

Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model

Matteo G. Ziliani^{a,*}, Muhammad U. Altaf^a, Bruno Aragon^b, Rasmus Houborg^c,
Trenton E. Franz^d, Yang Lu^e, Justin Sheffield^e, Ibrahim Hoteit^f, Matthew F. McCabe^a

^a Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology, Kingdom of Saudi Arabia

^b Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

^c Planet, San Francisco, California, USA

^d School of Natural Resources, University of Nebraska-Lincoln, USA

^e School of Geography and Environmental Sciences, University of Southampton, United Kingdom

^f Physical Science and Engineering (PSE), King Abdullah University of Science and Technology, Kingdom of Saudi Arabia

ARTICLE INFO

Keywords:

Crop yield prediction
Crop modeling
CubeSat
LAI
APSIM
Data assimilation

ABSTRACT

Accurate early season predictions of crop yield at the within-field scale can be used to address a range of crop production, management, and precision agricultural challenges. While the remote sensing of within-field insights has been a research goal for many years, it is only recently that observations with the required spatio-temporal resolutions, together with efficient assimilation methods to integrate these into modeling frameworks, have become available to advance yield prediction efforts. Here we explore a yield prediction approach that combines daily high-resolution CubeSat imagery with the APSIM crop model. The approach employs APSIM to train a linear regression that relates simulated yield to simulated leaf area index (LAI). That relationship is then used to identify the optimal regression date at which the LAI provides the best prediction of yield: in this case, approximately 14 weeks prior to harvest. Instead of applying the regression on satellite imagery that is coincident, or closest to, the regression date, our method implements a particle filter that integrates CubeSat-based LAI into APSIM to provide end-of-season high-resolution (3 m) yield maps weeks before the optimal regression date. The approach is demonstrated on a rainfed maize field located in Nebraska, USA, where suitable collections of both imagery and in-situ data were available for assessment. The procedure does not require in-field data to calibrate the regression model, with results showing that even with a single assimilation step, it is possible to provide yield estimates with good accuracy up to 21 days before the optimal regression date. Yield spatial variability was reproduced reasonably well, with a strong correlation to independently collected measurements ($R^2 = 0.73$ and $rRMSE = 12\%$). When the field averaged yield was compared, our approach reduced yield prediction error from 1 Mg/ha (control case based on a calibrated APSIM model), to 0.5 Mg/ha (using satellite imagery alone), and then to 0.2 Mg/ha (results with assimilation up to three weeks prior to the optimal regression date). Such a capacity to provide spatially explicit yield predictions early in the season has considerable potential to enhance digital agricultural goals and improve end-of-season yield predictions.

1. Introduction

The delivery of timely crop monitoring and accurate crop yield estimates, at both the field and regional scales, remains a precision agriculture and food security goal (Fritz et al., 2019). Accurate pre-harvest predictions of crop yield at the field scale would help farmers to tailor site-specific management decisions regarding crop inputs (i.e. timing and amount of fertilizer and irrigation), and also to forecast profit based

on spatially explicit yield estimates (Tewes et al., 2020). Put simply, by projecting end-of-season yield information as early as possible, farmers can take corrective measures to increase their yield. Policy makers and governments also benefit from yield forecasts, which can serve as benchmarks or provide information to drive food security efforts and reforms (Resnick, 2020).

Modeling of plant growth dynamics via crop models, in combination with remote sensing observations of agricultural variables through time,

* Corresponding author.

E-mail address: matteo.ziliani@kaust.edu.sa (M.G. Ziliani).

<https://doi.org/10.1016/j.agrformet.2021.108736>

Received 2 July 2021; Received in revised form 7 November 2021; Accepted 16 November 2021

Available online 29 November 2021

0168-1923/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

are increasingly being used to predict crop yield. Models such as the Agricultural Production System Simulator (APSIM) (Holzworth et al., 2014) or the Decision Support System for Agrotechnology Transfer (DSSAT) (Jones et al., 2003) have been applied to describe the temporal development of key crop processes throughout the growing season, providing insight into photosynthesis, soil dynamics, biomass partitioning, and yield accumulation, using climate, soil, and field management information (Hoogenboom, 2000). Although crop models are employed for farming operations at a range of scales (Tewes et al., 2020), they were originally developed for point-based applications that neglect spatial variation in yield forecast.

To overcome this limitation, considerable effort has been directed towards acquiring spatially distributed information on crop status and yield using satellite-based remote sensing platforms (Lobell, 2013). However, fine-scale yield predictions have been hindered by the limited availability of high spatiotemporal resolution images (Huang et al., 2019). Indeed, one of the major drawbacks of satellite platforms is their coarse spatial and temporal resolutions, which cannot resolve the finer spatial scale patterns that drive between and within-field variability (Jain et al., 2017). High spatial and temporal resolution information is needed for precision agriculture applications (McCabe et al., 2017b) to deliver critical farm management strategies such as nitrogen application and irrigation scheduling (Zhang et al., 2015; Aragon et al., 2018, 2021). Although previous studies have demonstrated the utility of using remote sensing for yield mapping operations (Lobell et al., 2015; Jain et al., 2019), field to within-field scale yield mapping applications remain underexplored. In large part, this is due to a paucity of ground-based observations that can be used to validate the derived remote sensing products, as well as the costs associated with obtaining and processing high spatiotemporal (and often commercial) resolution imagery (Jin et al., 2018).

Recent approaches to convert satellite images into yield maps have used crop models to simulate realistic field-level yield data, which are then used to train linear regressions that translate observed satellite vegetation indices. The trained regressions have then been applied to satellite imagery to produce spatially varying yield maps at both regional (Sibley et al., 2014; Lobell et al., 2015) and field scales (Burke and Lobell, 2017; Jeffries et al., 2019). These approaches are of interest both for their scalability and because they do not generally require ground-based measurements. Nevertheless, they can be limited by satellite revisit times, as in the case of Landsat and (to a lesser degree) Sentinel-2, or coarse spatial resolution (e.g. MODIS). One way to enhance yield prediction capabilities at within-field scales is to blend the predictive capacity of crop models with the spatial information retrievable from space. Data assimilation techniques facilitate the combination of observations with model predictions to obtain the best possible estimates of the underlying processes (Vettra-Carvalho et al., 2018). Amongst the variety of assimilation approaches, sequential Bayesian estimation methods employ a probabilistic framework to constrain model predictions with observations as they become available (van Velzen et al., 2016; Hoteit et al., 2018). Amongst the most standard of such methods is the Ensemble Kalman Filter (EnKF), which is based on linear and Gaussian assumptions (Evensen, 2003). Progressive increases in computational resources are enabling the use of more sophisticated methods that can deal with non-Gaussian distributions, such as the particle filter (PF). In contrast to EnKF, PF methods provide information of more than the first two statistical moments (mean and covariance) (van Leeuwen, 2010), and thus offers practical solutions for highly nonlinear dynamical systems.

Over the last decade or more, a variety of measurements have been assimilated into crop models, including remote-sensing based metrics of soil moisture (De Wit and Van Diepen, 2007; Bolten et al., 2009; Ines et al., 2013; Zhuo et al., 2019), canopy cover (Silvestro et al., 2017; Jin et al., 2020; Lu et al., 2021), and LAI (Maas, 1988; Fang et al., 2008; Huang et al., 2015; Mokhtari et al., 2018; Li et al., 2018). The latter is a particularly informative variable to assimilate due to its ability to

diagnose crop status, serve as an indicator of leaf abundance and phenological stage, and as a metric to evaluate different management approaches (Huang et al., 2019). While many of these studies have illustrated the potential that satellite information has in advancing yield predictions, the spatial and temporal resolutions have often been unable to fully resolve the within-field variabilities needed to drive precision agriculture applications, especially for small scale farmers (i.e., plots smaller than 2 ha) (Jain et al., 2017). For instance, even the 5-day revisit time of Sentinel-2 cannot completely negate the observation gaps due to frequent overcast conditions typical of subtropical and temperate environments (Beuchle et al., 2011). Recent advances in Earth observation platforms, specifically the development and proliferation of micro-satellites (referred to herein as CubeSats), have enabled previously unattainable capabilities of high-spatial resolution with daily frequency (McCabe et al., 2017a). As such, integrating daily CubeSat products into crop models presents an opportunity to enhance the predictive ability of these models and to drive digital agricultural advances in crop modeling (Peng et al., 2020), forecasting (Basso et al., 2013), and yield estimation (Ziliani et al., 2021).

Here we develop an approach to retrieve early season prediction of within-field crop yield by combining the spatiotemporal advantage of CubeSat imagery with the predictive capability of a crop model. For this purpose, we use APSIM crop model simulations to train a linear regression that relates model simulated and field-level yield to simulated LAI. Using this relationship, we then identify the optimal date during the growing season to predict yield (referred to as “regression date”). Importantly, the training of the regression does not require measured data, following a similar strategy employed by Sibley et al. (2014) and Lobell et al. (2015). Advancing upon these previous efforts, we incorporate CubeSat-based LAI maps (Johansen et al., 2021) into APSIM using a particle filter. The assimilation of CubeSat imagery enhances the APSIM predictive capacity of LAI by several weeks, delivering information that can then be used to predict yield via the previously trained linear regression. Comparison against spatially varying ground-based yield measurements over a rainfed maize field indicates that the procedure enables yield prediction ahead of time and without the need to wait for date-coincident satellite overpasses, advancing upon studies that may apply a regression-based solely on satellite imagery. The study is the first remote sensing and model-based approach to estimate yield at such high spatial resolution (3 m) using daily CubeSat data and without relying on ground-measured yield data.

2. Materials and methods

2.1. The study site

The experimental field used in this study is part of the Eastern Nebraska Research and Extension Center (ENREC) at the University of Nebraska-Lincoln (UNL). The region faces cold winters, hot summers, and overall humid conditions (Sharma and Irmak, 2012). Importantly, it has an extensive history of agronomic data collections and analysis (Baldocchi et al., 2001), with meteorological forcing, soil conditions data, and management information available to drive model simulations. The meteorological data were recorded from the Ithaca 3E weather station (<https://cropwatch.unl.edu/tags/nebraska-mesonet>) and consisted of daily measurements of air temperature ($T_{\min}$ and $T_{\max}$, in °C), precipitation (P , in mm), and solar radiation (G , $W\ m^{-2}$). A 65 ha rainfed field (US-Ne3) that operates under a no-till management policy was used to explore the developed approach (Fig. 1). The site alternates between maize (*Zea mays* L.) and soybeans (*Glycine max*) and has fertilizer applied before each planting cycle. Sowing typically occurs between late April and the middle of May, while harvesting is performed in late October/beginning of November. In this study, 10 maize-growing seasons (every alternating year from 2001 to 2019) were considered for the analysis.

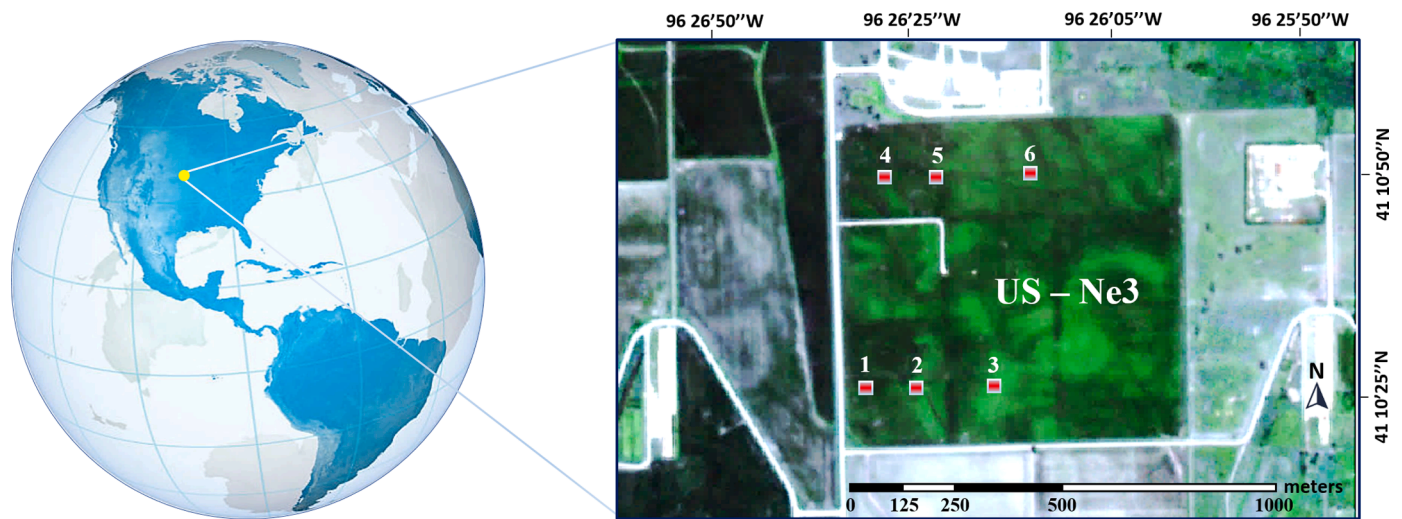


Fig. 1. Layout of the rainfed maize field (US-Ne3) located in Nebraska, USA (left, yellow dot). Each red box within the field (right) represents an Intensive Measurement Zone (IMZ) where destructive crop sampling occurs throughout the season for the collection of LAI.

