean makes intuitive sense, as NIR is sensitive to leaf structure [72],

and its interactions with other spectral bands can effectively capture crop LAI and canopy chlorophyll, signals of plant health [37,73].

4.3. Scalable Yield Mapping

As detailed above, based on held-out 30-m harvester data, the preferred SCYM model used biomass \times harvest index predictions with the harmonic “Peak + 2nd Window” implementation (Table 6, Figure 6). Although the harmonic implementation only slightly outperformed the previous two-window method (Table 6), the harmonics methodology will likely translate much better to regions across the world given that the percentage of clear (cloud-free) growing-season observations in many equatorial regions, such as Sub-Saharan Africa or India, is much lower than the US Midwest [74]. For these regions, harmonic regressions have proved effective in the context of small-holder yield prediction [17,75].

The relatively small difference between peak and cumulative VI-metrics (Table 6), particularly without weather covariates, differed from previous work showing that cumulative VI metrics explained more than twice as much variability in wheat yields compared to simple peak or two-window observations in models without weather data, and ~30% more when models included weather [53]. Here, we observe that cumulative metrics improve performance slightly compared to a single peak VI, but worsened it when compared with a harmonic-based two-window approach. In a more similar context, our finding aligns with other work on Midwestern soy yields for which cumulative VIs only marginally improved explained variability compared to peak VI, when using multiple satellite fusion to gain the requisite temporal resolution [8].

5. Conclusions

Using an extensive ground-truth data set of high-resolution yield data, we found that random forest models could explain up to 45% of pixel-level yield variability using satellite imagery and weather data. For Landsat-based models, applying harmonic regressions in this machine learning context markedly improved performance. Overall, pixel-based random forest models were not very sensitive to feature engineering approaches. Comparing these pixel-scale models with models trained on county data, we found that pixel-scale models performed similarly at the county-scale. On the other hand, county models performed relatively poorly on pixel-scale data by learning a simplified response function that did not effectively model fine-scale yields. Translating features which performed well in the random forests context did not significantly improve performance of the simulations-based SCYM model, which performed similarly to county-based random forest models at explaining pixel-scale yields. Due to potential shortcomings of the model simulations or linear regression approach, the simulations-based models output low variance estimates that performed poorly farther away from the mean. Additional work on these shortcomings is needed, in order to build a highly accurate—in addition to scalable—model.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-4292/12/21/3471/s1>, Table S1: All County-Year Yields and Model Predictions. Table S1 includes observed county yields, by county code and year, for all county-years in the study area and study period, along with model predictions at the county level from the baseline Landsat harvester-trained model, the Landsat county-trained model, and the preferred SCYM model. On-farm yield monitor data are not available, to ensure privacy in accordance with the data use agreement. All other data are accessible as cited and described in the Methods section.

Author Contributions: Conceptualization, W.T.D., D.B.L., J.M.D.; methodology, W.T.D., D.B.L., J.M.D.; software, W.T.D., J.M.D.; validation, W.T.D., D.B.L., J.M.D.; formal analysis, W.T.D.; resources, W.T.D., R.P., S.-Z.L.; data curation, R.P., S.-Z.L.; writing—original draft preparation, W.T.D.; writing—review and editing, W.T.D., D.B.L., J.M.D., R.P., S.-Z.L.; visualization, W.T.D.; supervision, D.B.L.; project administration, D.B.L.; funding acquisition, D.B.L., J.M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NASA Harvest Consortium, NASA Applied Sciences Grant No. 80NSSC17K0625, sub-award 54308-Z6059203 to DBL.