Soil information for the study site, retrieved from the United States Web Soil Survey agency (<http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm>), reveals a high percentage of silt loam and silty clay loam (Foolad et al., 2017), which is typical of eastern Nebraska. Additional information on the site characteristics and conditions are discussed in Suyker and Verma (2008) and Verma et al. (2005). During each growing season, in-situ crop sampling is performed at six Intensive Measurement Zones (IMZs) within the field (Fig. 1, right) every 10 to 14 days. Samples of LAI (m^2/m^2) were collected throughout the season using an area meter (Model LI-3100, LI-Cor, Lincoln, NE) (Nguy-Robertson et al., 2015), while spatially varying yield was collected at the end of the growth cycle using harvesting tractors equipped with GPS technologies (see Franz et al. (2020) for additional details on yield data).

2.2. APSIM model

Continuous development of crop models over the last three decades has contributed to a range of sophisticated crop modeling systems. Amongst these, APSIM is one of the most widely employed crop models (McCown et al., 1996; Keating et al., 2003). The model is open-source, allowing access to source codes that incentivise community-wide module development for possible inclusion in official versions (Holzworth et al., 2014). It should be noted that even though our study focuses on yield predictions using APSIM, the approach is generic, and other models could be used with minor framework modifications. For example, the DSSAT (Jones et al., 2003) and the WOFOST (World Food Studies) model (Van Diepen et al., 1989), have been shown to offer comparable performance in predicting crop growth and yield.

APSIM is a modeling framework designed to simulate crop growth and development through fundamental processes such as phenology, leaf development, underlying soil dynamics, biomass production, and partitioning, which are evaluated on a daily time step (Wang et al., 2002). The model was developed from a combination of approaches deriving from the CERES-Maize model (Keating et al., 1991; Carberry et al., 1989), with a major difference being the routines that kill the crops in response to severe water deficit during the early- to mid-vegetative stage (Carberry and Abrecht, 1991). Daily temperature (daily min/max), precipitation (daily sum), and daily sum of solar

radiation are required to drive the model. Management practices are defined by sowing variables (e.g. crop type, sowing date, sowing depth, density) and application strategies (e.g. tillage, fertilizer application, and irrigation). APSIM divides crop phenology into sub-phases. The duration of each phase is primarily determined by temperature and photoperiod and can also be affected by water and nitrogen stresses. When water and nitrogen-limiting conditions affect the crop, leaf area expansion is considerably reduced, as is the ability to intercept radiation (Vos et al., 2005; Lemaire et al., 2008; Massignam et al., 2009). In most cases, temperature drives canopy development, with the final daily estimate of LAI derived as the product of plant density and leaf area per plant (Soufizadeh et al., 2018). The product of grain size and grain number is used to simulate grain yield. Grain number is estimated from the average daily growth rate per plant between tassel initiation and the start of grain filling and the potential grain number per ear, using the function developed by Edmeades and Daynard (1979). A detailed description of the APSIM model and all its components can be found in the literature (Keating et al., 2003; Holzworth et al., 2014, 2018).

2.3. Planet CubeSat and LAI estimation

The study employs a gap-free LAI product that leverages Planet multi-constellation CubeSat imagery (www.planet.com), publicly accessible multi-spectral satellites, and the CubeSat Enabled Spatio-Temporal Enhancement Method (CESTEM; (Houborg and McCabe, 2018a) to provide a high spatiotemporal fusion dataset (i.e., Planet Fusion) at daily and at 3 m resolution starting from DOY 91 to DOY 334 (April 1 to November 30, 2019). CESTEM is a data fusion model that cleans, atmospherically corrects, and normalizes CubeSat imagery to a target product by constructing relationships between the CubeSat bands and the reference imagery. The target imagery used for the CESTEM normalization came from the FORCE methodology that combines Landsat 8 and Sentinel-2 (A/B) to produce analysis ready surface reflectance data (Frantz, 2019). Additionally, daily Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance (NBAR) data from the Moderate Resolution Imaging Spectroradiometer (MODIS) was used for calibration and gap-filling purposes in synergy with multi-constellation CubeSat, Landsat 8, and Sentinel-2 imagery. The resulting dataset after applying the CESTEM framework is a Level 3

surface reflectance (SR) product. Further details on the CESTEM methodology for radiometric normalization can be found in [Houborg and McCabe \(2018b\)](#). To produce LAI, CESTEM is used again, with modeled LAI as the target using the Level 3 SR product as an explanatory variable ([Johansen et al., 2021](#)). The reference LAI dataset is modeled via a random forest approach that combines the Sentinel-2 BRDF adjusted Surface Reflectance product (30 m) from the FORCE methodology and derived vegetation indices together with forward runs of the PROSAIL canopy reflectance model ([Jacquemoud et al., 2009](#)). The result is a Level 4 LAI product, which has the same spatial and temporal resolution of the Level 3 SR data. A more detailed description of the CESTEM-LAI methodology is presented in [Houborg and McCabe \(2018a,c\)](#). It should be noted that the Level 4 LAI product used in this study was not calibrated against in-situ data.

2.4. Yield estimation with regression

Empirical models can be directly applied to satellite-derived variables to generate predictions of end-of-season yield. The regression approach employed here uses APSIM to train a simple model that relates yield to a vegetation index ([Clevers, 1997](#)). To do so, a large number of crop model simulations are performed to simulate realistic field-level yield data, which are then used to train a linear regression that translates an observed satellite vegetation index into yield. The simulations span a realistic range of weather, soil, and management conditions, all of which are based on previous literature for the study region ([Lobell et al., 2015](#); [Jeffries et al., 2019](#)), as well as information gathered from ENREC ([Section 2.1](#)). Simulations were produced for alternating years between 2001 and 2018 (recall that US-Ne3 is planted with maize every other year), providing daily values of model outputs such as LAI and yield. A total of 7776 APSIM simulations were run based on all possible combinations of model inputs and parameters, which are listed in [Table 1](#). Daily weather data from the in-situ station at US-Ne3, which included maximum and minimum temperatures, solar radiation, and precipitation, were used to initialize the model. The simulated yield and the pseudo remote sensing observations (i.e. LAI) can then be combined to train a simple regression model:

Table 1

Description of APSIM model input parameters and variables selected to train the regression model for the rainfed maize field in this study.

Factors	Values	Units	Note
Years	2001–2018		Daily weather files obtained from ENREC
Fertilizer rates	90,200	Kg of urea N/ha	Application 1 week before planting
Sowing dates	April 24, May 1, May 15		10th, 50th, and 90th percentile of observed sow dates in the region, from Ziliani et al. (2021)
Sowing densities	4.5, 6, 7.5, 9	Plants per m ²	
Cultivar parameter 1 (tt_emerg_to_endjuv)	160, 185, 220	Degree days	thermal time from emergence till end juvenile phase, based on Ziliani et al. (2021)
Cultivar parameter 2 (tt_flower_to_maturity)	630, 730, 780	Degree days	thermal time from silking to physiological maturity, based on Jeffries et al. (2019)
Soil types	Tomek silt loam, Filbert silt loam, Yutan silty clay, Fillmore silt loam		Estimated properties of four main soil profiles in the region, based on https://websoilsurvey.sc.egov.usda.gov/

$$yield = a_i + b_i * VI_i, \quad (1)$$

where a and b are the regression coefficients to be estimated, and VI is the LAI map used to predict yield at a specific date t . An important consideration of the approach is that it does not rely on any field measurements of yield, but only on the crop model's ability to relate variables (such as LAI) to model outputs (such as yield). As with the above choice of statistical model, it would be possible to make many other decisions on the model and the number of predictors to calculate the yield. In our case, these were kept to a minimum, with the aim of developing a simple and easy-to-use approach that can be employed with the least amount of data ([Baez-Gonzalez et al., 2005](#); [Clevers, 1997](#); [Sibley et al., 2014](#)).

A separate regression is run for each day of the growing season, with the simulated vegetation index of that day used as a predictor and the end-of-season yield as a response. Besides providing the coefficient estimates, the training of the regression provides information about the model via the coefficient of determination (R^2), which can be used to identify which day of the season (i.e. regression date) is most suited to predict yield. This “regression date” is determined based on the day with the highest correlation between simulated VI and yield. Once the day with the highest correlation has been identified, the regression equation can be applied to actual measurements obtained from satellite data or, as our method proposes, to LAI forecasts obtained from a data assimilation approach (see [Section 2.5](#)). This step can be done on a pixel-by-pixel basis to generate within-field spatially explicit predictions.

As discussed above, regression approaches used in the literature tend to rely on the satellite data acquisition on – or closest to – the date of maximum correlation i.e. the regression can only be applied once the satellite image becomes available. To bypass this limitation, we implement a particle filter to assimilate Planet Fusion CubeSat imagery into APSIM and then forecast the LAI map until the day of maximum correlation ([Section 2.5](#)). The end-of-season yield is then calculated by applying the previously trained regression on LAI forecasts, hence allowing the prediction of yield ahead of time, even before the satellite image becomes available. Also, as opposed to previous studies where the regression was calculated considering the weather for a particular year and then used to predict yield in the same year ([Sibley et al., 2014](#); [Lobell et al., 2015](#); [Jeffries et al., 2019](#)), here we utilize prior knowledge of climate conditions (i.e. data from 2001 to 2018) to extrapolate a regression that can be used to test the method in later years (i.e. 2019 was used for validation). Such an approach reflects operational considerations where future weather is generally unknown, or if available from forecasts, not necessarily reliable for the 5–6 months growth cycle typical of the Nebraskan region.

2.5. Data assimilation with a particle filter

The Particle filter (PF) approximates the probability distribution of the model state by a set of samples called particles. The samples are first generated from a given prior. An importance sampling is then applied to turn them into samples from the posterior, where each sample is weighted by its likelihood value based on incoming observations ([van Leeuwen, 2010](#); [Vetra-Carvalho et al., 2018](#)). When observations are not available, the PF integrates the particles forward in times for forecasting. This concept is referred to as sequential, or Bayesian, importance sampling.

Let x_i^k be each LAI pixel value of the system state integrated forward in time and k be the time of the current observation. The PF framework considers a stochastic state-space system comprised of a stochastic model, for each particle $i = 1 \dots N$ is given by:

$$x_i^k = \mathbf{M}(x_i^{k-1}) + \beta_i^k, \quad (2)$$

where $\mathbf{M}$ is the deterministic crop model APSIM, β_i^k is a noise term

assumed here Gaussian, with zero mean and covariance matrix $\mathbf{Q}$. Now let $\mathbf{x}^m$ be the mean of $\mathbf{N}$ particles. The weight w_i^k for each LAI pixel value x_i^k of the i th particle at time k of the satellite observation acquisition is therefore computed as:

$$w_i^k = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i^k - x_o^k)^2}{2\sigma^2}\right). \quad (3)$$

The observed LAI x_o^k (i.e. each pixel x of the satellite image at time k) with its associated standard deviation σ (which quantifies the observation uncertainty) represents the discrete observation at a specific time step. Here we assumed a Gaussian likelihood with independent observations. Subsequently, all the weights w_i^k are normalized by the sum of the weights of all particles as:

$$\text{norm_}w_i^k = \frac{w_i^k}{\sum_{i=1}^N w_i^k}. \quad (4)$$

The posterior probability distribution of the state is then described by the updated weights which allow the computation of an expected weight E of the LAI values x_i^k (referred to herein as rs_{PF}):

$$rs_{PF} = \sum_{i=1}^N x_i^k w_i^k. \quad (5)$$

Note that in Bayesian importance sampling, the state variables are not updated, only the weights. In practice, without applying any re-initialization of the model states in APSIM, the computed weights were applied to the APSIM outputs to obtain the mean LAI predictions. The expected weights are assumed to be valid from time k until a new observation becomes available and new weights are estimated, keeping memory of the previous assimilation time steps. To verify the capacity of the particle filter to predict LAI maps, multiple assimilation starting dates were tested to understand how long in advance it is possible to generate a reliable VI map that can be used to predict the yield.

To obtain an ensemble representing model uncertainty, a number of APSIM variables were selected to be perturbed. These variables were selected following previous studies (Machwitz et al., 2014; Lobell et al.,

2015) and based on their impact on the model outputs (Ziliani et al., 2021). For the generation of the ensemble, input variables related to soil conditions and management strategies were used. Specifically, 100 particles were created by running the model with varying combinations of sowing dates, different amounts of applied fertilizer, and soil conditions whose ranges of variation were set according to standard farming practices at the ENREC facilities (and consistent with those listed in Table 1). All other management factors remained constant, including maize variety parameters used to run the model (*Pioneer 3237*, which possesses similar phenological characteristics to most of the cultivars planted in the survey region; Ziliani et al., 2021). Weather conditions for the year 2019 were used to drive the simulations for generating the particles. As mentioned, only the meteorology for those years planted with maize (between 2001 and 2018) was used to develop the underlying regressions that are used for yield prediction. A schematic representation of the two-step (regression and particle filter) approach proposed here is outlined in Fig. 2.

2.6. Statistical evaluation

The accuracy of CubeSat-retrieved LAI was assessed against LAI field data collected at each of the six IMZs within the field. CubeSat-based yield predictions were evaluated against in situ spatially varying yield observations collected during harvest at the end of the 2019 growing season. Since the field-collected yield has a resolution of 10 m (see Franz et al. (2020) on yield data preparation), the predicted yield was upscaled (i.e. averaging) from 3 m to 10 m to enable a meaningful pixel-by-pixel comparison. Regression analysis was performed between predicted and measured gridded datasets. Four statistical metrics were used to evaluate the goodness of fit for each prediction, including the coefficient of determination (R^2), the root mean square error (RMSE) and its relative value (rRMSE), as well as the bias, with the metrics calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}, \quad (6)$$

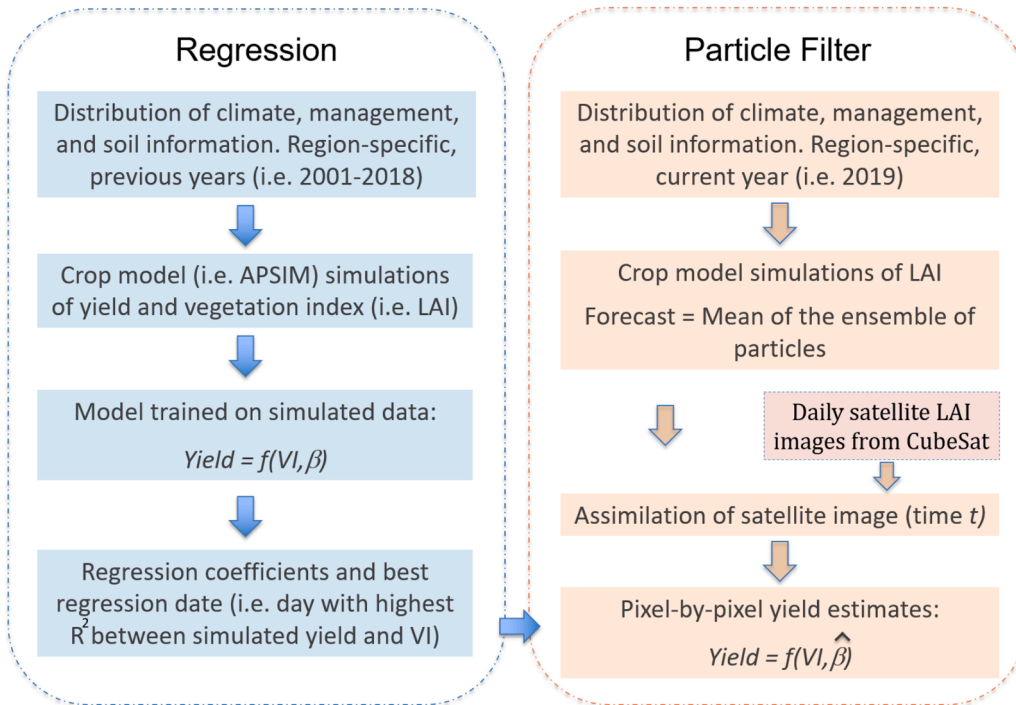


Fig. 2. Schematic of the approach for training a linear regression that translates an observed vegetation index (i.e. LAI) into yield (left side; regression) and the assimilation scheme to drive LAI predictions (right side; particle filter) which uses CubeSat generated LAI maps. Each box represents a specific step in the process.

$$rRMSE = RMSE/\bar{x}^* 100, \quad (7)$$

$$R^2 = \left(\frac{cov(x, y)}{\sigma_y \sigma_x} \right)^2, \quad (8)$$

$$bias = \frac{1}{n} \sum_{i=1}^n (y_i - x_i), \quad (9)$$

where n is the number of pixels in each dataset, $\bar{x}$ is the mean value of the observed values, σ is the standard deviation and $cov(x, y)$ is the covariance between the observed data x and the predicted y . Each metric provides useful information on the goodness of fit, in which R^2 depicts the proportion of variance between observed and predicted variables, RMSE and rRMSE illustrate the absolute and relative magnitude of the difference between prediction and observation, and the bias refers to the tendency of the prediction to under- or over-estimate the observation.

3. Results

3.1. Remotely sensed LAI

Achieving high-quality input data from satellite platforms is key for developing an efficient assimilation system. Moreover, to provide within-field estimates of crop yield needed for precision agricultural applications, high spatial resolution images are necessary. Fig. 3 shows a daily LAI sequence of 3 m CubeSat LAI retrievals that highlights the spatial heterogeneity of the rainfed US-Ne3 maize field, with the time series corresponding to a green-up period from DOY 173 to 190. During this 18-day window, the field undergoes rapid development, characterized by a doubling of leaf area, thereby increasing the field average LAI from approximately 2 to 4 m²/m². The within-field variability

becomes apparent after DOY 180, where some of the darker patches fail to increase in the same way as the central parts of the field, which may lead to potential yield losses.

The accuracy of the CubeSat LAI retrievals was assessed by comparison against in situ LAI measurements at the IMZs within the field. Fig. 4 shows the daily CubeSat LAI retrieval at the coordinates (latitude and longitude) corresponding to the IMZ 1 from DOY 100 till DOY 320. Overall, the LAI trend is well reproduced by the satellite observations, with the majority of CubeSat retrieved values falling within the LAI in situ ranges and close to their average values (blue dots in Fig. 4). Interestingly, the rapid development of LAI occurring between the V6 and V15 vegetative stages is accurately reproduced by the CubeSat LAI. Such an increasing trend reflects the crop reaching the end of its vegetative stage, which starts approximately two weeks before flowering (around DOY 200 in our case) (Ritchie and Hanway, 1989; Abendroth et al., 2011).

During this rapid growth phase, the stalk undergoes substantial development, leading leaf area to increase dramatically (Bitzer et al., 2000; Ziliani et al., 2018). During the first part of the reproductive stage (from R1 till R4), the LAI response remains mostly unchanged (with values around 4.5 m²/m², on average). After this, the crop reaches physiological maturity (around R6) by accumulating above-ground biomass while reducing leaf area. The latter subsequently decreases rapidly until almost zero. A related study focused on assessing the performance of the CubeSat LAI retrieval at the same study site (Johansen et al., 2021), reported an R^2 of 0.92 between the in situ and the CubeSat LAI, with reasonably low RMSE and rRMSE (0.4 m²/m² and 13%, respectively). Since the CubeSat retrievals in the other intensive measurement zones exhibit similarities to the results of Fig. 4, they are presented in Appendix 1 for the sake of brevity. Further information on the accuracy of the CubeSat-based LAI retrievals, along with a thorough

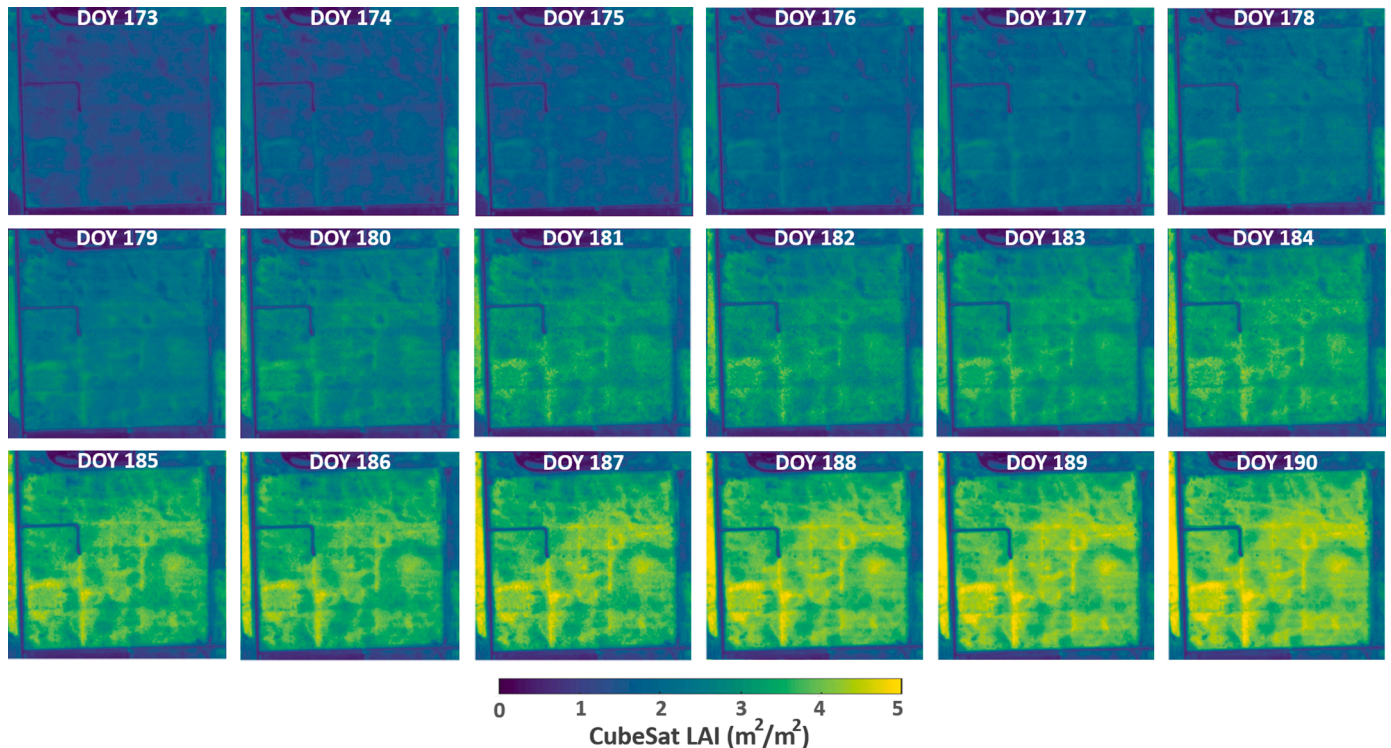


Fig. 3. Daily sequence of 3 m resolution CubeSat LAI retrievals of the field US-Ne3 from DOY 173 to 190. The imagery captures the vegetative phase at US-Ne3 and the progressive increase in spatially varying LAI values within the field.