Acknowledgments: We thank George Azzari for work on the initial SCYM implementation, Zhenong Jin for his help with APSIM simulations, and Jake Campolo, Sherrie Wang, Stefania Di Tomasso, Hemant Pullabhotla, and Matthieu Stigler for helpful comments and support throughout. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Jain, M.; Singh, B.; Rao, P.; Srivastava, A.K.; Poonia, S.; Blesh, J.; Azzari, G.; McDonald, A.J.; Lobell, D.B. The impact of agricultural interventions can be doubled by using satellite data. *Nat. Sustain.* **2019**, *2*, 931–934. [\[CrossRef\]](#)
- Lobell, D.B. The use of satellite data for crop yield gap analysis. *Field Crop. Res.* **2013**, *143*, 56–64. [\[CrossRef\]](#)
- Basso, B.; Dumont, B.; Cammarano, D.; Pezzuolo, A.; Marinello, F.; Sartori, L. Environmental and economic benefits of variable rate nitrogen fertilization in a nitrate vulnerable zone. *Sci. Total Environ.* **2016**, 227–235. [\[CrossRef\]](#) [\[PubMed\]](#)
- Diker, K.; Heermann, D.; Brodahl, M. Frequency Analysis of Yield for Delineating Yield Response Zones. *Precis. Agric.* **2004**, *5*, 435–444. [\[CrossRef\]](#)
- Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410. [\[CrossRef\]](#)
- Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. *Remote Sens.* **2019**, *11*, 2873. [\[CrossRef\]](#)
- You, J.; Li, X.; Low, M.; Lobell, D.; Ermon, S. Deep Gaussian process for crop yield prediction based on remote sensing data. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4559–4565.
- Gao, F.; Anderson, M.; Daughtry, C.S.T.; Johnson, D. Assessing the Variability of Corn and Soybean Yields in Central Iowa Using High Spatiotemporal Resolution Multi-Satellite Imagery. *Remote Sens.* **2018**, *10*, 1489. [\[CrossRef\]](#)
- 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.* **2017**, *247*, 207–220. [\[CrossRef\]](#)
- Kang, Y.; Ozdogan, M. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sens. Environ.* **2019**, *228*, 144–163. [\[CrossRef\]](#)
- Maestrini, B.; Basso, B. Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. *Sci. Rep.* **2018**, *8*, 14833. [\[CrossRef\]](#)
- Robertson, M.J.; Llewellyn, R.S.; Mandel, R.; Lawes, R.; Bramley, R.G.V.; Swift, L.; Metz, N.; O’Callaghan, C. Adoption of variable rate fertiliser application in the Australian grains industry: Status, issues and prospects. *Precis. Agric.* **2011**, *13*, 181–199. [\[CrossRef\]](#)
- Rodriguez-Galiano, V.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [\[CrossRef\]](#)
- Wang, S.; Azzari, G.; Lobell, D.B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* **2019**, *222*, 303–317. [\[CrossRef\]](#)
- Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [\[CrossRef\]](#)
- Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [\[CrossRef\]](#)
- Jin, Z.; Azzari, G.; You, C.; Di Tommaso, S.; Aston, S.; Burke, M.; Lobell, D.B. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* **2019**, *228*, 115–128. [\[CrossRef\]](#)
- Saeed, U.; Dempewolf, J.; Becker-Reshef, I.; Khan, A.; Ahmad, A.; Wajid, S.A. Forecasting wheat yield from weather data and MODIS NDVI using Random Forests for Punjab province, Pakistan. *Int. J. Remote Sens.* **2017**, *38*, 4831–4854. [\[CrossRef\]](#)

19. Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.; Ciampitti, I.A. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. For. Meteorol.* **2020**, *284*, 107886. [[CrossRef](#)]
20. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
21. Aghighi, H.; Azadbakht, M.; Ashourloo, D.; Shahrabi, H.S.; Radiom, S. Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4563–4577. [[CrossRef](#)]
22. Yue, J.; Feng, H.; Yang, G.; Li, Z. A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy. *Remote Sens.* **2018**, *10*, 66. [[CrossRef](#)]
23. Cai, Y.; Guan, K.; Lobell, D.B.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [[CrossRef](#)]
24. Richetti, J.; Judge, J.; Boote, K.J.; Johann, J.A.; Opazo, M.A.U.; Becker, W.R.; Paludo, A.; Silva, L.C.D.A. Using phenology-based enhanced vegetation index and machine learning for soybean yield estimation in Paraná State, Brazil. *J. Appl. Remote Sens.* **2018**, *12*, 026029. [[CrossRef](#)]
25. Han, L.; Yang, G.; Dai, H.; Xu, B.; Yang, H.; Feng, H.; Li, Z.; Yang, X. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods* **2019**, *15*, 1–19. [[CrossRef](#)] [[PubMed](#)]
26. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [[CrossRef](#)]
27. Burke, M.; Lobell, D.B. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2189–2194. [[CrossRef](#)]
28. Jain, M.; Srivastava, A.K.; Singh, B.; Joon, R.K.; McDonald, A.; Royal, K.; Lisaius, M.C.; Lobell, D.B. Mapping Smallholder Wheat Yields and Sowing Dates Using Micro-Satellite Data. *Remote Sens.* **2016**, *8*, 860. [[CrossRef](#)]
29. Sun, J.; Di, L.; Sun, Z.; Shen, Y.; Lai, Z. County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. *Sensors* **2019**, *19*, 4363. [[CrossRef](#)] [[PubMed](#)]
30. Margono, B.A.; Turubanova, S.; Zhuravleva, I.; Potapov, P.; Tyukavina, A.; Baccini, A.; Goetz, S.; Hansen, M.C. Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. *Environ. Res. Lett.* **2012**, *7*, 034010. [[CrossRef](#)]
31. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171. [[CrossRef](#)]
32. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [[CrossRef](#)]
33. Clevers, J.; Gitelson, A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 344–351. [[CrossRef](#)]
34. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors* **2011**, *11*, 7063–7081. [[CrossRef](#)]
35. Kira, O.; Nguy-Robertson, A.L.; Arkebauer, T.J.; Linker, R.; Gitelson, A.A. Informative spectral bands for remote green LAI estimation in C3 and C4 crops. *Agric. For. Meteorol.* **2016**, 243–249. [[CrossRef](#)]
36. Sun, Y.; Qin, Q.; Ren, H.; Zhang, T.; Chen, S. Red-Edge Band Vegetation Indices for Leaf Area Index Estimation From Sentinel-2/MSI Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 826–840. [[CrossRef](#)]
37. Nguy-Robertson, A.; Gitelson, A.A.; Peng, Y.; Viña, A.; Arkebauer, T.; Rundquist, D. Green Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to Achieve Maximal Sensitivity. *Agron. J.* **2012**, *104*, 1336–1347. [[CrossRef](#)]
38. USDA ERS. Soybeans & Oil Crops. Available online: <https://www.ers.usda.gov/topics/crops/soybeans-oil-crops/> (accessed on 14 May 2020).
39. USDA ERS. Oil Crops Sector at a Glance. Available online: <https://www.ers.usda.gov/topics/crops/soybeans-oil-crops/oil-crops-sector-at-a-glance/> (accessed on 29 March 2020).
40. National Agricultural Statistics Service United States Summary and State Data 2012 Census Agric. Available online: <https://www.nass.usda.gov/Publications/AgCensus/2012/> (accessed on 2 May 2014).

41. NASS. Quick Stats|Ag Data Commons. Available online: <https://data.nal.usda.gov/dataset/nass-quick-stats> (accessed on 5 September 2020).
42. Abatzoglou, J.T. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Clim.* **2011**, *33*, 121–131. [[CrossRef](#)]
43. Fulton, J.; Hawkins, E.; Taylor, R.; Franzen, A.; Shannon, D.; Clay, D.; Kitchen, N. Yield Monitoring and Mapping. *Precis. Agric. Basics* **2018**, 63–77. [[CrossRef](#)]
44. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
45. Vogelmann, J.E.; Gallant, A.L.; Shi, H.; Zhu, Z. Perspectives on monitoring gradual change across the continuity of Landsat sensors using time-series data. *Remote Sens. Environ.* **2016**, *185*, 258–270. [[CrossRef](#)]
46. Gitelson, A.A.; Viña, A.; Arkebauer, T.J.; Rundquist, D.; Keydan, G.; Leavitt, B. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
47. Sola, I.; Álvarez-Mozos, J.; González-Audicana, M. Inter-comparison of atmospheric correction methods on Sentinel-2 images applied to croplands. *Int. Geosci. Remote Sens. Symp.* **2018**, 5940–5943. [[CrossRef](#)]
48. Rumora, L.; Miler, M.; Medak, D. Impact of Various Atmospheric Corrections on Sentinel-2 Land Cover Classification Accuracy Using Machine Learning Classifiers. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 277. [[CrossRef](#)]
49. Sola, I.; García-Martín, A.; Sardonís-Pozo, L.; Álvarez-Mozos, J.; Pérez-Cabello, F.; González-Audicana, M.; Llovería, R.M. Assessment of atmospheric correction methods for Sentinel-2 images in Mediterranean landscapes. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 63–76. [[CrossRef](#)]
50. Friedman, J.H. Recent Advances in Predictive (Machine) Learning. *J. Classif.* **2006**, *23*, 175–197. [[CrossRef](#)]
51. Coluzzi, R.; Imbrenda, V.; Lanfredi, M.; Simoniello, T. A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses. *Remote Sens. Environ.* **2018**, *217*, 426–443. [[CrossRef](#)]
52. Lobell, D.B.; Di Di Tommaso, S.; You, C.; Djima, I.Y.; Burke, M.; Kilic, T. Sight for Sorghums: Comparisons of Satellite- and Ground-Based Sorghum Yield Estimates in Mali. *Remote Sens.* **2019**, *12*, 100. [[CrossRef](#)]
53. Waldner, F.; Horan, H.; Chen, Y.; Hochman, Z. High temporal resolution of leaf area data improves empirical estimation of grain yield. *Sci. Rep.* **2019**, *9*, 1–14. [[CrossRef](#)]
54. Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-learn. *GetMobile Mob. Comput. Commun.* **2015**, *19*, 29–33. [[CrossRef](#)]
55. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [[CrossRef](#)]
56. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
57. Badgley, G.; Field, C.B.; Berry, J.A. Canopy near-infrared reflectance and terrestrial photosynthesis. *Sci. Adv.* **2017**, *3*, 1602244. [[CrossRef](#)]
58. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [[CrossRef](#)]
59. Pasqualotto, N.; Delegido, J.; Van Wittenberghe, S.; Rinaldi, M.; Moreno, J. Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). *Sensors* **2019**, *19*, 904. [[CrossRef](#)] [[PubMed](#)]
60. Dash, J.; Curran, P.J. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413. [[CrossRef](#)]
61. Daughtry, C. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. [[CrossRef](#)]
62. Gitelson, A.; Merzlyak, M.N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252. [[CrossRef](#)]
63. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In Proceedings of the Third Earth Resources Technology Satellite Symposium, Washington, DC, USA, 10–14 December 1974; pp. 309–317.
64. Boryan, C.; Yang, Z.; Mueller, R.; Craig, M. Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.* **2011**, *26*, 341–358. [[CrossRef](#)]
65. Holzworth, D.P.; Huth, N.I.; Devoil, P.G.; Zurcher, E.J.; Herrmann, N.I.; McLean, G.; Chenu, K.; Van Oosterom, E.J.; Snow, V.; Murphy, C.; et al. APSIM—Evolution towards a new generation of agricultural systems simulation. *Environ. Model. Softw.* **2014**, *62*, 327–350. [[CrossRef](#)]
66. Lobell, D.B.; Azzari, G. Satellite detection of rising maize yield heterogeneity in the U.S. Midwest. *Environ. Res. Lett.* **2017**, *12*, 014014. [[CrossRef](#)]