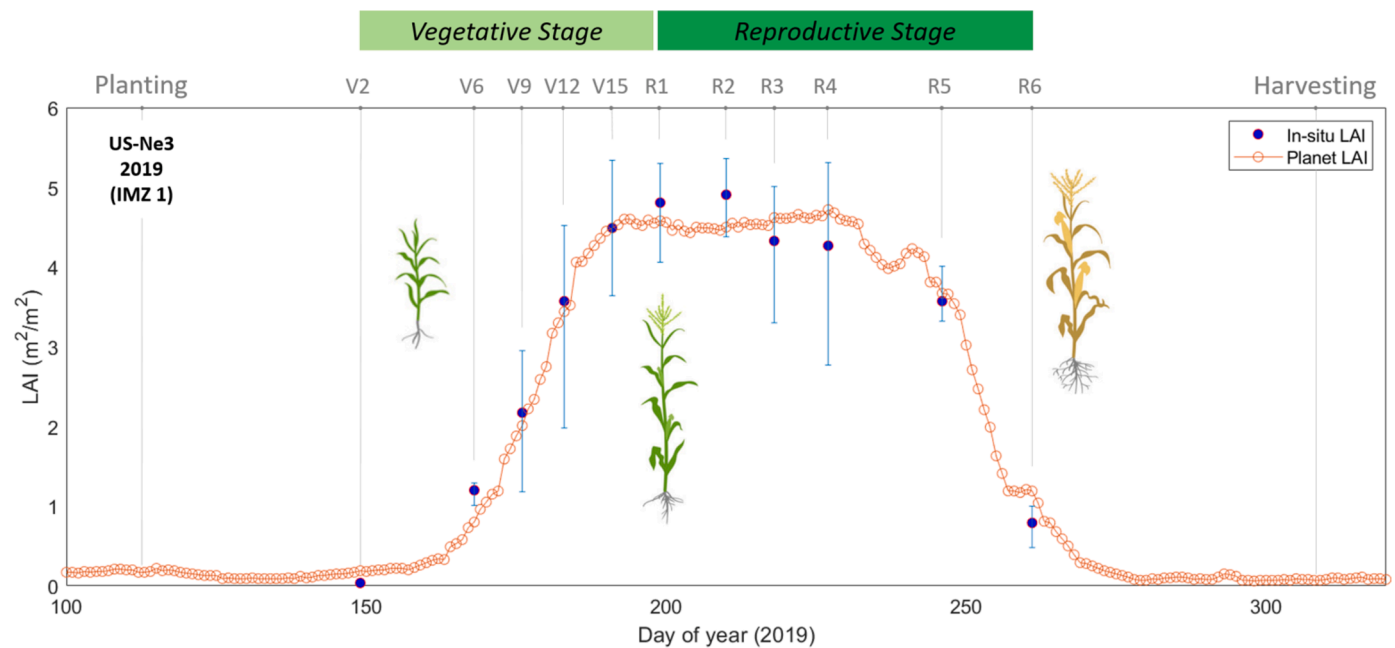


Fig. 4. Planet CubeSat LAI time-series (orange circles) at the first intensive measurement zone (IMZ1) of US-Ne3 against the LAI in situ measurements performed at the indicated vegetative and reproductive stages (blue circles). The range of the field measurements is depicted by blue bars.

explanation of LAI dynamics during the 2019 growing season can be found in [Johansen et al. \(2021\)](#).

3.2. Yield estimation through regression

The regression approach does not rely on ground calibration data, but instead uses crop model simulations as a training dataset to calibrate a regression model. To do this, crop simulations were run with the APSIM-Maize model for each of the maize-planted study years (i.e. between 2001 and 2018) independently, using the daily weather measurements from local station data and different combinations of inputs and parameters as listed in [Table 1](#). Each model run generates daily simulated maize yield and simulated LAI. These results were then used to estimate a regression model ([Eq. \(1\)](#)) to predict yields from the LAI (see [Section 2.2, Section 2.3](#)). The coefficient of determination (R^2) for the regression between simulated maize yield and simulated vegetation index (LAI) was computed for each day of each simulated model run. To develop (and select) the regression that can be used to predict yield for later years (i.e. in 2019), an average daily value of R^2 was then taken as representative from all the simulated model runs.

[Fig. 5a](#) (left side) shows the trend of R^2 during a typical May–October cropping season. From this, it is possible to identify that the LAI at DOY 212 (July 31) is the best predictor of future crop yield. In general, the correlation between LAI and yield starts to be apparent from the middle of June, when R^2 sharply increases until the end of July and then decreases until the end of October. The date of maximum correlation (DOY 212), referred to as the “regression date”, can be used to predict yield. A scatter plot in [Fig. 5b](#) depicts the ensemble of relationships on DOY 212 between simulated LAI and yield generated from all the model runs (using all possible combinations of inputs from [Table 1](#)), and presents a correlation of 0.72. The regression equation for DOY 212 can be determined, providing the coefficients a and b . In practice, this regression can be applied to all the pixels in a satellite-derived vegetation index map to retrieve an end-of-season yield map.

3.3. Assimilating LAI using a particle filter

Although every day between thirty days and one day prior to the regression date was examined, here we focus on illustrating the results for just four different satellite acquisition dates ([Fig. 6](#)) that include (a)

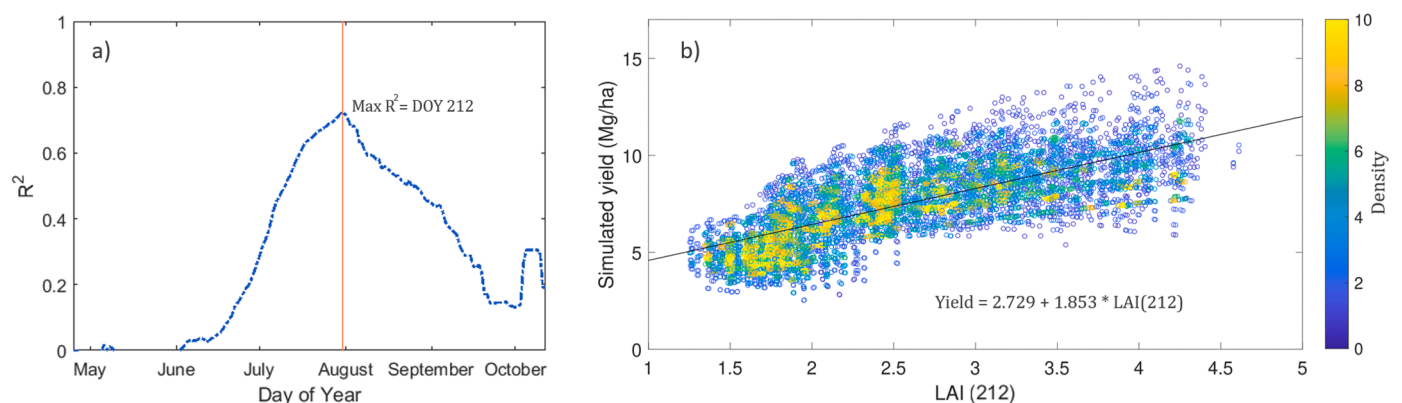


Fig. 5. (Left box) Trend of coefficient of determination derived from all APSIM simulation runs (years 2001–2018 and input combinations from [Table 1](#)) and used to detect the date where the VI is the best predictor of final yield. The date corresponding to the highest value of R^2 is used to predict crop yield by applying the regression equation (right box), which relates yield and LAI to that particular day.

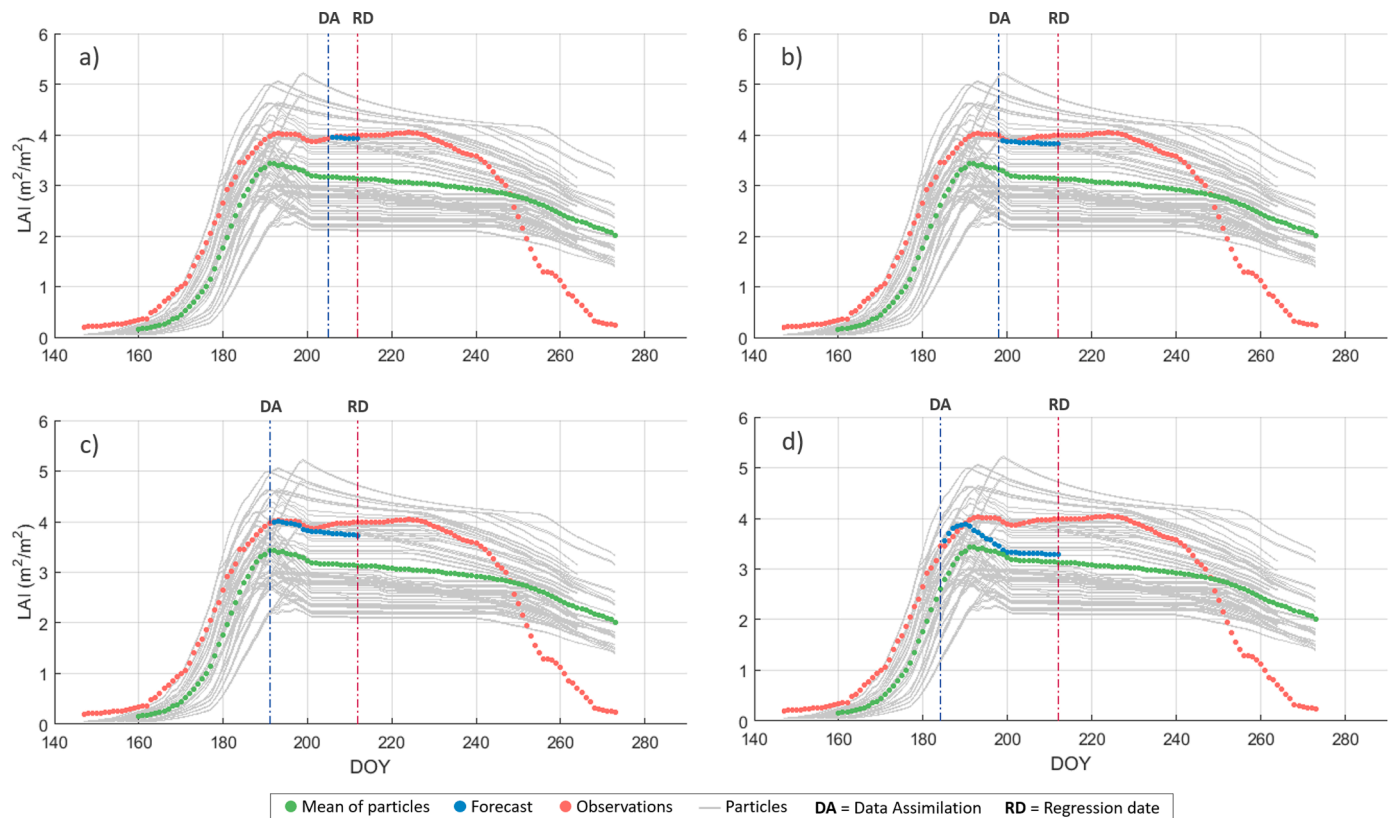


Fig. 6. LAI simulations (particles, gray lines) produced by the APSIM model. The simulations can be seen to contain the daily satellite observations (orange dots). The green dots represent the mean of the various particles, while the blue dots represent the daily LAI forecast from the day of assimilation until the regression date on DOY 212. Assimilation (a) one, (b) two, (c) three, and (d) four weeks before the regression date are evaluated.

one, (b) two, (c) three, (d) four weeks before the regression date (DOY 212). The acquisition day of the CubeSat data (which corresponds to the assimilation day) is displayed as a blue vertical line, while the regression

date is depicted as a red vertical line. The results show that after the assimilation step all the weighted mean LAI predictions (blue dots) clearly shift towards the mean of the observations (orange dots) and that