67. Jin, Z.; Ainsworth, E.A.; Leakey, A.D.B.; Lobell, D.B. Increasing drought and diminishing benefits of elevated carbon dioxide for soybean yields across the US Midwest. *Glob. Chang. Biol.* **2017**, *24*, 522–e533. [[CrossRef](#)] [[PubMed](#)]
68. Srinivasan, V.; Kumar, P.; Long, S.P. Decreasing, not increasing, leaf area will raise crop yields under global atmospheric change. *Glob. Chang. Biol.* **2016**, *23*, 1626–1635. [[CrossRef](#)] [[PubMed](#)]
69. Mourtzinis, S.; Specht, J.E.; Lindsey, L.E.; Wiebold, W.J.; Ross, J.; Nafziger, E.D.; Kandel, H.J.; Mueller, N.; DeVillez, P.L.; Arriaga, F.J.; et al. Climate-induced reduction in US-wide soybean yields underpinned by region- and in-season-specific responses. *Nat. Plants* **2015**, *1*, 14026. [[CrossRef](#)] [[PubMed](#)]
70. Tannura, M.A.; Irwin, S.H.; Good, D.L. Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt. *SSRN Electron. J.* **2008**. [[CrossRef](#)]
71. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritschi, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [[CrossRef](#)]
72. Guyot, G.; Baret, F.; Jacquemoud, S. Imaging spectroscopy for vegetation studies. *Imaging Spectrosc.* **1992**, 145–165.
73. Peng, Y.; Nguy-Robertson, A.; Linker, R.; Gitelson, A.A. Assessment of Canopy Chlorophyll Content Retrieval in Maize and Soybean: Implications of Hysteresis on the Development of Generic Algorithms. *Remote Sens.* **2017**, *9*, 226. [[CrossRef](#)]
74. Whitcraft, A.K.; Vermote, E.F.; Becker-Reshef, I.; Justice, C. Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sens. Environ.* **2015**, *156*, 438–447. [[CrossRef](#)]
75. Lobell, D.B.; Azzari, G.; Burke, M.; Gourlay, S.; Jin, Z.; Kilic, T.; Murray, S. Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis. *Am. J. Agric. Econ.* **2019**, *102*, 202–219. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).