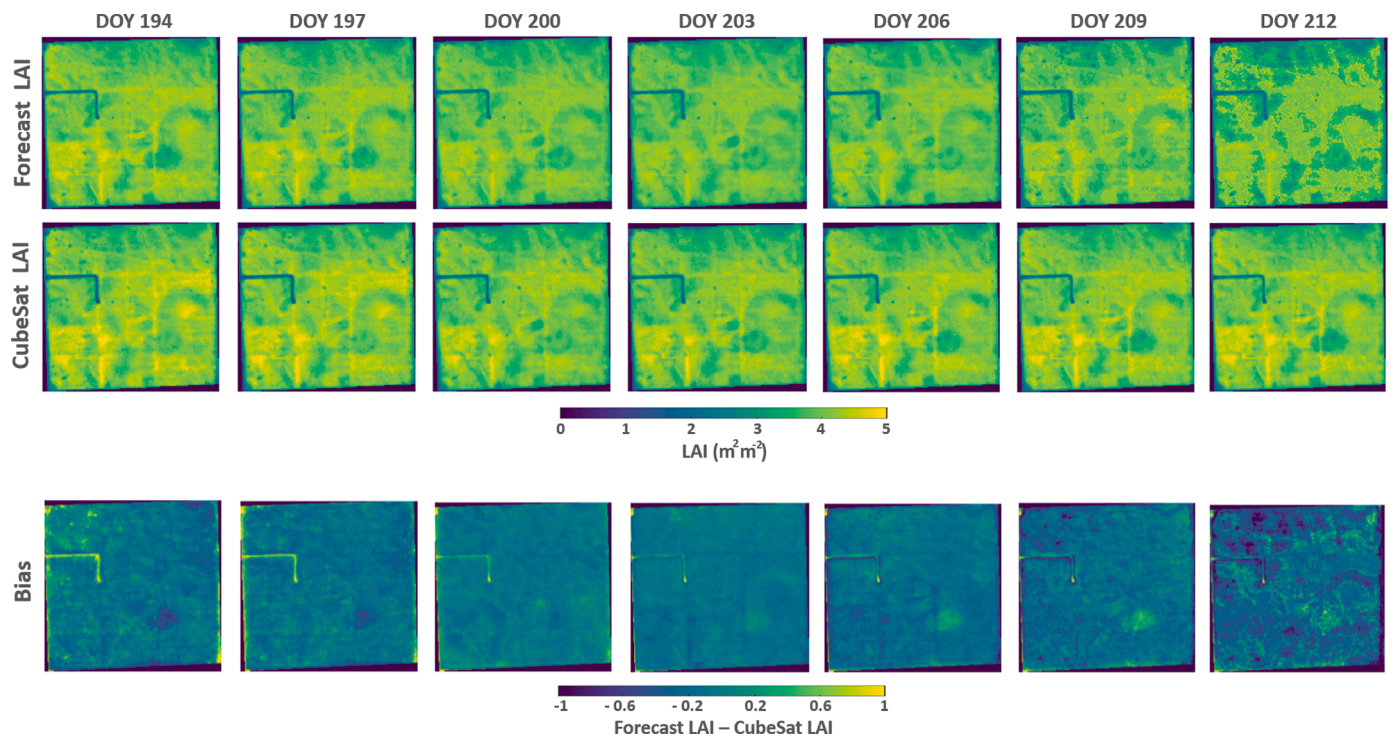


Fig. 7. Forecast LAI (top row) using the particle filter after assimilating at DOY 191 (21 days before the regression date). Forecast results are shown every three days together with the corresponding CubeSat-LAI observations (second row from top). The bias between the forecast LAI and observed LAI is depicted in the bottom row.

overall, assimilating up to three weeks before the identified regression date (Fig. 6c) provides accurate LAI predictions throughout the forecast period. In particular, assimilating one and two weeks before the regression date (Fig. 6a, 6b) produce the best LAI predictions with well contained errors ($< 0.1 \text{ m}^2/\text{m}^2$ and $0.13 \text{ m}^2/\text{m}^2$, respectively), while assimilating three weeks ahead (Fig. 6c) generate good prediction but with an increased error that rises up to $0.3 \text{ m}^2/\text{m}^2$. On the other hand, the assimilation of CubeSat observation 4 weeks in advance (Fig. 6d) produces a shift towards the observation only for the coming seven days. After that, the weighted mean worsens and converges closer to the particles' mean (green dots). The RMSE of the weighted mean against the observation nearly doubles ($0.81 \text{ m}^2/\text{m}^2$) when the assimilation is performed four weeks prior to the regression date.

The particle filter is applied on each pixel of the satellite image, to get the best estimate of the LAI map which is then used to predict end-of-season yield. As such, a reliable prediction of LAI is a requirement to accurately estimate within-field yield. Fig. 7 (top row) shows 21 days of LAI (forecast every 3 days) obtained after assimilating one CubeSat observation at DOY 191 (three weeks before the regression date). A comparison is presented for the corresponding CubeSat LAI observations (Fig. 7 middle row). In general, the LAI forecasts show a high level of

agreement when compared to the CubeSat LAI for much of the forecast period, with a less defined pattern identified only towards the end. The tendency of the forecast LAI to over- and under-estimate the CubeSat LAI can be seen in the bias maps (bottom row in Fig. 7), which reveal relatively small differences between the forecast and CubeSat LAI. Interestingly, the forecast LAI slightly underestimates the low (circular) LAI pattern in the bottom portion of the field only at DOY 194 and 197. In contrast, this patch is gradually overestimated in subsequent days (i. e., DOY 206, 209, 212). Although localized to a small area of the field, this behavior is likely a consequence of the inability of the model/assimilation scheme to correctly predict the lowest values of the LAI in a spatially-explicit domain.

The smoother LAI pattern at DOY 212 reflects the highest error (especially in the upper part of the field where lower LAI values were observed), which reaches $0.45 \text{ m}^2/\text{m}^2$ (on average). In general, the bias depicts how the areas with the lowest LAI values are those subject to the highest errors, which, as mentioned, are likely related to the ineffectiveness of the approach in reproducing the extreme LAI values.

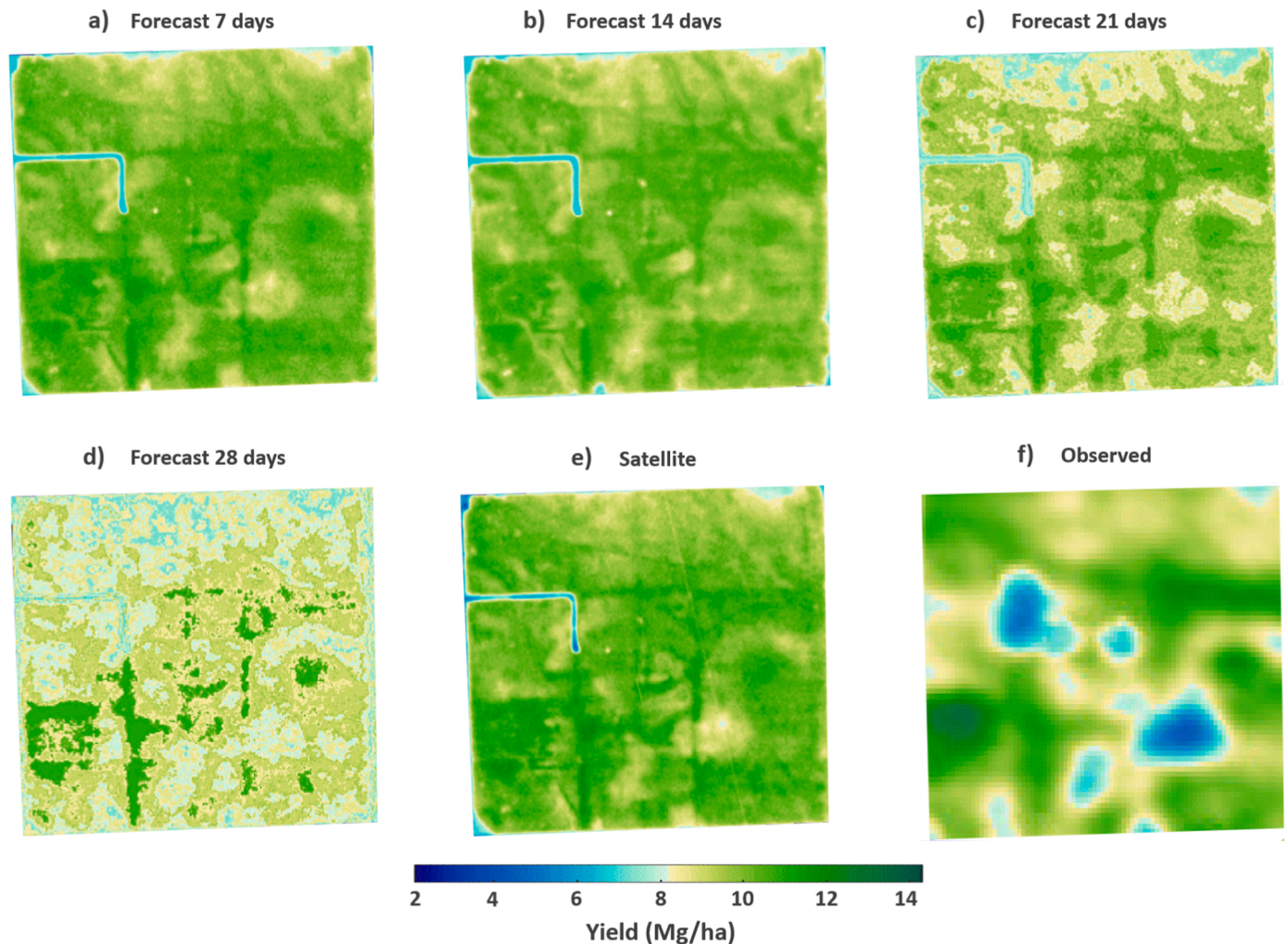


Fig. 8. Yield predictions after assimilating one single satellite observation (a) 7 days, (b) 14 days, (c) 21 days, and (d) 28 days prior to the regression date and by using the satellite observation at DOY 212 (e). Measured yield using harvesters equipped with yield monitor sensors (f).

Table 2Metrics (RMSE, rRMSE, R^2 , and bias) calculated for each of the four assimilation steps depicted in Fig. 8.

Forecast step	7 days	14 days	21 days	28 days
Evaluation Metrics	RMSE = 1.69 rRMSE = 11.75 R^2 = 0.73 bias = -0.0917	RMSE = 1.71 rRMSE = 11.82 R^2 = 0.7 bias = -0.1162	RMSE = 1.82 rRMSE = 12.71 R^2 = 0.68 bias = -0.2271	RMSE = 2.06 rRMSE = 14.42 R^2 = 0.65 bias = -1.0810

3.4. Integrating CubeSat LAI into APSIM allows accurate mapping of within-field yield

Fig. 8 demonstrates that the spatial yield estimates are similar to the measured yield, effectively capturing spatial differences with correlation values ranging from 0.65 (assimilating 4 weeks before the regression date) to 0.73 (assimilating 1 week before the regression date). The correlation was performed using a pixel-by-pixel comparison between the measured in situ yield map and the predicted values. The yield prediction results confirm that after the assimilation step the model reproduces the sub-field yield variations reasonably well. As expected, assimilating closer to the regression date (Fig. 8a) enables the prediction of within-field yield with greater accuracy compared to earlier assimilations (2–4 weeks), both in terms of spatial variability and in terms of actual yield pixel values (R^2 = 0.73). Assimilating two (Fig. 8b) and three (Fig. 8c) weeks before the regression date still provides a reliable yield map, delivering within-field patches that resemble those of the measured yield, albeit with a slight decrease in R^2 to 0.7. When the assimilation is performed four weeks before the regression date (Fig. 8d) the within-field yield variability underestimates the measured yield (RMSE = 2.06 Mg/ha, bias = -1.08). It should be noted that the four-week assimilation is performed towards the end of the green-up period (see Fig. 6d), where LAI undergoes rapid development and sharply increase in values (see Fig. 4). It is likely that the assimilation

during such a rapidly changing LAI period is unable to provide accurate forecast estimates of LAI, which results in inaccurate within-field yield predictions. The predicted yield after assimilation was also compared with the yield prediction obtained using the CubeSat imagery (Fig. 8e) corresponding to DOY 212, suggesting similar within-field variability with respect to the assimilation results: at least until assimilating 3 weeks before the regression date.

In general, although the spatial variability is fairly well reproduced, the predicted yields do not well replicate the extreme values of the measured yield, such as the lowest yield-value patches that appear in the in situ yield maps (e.g., the darker blue areas in Fig. 8f). Such behavior has been explained in previous studies such as in Corti et al. (2018), who found that vegetation indices can saturate at high LAI values, i.e. become insensitive to variations in LAI values above a certain threshold, due to the horizontal leaf stratification. However, it may also be related to the inability of the simple linear regression model to predict non-linear patterns within the field. The progressive improvement in the predictive power of the four assimilating dates (i.e. from 4 to 1 week before the regression date) is also displayed by the evaluation metrics of Table 2.

Fig. 9 presents the full comparison of end-of-season yield predictions when assimilating a single CubeSat observation 30 days to 1 day before the regression date, with the forecast LAI on DOY 212 converted into

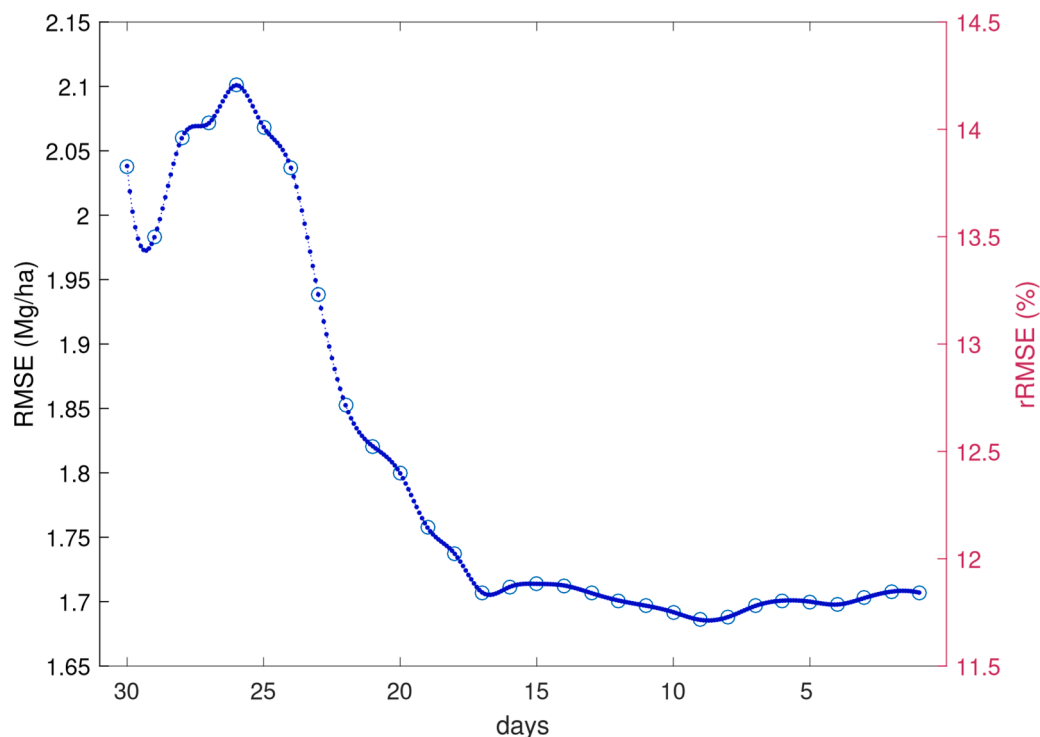


Fig. 9. RMSE (left y-axis) and rRMSE (right y-axis) were calculated on a pixel-by-pixel basis between predicted and in-field yield. The predicted yield was calculated by applying the regression equation on the forecasted LAI at DOY 212 after assimilating one satellite observation from one day (DOY 211) until thirty days (DOY 182) before the regression date.

yield through the previously trained regression. It should be noted that although we tested our method by assimilating every day from 30 to 1 day before the regression date, the assimilation was performed on a single time step only. The trends of the RMSE and rRMSE reveal that in this particular scenario, assimilating up to 17 days before the regression date generates yield predictions with RMSE and rRMSE less than 1.7 Mg/ha and 12% respectively. The errors sharply increase when the assimilation is performed earlier than 17 days, with up to 2.5% increase in rRMSE, and 0.4 Mg/ha in RMSE. As noted earlier, this error increase corresponds to the assimilation of CubeSat observations during the vegetative stage. Assimilating during such a period (see Fig. 6d) produces less accurate LAI forecasts and therefore less accurate end-of-season yield predictions, since yield is linearly related to LAI through the regression. Interestingly, although the yield predictions are impacted in terms of spatial variability (i.e., the predicted within-field variability does not mirror that of the in situ yield; e.g. see the comparison between Fig. 8d and 8f), the comparison between field-averaged prediction and measured yield shows a reduced error.

To test the effectiveness of the assimilation step in providing improved estimates of field-averaged yield, we compare the results against the outcomes of the APSIM model after calibrating its cultivar parameters (see Ziliani et al., 2021). The assimilation reduced the field-averaged yield error from 1 Mg/ha (result with calibrated model) and 0.5 Mg/ha (result using satellite imagery, Fig. 8e) to 0.2 Mg/ha, when the assimilation was performed up to three weeks before the regression date. The average yield worsens when the assimilation is

performed after 21 days, with errors reaching up to 1.1 Mg/ha.

Finally, to demonstrate the scalability of the approach, we develop a maize yield map for the 2019 season for fields surrounding the studied US-Ne3 site. The region includes two additional maize fields (US-Ne1 and US-Ne2 in Fig. 10) for which spatially varying yield was also available for evaluation. The map was constructed by using the trained regression of Eqn 1 (i.e. the same as used for US-Ne3) to estimate yield for each assimilated 3 m CubeSat pixel in the region. Plot outlines are clearly distinguishable in the yield map, with large variation in productivity visible both across and within fields. For instance, yield estimates indicate that productivity can differ on adjacent fields, consistent with productivity dispersion observed in other agricultural and non-agricultural systems (Burke and Lobell, 2017; Lobell et al., 2009) and suggest that management differences are a key determinant in overall yield variation in the region. Indeed, the irrigated pivots US-Ne1 and US-Ne2 (adjacent round pivots in the bottom left of Fig. 10) show a much more uniform yield map (and with higher values) compared to that retrieved for the rainfed US-Ne3 (top right of the figure). Pixel-by-pixel comparison between the measured in situ yield map and the predicted values for US-Ne1 and US-Ne2 reported a correlation of 0.72 (bias = -0.1009) and 0.74 (bias = -0.0812), respectively (which is comparable to that of US-Ne3 for the same date; 0.73, bias = -0.1162), confirming that the approach was able to predict within-field yield with good accuracy also for other maize fields (albeit considering that they were spatially close to each other).

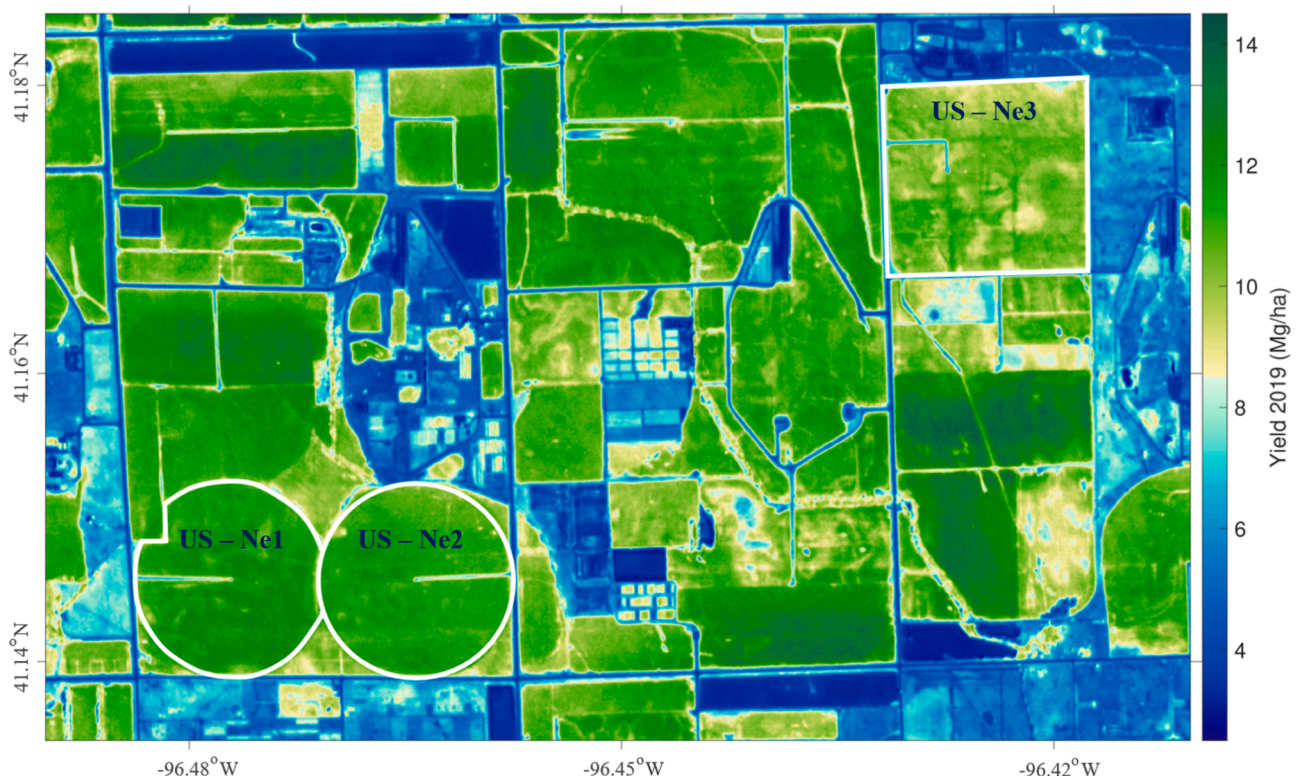


Fig. 10. Yield predictions for the 2019 season of a larger region close to US-Ne3 after assimilating one single satellite observation 14 days prior to the regression date. Field US-Ne1 and US-Ne2 (two center pivot irrigation systems) are shown in the bottom left of the region.

4. Discussion

Our results demonstrate the benefit of integrating CubeSat data into the APSIM crop model to enhance yield prediction at both the field and within-field scales. The within-field variability of the studied maize field was well represented by CubeSat Fusion data, indicating its utility for agricultural applications. The benefit of the high temporal resolution in the CubeSat data is that it allows the acquisition of daily and cloud-free data that can be efficiently integrated into a crop model whenever needed throughout the growth cycle. This allows farmers to predict their yield ahead of time while undertaking necessary measures to address problems as they arise (and thereby maximize output and profit). While previous studies have highlighted the limitations of remote sensing platforms with lower temporal resolution (as they were unable to provide crop information in a timely manner throughout the season) (Mulla, 2013), CubeSat data overcome this barrier as they made available crop biophysical information at any day during the season (Poghosyan and Golkar, 2017). Equally important is the spatial resolution, with these CubeSat platforms offering pixel-level retrievals at the 3 m scale, providing two orders of magnitude improvement over the area sampled by a comparable Landsat pixel (Houborg and McCabe, 2018a). In addition, the spatiotemporal advantage of CubeSats may be of particular interest to smallholder studies (especially related to developing countries), where the size of plots can be less than 1 ha in area (Lowder et al., 2016).

Thematically similar studies that have applied regression equations to satellite images in order to predict yield have shown variable correlations. For instance, a scalable crop yield mapping approach (SCYM, Lobell et al., 2015) was tested to predict field-level yield across a number of states and years (2000–2013) in the US Corn Belt using Landsat imagery, reporting a maximum R^2 value of 0.58. More recently, Jeffries et al. (2019) developed a variation of the SCYM to estimate sub-field maize yield at 10–30 m spatial resolution and applied it over three fields in Nebraska (including the US-Ne3 that is part of our study). Jeffries et al. (2019) employed both Landsat and aerial hyperspectral imagery, and validated their results at the pixel level using harvester yield monitor records (rather than field or county level) from 2002 to 2013 (maximum R^2 value of 0.37). The different experimental settings and conditions of previous studies make direct comparisons complicated. However, the combination of the assimilation scheme and regression approach used in our study points towards an increase in statistical performance and the possibility to improve yield prediction accuracy. Nevertheless, the DA framework must be evaluated under a wider range of field conditions, crop types, and over longer periods to better determine the increase in predictive power. With the increasing availability of high spatiotemporal resolution satellite data, and opportunities for cloud-based analysis (Ampatzidis et al., 2020), there is an expectation that more frequent observations can improve yield predictions (Burke and Lobell, 2017). However, further analysis and additional data-model fusion and assimilation-related studies are required to determine precisely how much information gain is achieved by regular, as opposed to occasional, in-time model updating.

In the particle filter step, the particles generated with multiple APSIM runs show that the LAI dynamics were relatively stable during the mid-season (i.e. parallel trajectories from DOY 200 to 240, see Fig. 6). The stability in LAI particles is most likely related to the characteristics of modern maize cultivars (typical of the study region), which have a longer growth period, large leaf area, and slower leaf senescence. These characteristics lead to a significant increase in yield accumulation

(Sun et al., 2019). It follows that, for this location and agronomy, assimilating a single observation up to three weeks before the regression date (see Figs. 6c and 8c) could be sufficient to predict within-field yield if the acquisition date is optimal (i.e. no cloud cover) (Machwitz et al., 2014). It is also worth noting that only a single metric of crop condition (i.e. LAI) was explored here, whereas there are numerous vegetation indices available from satellite platforms that may provide greater insight into crop yield (Angel et al., 2021). In recent studies on crop yield prediction using remote sensing data (Burke and Lobell, 2017; Franz et al., 2020) the Green Chlorophyll Vegetation Index (GCVI) was shown to outperform more commonly used indices like the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974). Indeed, some parallel analyses performed during the current study using the GCVI provided slightly improved results relative to the RMSE (around 0.1 Mg/ha). Further investigation is required to understand the benefits of assimilating related vegetation metrics into the APSIM model. Likewise, future studies should also assess the relative information gain (or performance improvement) that can be realized from the incorporation of multiple vegetation indices into the assimilation step.

Although this work represents one of the first attempts to predict the within-field yield variability at such high spatial resolutions and without the need for ground calibration, further research is required to assess the scalability of the method to other crops, different latitudes, and larger regions. Previous literature has illustrated that changes in environments and farm management practices, which result in local variations of crop canopies and therefore crop yields (Delmotte et al., 2011; Knox et al., 2012), are likely to affect or limit the generality (or transferability) of any yield prediction approach (Prasad et al., 2006). The approach developed in this study integrates satellite-derived indices whose spatial patterns are the main contributors to the observed within-field yield variability. Part of the approach's discriminative power is driven by the selection of the regression date (from the peak of maximum correlation between yield and LAI, as in Fig. 5). The timing of when that peak occurs can vary between fields and regions, due to the interaction between cultivated varieties (maturity types) and growing degree days (temperature) (Evans, 1996).

Generally speaking, our combined data assimilation and regression approach could be readily transferred to other regions. The data requirements remain mostly the same, with minor changes needed to accommodate different crops. For instance, a specific crop-type model must be used (e.g. APSIM-soybean or APSIM-wheat) as the regression is trained using the crop model's simulated variables (such as LAI) against simulated yield (see Section 2.4 for more details). Similarly, the CubeSat based LAI would need to undergo validation against measured values to ensure its accuracy for other crops. While Planet's near-daily coverage at 3 m spatial resolution could allow for on-demand yield predictions, ancillary data required by the model (i.e., soil and weather components) is generally only available at coarser resolutions. Nevertheless, the approach presented here demonstrates how it is possible to retrieve within-field yield prediction with good accuracy even when using coarse resolution soil information. For regions where detailed soil data are lacking, low spatial resolution soil profiles (or soil profiles of a comparable/nearby location) could be used to deliver field-scale yield predictions. An example of such an approach was tested in a smallholder agricultural systems by Burke and Lobell (2017) and Jain et al. (2017).

Overall, this approach illustrates the clear benefit of utilizing the high-resolution daily Planet Fusion products to predict within-field yield variability. The developed approach has considerable utility for precision agricultural applications in not only identifying patches of lower

and higher yield within a field (as shown in Fig. 8), but also in driving strategic management decisions aimed at increasing profitability. While not exhaustive, this study demonstrated the high predictability of maize yield at the field and within-field scales from remotely-sensed and modeled vegetation indices.

5. Conclusion

An approach for within-field crop yield mapping and model prediction was developed to leverage the information content of new high spatiotemporal resolution satellite data. The procedure uses the APSIM crop model to simulate plant physiological response to environmental and management conditions, and integrates data using a particle filter scheme. By applying a retrospectively trained regression to APSIM model forecasts of LAI after a single assimilation step, good agreement was achieved compared to field-collected yield. Importantly, the approach makes no use of in-field yield data to calibrate the regression model and has demonstrated that even with a single assimilation step, it is possible to provide yield estimates with good accuracy up to three weeks prior to an identified “regression date”. This is an important outcome, as it allows the development of an accurate end-of-season yield forecast without the need to wait for the actual satellite overpass (which may be interrupted due to cloud cover or other meteorological conditions).

Efficient and accurate estimation of yield at the within-field scale could enable improved strategies for agricultural intervention as well as an enhanced evaluation of farmer profit based on spatially explicit yield forecast. There are a number of areas for potential developments of the proposed approach, including the sequential assimilation of multiple images into crop models. As CubeSat data provides the capacity for daily observation, further research is also needed to evaluate the influence of the LAI assimilation window on yield estimation and perhaps an exploration of the integration of related biophysical metrics (i.e. evapotranspiration, canopy cover) either individually or in combination (i.e. multi-variable assimilation). As additional observations can be

assimilated into the model, other assimilation tools can also be explored, such as ensemble smoothers. Further targeted research will be needed to explore the broader utility of the approach to aid in the field management and decision-making process. For instance, apart from the obvious direct benefit to producers, access to spatially distributed information on end-of-season yield could have a range of applications for both government and commercial entities responsible for commodity markets forecasts and crop insurance products, both at the regional and national scales.

Declaration of Competing Interest

None.

Acknowledgments

Research was supported by the King Abdullah University of Science and Technology (Grant number OSR-2017-CRG6) and in collaboration with Planet Labs corporation and the Eastern Nebraska Research and Extension Center at the University of Nebraska-Lincoln (Lincoln, United States), who provided datasets for model calibration and validation. Funding for the AmeriFlux core site was provided by the U.S. Department of Energy's Office of Science. Field data used within this research was a contribution from the Long-Term Agroecosystem Research (LTAR) network. LTAR is supported by the United States Department of Agriculture. T.E.F. acknowledges the financial support of the USDA National Institute of Food and Agriculture, Hatch project #1009760, #1020768 and project #2019-67021-29312. The author would like to thank Prof. Scott Chapman (University of Queensland) and Dr. BangYou Zhang (CSRIO) for their advice with the APSIM crop model simulations.

Appendix

Fig. 11.

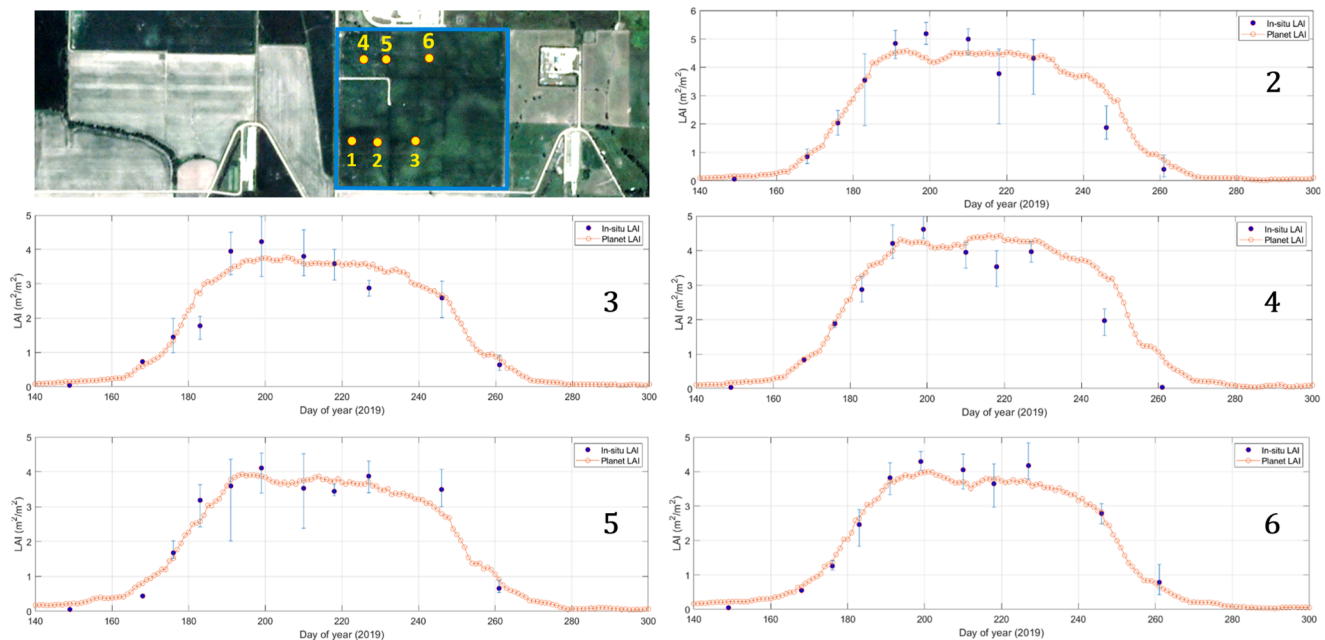


Fig. 11. CESTEM-based time series of PlanetScope (PS) LAI at the 2–5 intensive measurement zones (IMZs) in the corn study-unit (Figure 1, orange circles) against in-situ destructive measurements performed throughout the season. The displayed ranges are the results of multiple samples within the IMZs.

References

- Abendroth, L.J., Elmore, R.W., Boyer, M.J., Marlay, S.K., 2011. Corn Growth and Development. *PMR*.
- Amptatzidis, Y., Partel, V., Costa, L., 2020. Agroviz: cloud-based application to process, analyze and visualize UAV-collected data for precision agriculture applications utilizing artificial intelligence. *Comput. Electron. Agric.* 174, 105457.
- Aragon, B., Houborg, R., Tu, K., Fisher, J.B., McCabe, M., 2018. CubeSats enable high spatiotemporal retrievals of crop-water use for precision agriculture. *Remote Sens. (Basel)* 10 (12), 1867.
- Aragon, B., Ziliani, M.G., Houborg, R., Franz, T.E., McCabe, M.F., 2021. CubeSats deliver new insights into agricultural water use at daily and 3m resolutions. *Sci. Rep.* 11 (1), 1–12.
- Baez-Gonzalez, A.D., Kiniry, J.R., Maas, S.J., Tiscareno, M.L., Macias, C., Mendoza, J.L., Manjarrez, J.R., 2005. Large-area maize yield forecasting using leaf area index based yield model. *Agron. J.* 97 (2), 418–425.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Evans, R., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* 82 (11), 2415–2434.
- Basso, B., Cammarano, D., Carfagna, E., 2013. In: *Review of crop yield forecasting methods and early warning systems* Paper presented at the Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics. FAO Headquarters, Rome, Italy.
- Beuchle, R., Eva, H.D., Stibig, H.-J., Bodart, C., Brink, A., Mayaux, P., Belward, A., 2011. A satellite data set for tropical forest area change assessment. *Int. J. Remote Sens.* 32 (22), 7009–7031.
- Bitzer, M., Herbek, J., Bessin, R., Green, J., Ibendahl, G., Martin, J., Vincelli, P., 2000. A Comprehensive Guide to Corn Management in Kentucky. University of Kentucky, Lexington, KY, USA.
- Bolten, J.D., Crow, W.T., Zhan, X., Jackson, T.J., Reynolds, C.A., 2009. Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. *IEEE J. Sel. Topic. Appl. Earth Observat. Remote Sens.* 3 (1), 57–66.
- Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci.* 114 (9), 2189–2194.
- Carberry, P., Abrecht, D., 1991. Tailoring Crop Models to the Semiarid Tropics.
- Carberry, P., Muchow, R., McCown, R., 1989. Testing the CERES-Maize simulation model in a semi-arid tropical environment. *Field Crops Res.* 20 (4), 297–315.
- Clevers, J., 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sens. Environ.* 61 (2), 221–228.
- Corti, M., Cavalli, D., Cabassi, G., Gallina, P.M., Bechini, L., 2018. Does remote and proximal optical sensing successfully estimate maize variables? A review. *Eur. J. Agron.* 99, 37–50.
- De Wit, A.d., Van Diepen, C., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agric. For. Meteorol.* 146 (1–2), 38–56.
- Delmotte, S., Tittonell, P., Mouret, J.-C., Hammond, R., Lopez-Ridaura, S., 2011. On farm assessment of rice yield variability and productivity gaps between organic and conventional cropping systems under Mediterranean climate. *Eur. J. Agron.* 35 (4), 223–236.
- Edmeades, G., Daynard, T., 1979. The relationship between final yield and photosynthesis at flowering in individual maize plants. *Can. J. Plant Sci.* 59 (3), 585–601.
- Evans, L.T., 1996. *Crop Evolution, Adaptation and Yield*. Cambridge university press.
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53 (4), 343–367.
- Fang, H., Liang, S., Hoogenboom, G., Teasdale, J., Cavigelli, M., 2008. Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. *Int. J. Remote Sens.* 29 (10), 3011–3032.
- Foolad, F., Franz, T., Wang, T., Gibson, J., Kilic, A., Allen, R., Suyker, A., 2017. Feasibility Analysis of Using Inverse Modeling for Estimating Field-Scale Evapotranspiration in Maize and Soybean Fields from Soil Water Content Monitoring Networks.
- Frantz, D., 2019. *FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. Remote Sens. (Basel)* (9), 11. <https://doi.org/10.3390/rs11091124>.
- Franz, T.E., Pokal, S., Gibson, J.P., Zhou, Y., Gholizadeh, H., Tenorio, F.A., Ziliani, M., 2020. The role of topography, soil, and remotely sensed vegetation condition towards predicting crop yield. *Field Crops Res.* 252, 107788.
- Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Crutchfield, J., 2019. A comparison of global agricultural monitoring systems and current gaps. *Agric. Syst.* 168, 258–272.
- Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Snow, V., 2018. APSIM Next Generation: overcoming challenges in modernising a farming systems model. *Environ. Modell. Softw.* 103, 43–51.
- Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Murphy, C., 2014. APSIM—evolution towards a new generation of agricultural systems simulation. *Environ. Modell. Softw.* 62, 327–350.
- Hoogenboom, G., 2000. Contribution of agrometeorology to the simulation of crop production and its applications. *Agric. For. Meteorol.* 103 (1–2), 137–157.
- Hoteit, I., Luo, X., Bocquet, M., Kohl, A., Ait-El-Fquih, B., 2018. Data assimilation in oceanography: current status and new directions. *New Front. Oper. Oceanogr.* 465–512.
- Houborg, R., McCabe, M., 2018a. Daily retrieval of NDVI and LAI at 3m resolution via the fusion of CubeSat, Landsat, and MODIS data. *Remote Sens. (Basel)* 10 (6), 890. <https://doi.org/10.3390/rs10060890>.
- Houborg, R., McCabe, M.F., 2018b. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sens. Environ.* 209, 211–226. <https://doi.org/10.1016/j.rse.2018.02.067>.
- Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Wu, Y., 2019. Assimilation of remote sensing into crop growth models: current status and perspectives. *Agric. For. Meteorol.* 276, 107609.
- Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Wu, W., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* 204, 106–121.
- Ines, A.V., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* 138, 149–164.
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., Ustin, S.L., 2009. PROSPECT+ SAIL models: a review of use for vegetation characterization. *Remote Sens. Environ.* 113, S56–S66.
- Jain, M., Rao, P., Srivastava, A.K., Poonia, S., Blesh, J., Azzari, G., Lobell, D.B., 2019. The impact of agricultural interventions can be doubled by using satellite data. *Nature Sustain.* 2 (10), 931–934.
- Jain, M., Singh, B., Srivastava, A., Malik, R.K., McDonald, A., Lobell, D.B., 2017. Using satellite data to identify the causes of and potential solutions for yield gaps in India's Wheat Belt. *Environ. Res. Lett.* 12 (9), 094011.
- Jeffries, G.R., Griffin, T.S., Fleisher, D.H., Naumova, E.N., Koch, M., Wardlow, B.D., 2019. Mapping sub-field maize yields in Nebraska, USA by combining remote sensing imagery, crop simulation models, and machine learning. *Precis. Agric.* 1–17.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data assimilation of remote sensing and crop models. *Eur. J. Agron.* 92, 141–152.
- Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020. Estimation of maize yield by assimilating biomass and canopy cover derived from hyperspectral data into the AquaCrop model. *Agric. Water Manage.* 227, 105846.
- Johansen, K., Ziliani, M.G., Houborg, R., Franz, T., McCabe, M.F., 2021. Leaf area index from CubeSats at 3m resolution every day. *Environ. Res. Lett.* in review.
- Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L., Ritchie, J.T., 2003. The DSSAT cropping system model. *Eur. J. Agron.* 18 (3–4), 235–265.
- Keating, B., Godwin, D., Watiki, J., 1991. Optimising nitrogen inputs in response to climatic risk. *Climatic Risk in Crop Production-Models and Management for the Semi-arid Tropics and Subtropics*. CAB International. Wallingford.
- Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Hochman, Z., 2003. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* 18 (3–4), 267–288.
- Knox, J., Hess, T., Daccache, A., Wheeler, T., 2012. Climate change impacts on crop productivity in Africa and South Asia. *Environ. Res. Lett.* 7 (3), 034032.
- Lemaire, G., van Oosterom, E., Jeuffroy, M.-H., Gastal, F., Massignam, A., 2008. Crop species present different qualitative types of response to N deficiency during their vegetative growth. *Field Crops Res.* 105 (3), 253–265.
- Li, X., Du, H., Mao, F., Zhou, G., Chen, L., Xing, L., Cui, L., 2018. Estimating bamboo forest aboveground biomass using EnKF-assimilated MODIS LAI spatiotemporal data and machine learning algorithms. *Agric. For. Meteorol.* 256, 445–457.
- Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Res.* 143, 56–64.
- Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop yield gaps: their importance, magnitudes, and causes. *Annual Review of Environment and Resources*, p. 34.
- Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* 164, 324–333.
- Lowder, S.K., Skoet, J., Raney, T., 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Dev.* 87, 16–29.
- Lu, Y., Chibarabada, T.P., Ziliani, M.G., Onema, J.-M.K., McCabe, M.F., Sheffield, J., 2021. Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model. *Agric. Water Manage.* 252, 106884.
- Maas, S.J., 1988. Use of remotely-sensed information in agricultural crop growth models. *Ecol. Modell.* 41 (3–4), 247–268.
- Machwitz, M., Giustarini, L., Bossung, C., Frantz, D., Schlerf, M., Lilienthal, H., Udelhoven, T., 2014. Enhanced biomass prediction by assimilating satellite data into a crop growth model. *Environ. Modell. Softw.* 62, 437–453.
- Massignam, A., Chapman, S., Hammer, G., Fukai, S., 2009. Physiological determinants of maize and sunflower grain yield as affected by nitrogen supply. *Field Crops Res.* 113 (3), 256–267.
- McCabe, M.F., Aragon, B., Houborg, R., Mascaro, J., 2017a. CubeSats in hydrology: ultrahigh-resolution insights into vegetation dynamics and terrestrial evaporation. *Environ. Res.* 53 (12), 10017–10024.
- McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Franz, T.E., 2017b. The future of Earth observation in hydrology. *Hydrol. Earth Syst. Sci.* 21 (7), 3879.
- McCown, R.L., Hammer, G.L., Hargreaves, J.N.G., Holzworth, D.P., Freebairn, D.M., 1996. APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. *Agric. Syst.* 50 (3), 255–272.
- Mokhtari, A., Noory, H., Vazifedoust, M., 2018. Improving crop yield estimation by assimilating LAI and inputting satellite-based surface incoming solar radiation into SWAP model. *Agric. For. Meteorol.* 250, 159–170.
- Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosyst. Eng.* 114 (4), 358–371.

- Nguy-Robertson, A., Suyker, A., Xiao, X., 2015. Modeling gross primary production of maize and soybean croplands using light quality, temperature, water stress, and phenology. *Agric. For. Meteorol.* 213, 160–172.
- Peng, B., Guan, K., Tang, J., Ainsworth, E.A., Asseng, S., Bernacchi, C.J., Ewert, F., 2020. Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nat. Plants* 6 (4), 338–348.
- Poghosyan, A., Golkar, A., 2017. CubeSat evolution: analyzing CubeSat capabilities for conducting science missions. *Prog. Aerosp. Sci.* 88, 59–83.
- Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. *Int. J. Appl. Earth Obs. Geoinf.* 8 (1), 26–33.
- Resnick, D., 2020. Political economy of food system reform. *Nature Food* 1 (3), 154.
- Ritchie, S.W., Hanway, J.J., 1989. *How a Corn Plant develops*. Retrieved from.
- Rouse, J., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA spec. Publ.* 351 (1974), 309.
- Sharma, V., Irmak, S., 2012. Mapping spatially interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirements in Nebraska: part I. Precipitation and reference evapotranspiration. *Trans. ASABE* 55 (3), 907–921.
- Sibley, A.M., Grassini, P., Thomas, N.E., Cassman, K.G., Lobell, D.B., 2014. Testing remote sensing approaches for assessing yield variability among maize fields. *Agron. J* 106 (1), 24–32.
- Silvestro, P.C., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Casa, R., 2017. Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the Aquacrop and simple algorithm for yield (SAFY) models. *Remote Sens. (Basel)* 9 (5), 509.
- Soufizadeh, S., Munaro, E., McLean, G., Massignam, A., van Oosterom, E., Chapman, S., Hammer, G., 2018. Modelling the nitrogen dynamics of maize crops—Enhancing the APSIM maize model. *Eur. J. Agron.* 100, 118–131.
- Sun, J., Gao, J., Wang, Z., Hu, S., Zhang, F., Bao, H., Fan, Y., 2019. Maize canopy photosynthetic efficiency, plant growth, and yield responses to tillage depth. *Agronomy* 9 (1), 3.
- Suyker, A.E., Verma, S.B., 2008. Interannual water vapor and energy exchange in an irrigated maize-based agroecosystem. *Agric. For. Meteorol.* 148 (3), 417–427.
- Tewes, A., Hoffmann, H., Krauss, G., Schäfer, F., Kerkhoff, C., Gaiser, T., 2020. New approaches for the assimilation of LAI measurements into a crop model ensemble to improve wheat biomass estimations. *Agronomy* 10 (3), 446.
- Van Diepen, C.v., Wolf, J., Van Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. *Soil use Manag.* 5 (1), 16–24.
- van Leeuwen, P.J., 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Q. J. R. Meteorolog. Soc.* 136 (653), 1991–1999.
- van Velzen, N., Altaf, M.U., Verlaan, M., 2016. OpenDA-NEMO framework for ocean data assimilation. *Ocean Dyn* 66 (5), 691–702.
- Verma, S.B., Dobermann, A., Cassman, K.G., Walters, D.T., Knops, J.M., Arkebauer, T.J., Yang, H., 2005. Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems. *Agric. For. Meteorol.* 131 (1–2), 77–96.
- Vetra-Carvalho, S., Van Leeuwen, P.J., Nerger, L., Barth, A., Altaf, M.U., Brasseur, P., Beckers, J.-M., 2018. State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus A: Dyn. Meteorol. Oceanogr.* 70 (1), 1–43.
- Vos, J., Van Der Putten, P., Birch, C., 2005. Effect of nitrogen supply on leaf appearance, leaf growth, leaf nitrogen economy and photosynthetic capacity in maize (*Zea mays* L.). *Field Crops Res.* 93 (1), 64–73.
- Wang, E., Robertson, M., Hammer, G., Carberry, P.S., Holzworth, D., Meinke, H., McLean, G., 2002. Development of a generic crop model template in the cropping system model APSIM. *Eur. J. Agron.* 18 (1–2), 121–140.
- Zhang, X., Davidson, E.A., Mauzerall, D.L., Searchinger, T.D., Dumas, P., Shen, Y., 2015. Managing nitrogen for sustainable development. *Nature* 528 (7580), 51–59.
- Zhuo, W., Huang, J., Li, L., Zhang, X., Ma, H., Gao, X., Xiao, X., 2019. Assimilating soil moisture retrieved from Sentinel-1 and Sentinel-2 data into WOFOST model to improve winter wheat yield estimation. *Remote Sens. (Basel)* 11 (13), 1618.
- Ziliani, M.G., Parkes, S.D., Hoteit, I., McCabe, M.F., 2018. Intra-season crop height variability at commercial farm scales using a fixed-wing UAV. *Remote Sens. (Basel)* 10 (12), 2007.
- Ziliani, M.G., Altaf, M.U., Franz, T., Zheng, B., Chapman, S., Sheffield, J., Hoteit, I., McCabe, M.F., 2021. Improving APSIM crop model simulations using sensitivity analysis and Bayesian inference. *Agric. For. Meteorol.* in review.