INTRA-FIELD CROP YIELD VARIABILITY BY ASSIMILATING CUBESAT LAI IN THE APSIM CROP MODEL

Matteo G. Ziliani*[1], Muhammad U. Altaf[1], Bruno Aragon[2], Rasmus Houborg[2], Trenton E. Franz[3], Yang Lu[4], Justin Sheffield[4], Ibrahim Hoteit[5], Matthew F. McCabe[1]

[1]Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology, Kingdom of Saudi Arabia – (matteo.ziliani, umer.altaf, matthew.mccabe)@kaust.edu.sa;
[2]Planet, San Francisco, California, USA - rasmus.houborg@planet.com
[3]School of Natural Resources, University of Nebraska-Lincoln, USA - tfranz70@gmail.com
[4]School of Geography and Environmental Sciences, University of Southampton, United Kingdom – (Yang.Lu, Justin.Sheffield)@soton.ac.uk
[5]Physical Science and Engineering (PSE), King Abdullah University of Science and Technology, Kingdom of Saudi Arabia - ibrahim.hoteit@kaust.edu.sa

KEY WORDS: Crop yield prediction, Crop modeling, CubeSat, LAI, APSIM, Particle filter.

ABSTRACT:

Predicting within-field crop yield early in the season can help address crop production challenges to improve farmers' economic return. While yield prediction with remote sensing has been a research aim for years, it is only recently that observations with the suited spatial and temporal resolutions have become accessible to improve crop yield predictions.
Here we developed a yield prediction framework that integrates daily high-resolution (3 m) CubeSat imagery into the APSIM crop model. The approach trains a regression model that correlates simulated yield to simulated leaf area index (LAI) from APSIM. That relationship is then employed to determine the optimum date at which the regression best predicts yield from the LAI. Additionally, our approach can forecast crop yield by utilizing a particle filter to assimilate CubeSat-based LAI in the model APSIM to generate yield maps at 3 m several weeks before the optimum regression date. Our method was evaluated for a rainfed site located in the US Corn belt, using a collection of spatially varying yield data. The proposed approach does not need *in situ* data to rain the regression, with outcomes reporting that even with a single assimilation step, accurate yield predictions were provided up to 21 days before the optimum regression date. The spatial variability of crop yield was reproduced fairly well, with a good correlation against *in situ* measurements (R^2 = 0.73 and RMSE = 1.69), demonstrating that high-resolution yield predictions early in the season have great potential to meet and improve upon digital agricultural goals.

1. INTRODUCTION

Ensuring food security via timely crop monitoring and accurate crop yield estimation from field to regional and global scale is a priority policy goal for many countries (Cordell et al., 2009). While at the national or larger scales it is possible to study crop-climate interactions (Yu et al., 2012), accurate crop yield predictions at the field level earlier in the season can support farmers to tailor management choices of crop inputs (e.g. fertilizer and irrigation timing and amount (Basso et al., 2001), and also to estimate their net profit based on spatially explicit yield forecast (Tewes et al., 2020).
Modeling dynamic changes in plant growth and development through crop models, and remotely sensed key canopy state variables from satellites are essential tools to predict crop yield. Complex crop models such as, for instance, the Agricultural Production System Simulator (APSIM) (Holzworth et al., 2014) and the WOrld FOod STudies (WOFOST) model (Van Diepen et al., 1989), can temporally and dynamically describe key plant processes throughout the season including photosynthesis, soil dynamics, biomass, and yield formation, based on weather conditions (i.e. daily temperature, precipitation, and down-welling shortwave radiation), soil, and management information (Hoogenboom, 2000). Crop models have been increasingly employed for many precision farming operations and at different scales (Jin et al., 2018), but they are primarily employed for point-based applications that do not account for spatial variation in yield forecast. One way to bypass this constraint is to integrate spatial information on crop status (i.e. Vegetation Indices, VIs) from satellite platforms into crop models, which can increase the reliability and usefulness of these models (Huang et al., 2019).
Satellite data has been increasingly used to monitor agricultural systems from space, with yield estimation being one of the primary goals (Lobell, 2013). Yet, fine-level yield predictions have been compromised by the unavailability of high spatiotemporal resolution images. Indeed, the main drawback of using satellite platforms for fine-scale yield predictions has been their low spatial resolutions and temporal revisit time, which may hide main patterns and variability that happen at finer spatial scales (i.e. field and within-field levels). High spatial and temporal resolution data is key for precision agriculture applications, where farm management tactics (i.e. nitrogen and irrigation application) represent critical control variables to improve yield (Aragon et al., 2018). Although previous literature has shown the applicability of remote sensing technologies for yield monitoring (Lobell et al., 2015), field-level yield mapping for precision agricultural applications has not yet become widespread. In fact, there is a lack of in-field measurements available to assess the accuracy of remote sensing data, and the cost related to obtaining and processing high spatiotemporal resolution imagery - available from commercial platforms for the most part - can often be substantial. The combination of these factors can make remote sensing products unsuitable for practical applications.

* Corresponding author

Recent methods to translate satellite imagery into yield products have employed crop models to generate yield and crop variables (i.e. LAI, GCVI) values at the field level, which are then used to train a linear regression that translate observed satellite vegetation indices into yield. The trained regressions were applied to satellite images to generate yield maps both at the regional (Sibley et al., 2014; Lobell et al., 2015) and at the field scale (Jeffries et al., 2019). These applications have been useful for their scalability and because they do not rely on *in situ* yield observations. However, they are still constrained by the long satellite revisit times and coarse spatial resolution.

Providing expected yield values as early as possible during the season can enable farmers to undertake early corrective measures to improve their yield hence their economic returns. One avenue to improve yield prediction efforts is to combine the predictive abilities of crop models with the spatial information retrieved from space through data assimilation (DA) methods. DA combines observation of a system with the estimates from a dynamical model to create a more accurate description of that system including an uncertainty estimate (Vetra-Carvalho et al., 2018). DA can incorporate one or multiple observations of model state variables during the current crop growing season to improve crop predictions. These crop predictions (in the form of spatially varying VIs) can then serve as a base to predict end-of-season yield through regression models.

Amongst the DA methods, Ensemble Kalman Filters (EnKFs) are currently very popular due to their ease of filter implementation (Evensen, 2003). These methods employ a probabilistic framework and give estimates of the whole system state sequentially by propagating information forward in time. Standard EnKF is based on linear and Gaussian assumptions and with the increase in computational resources in the last decade or so, DA methods that can deal with non-Gaussian distribution such as particle filters became very popular (Carpenter et al., 1999). In contrast to the EnKF, the PF does not assume a Gaussian distribution for the observations nor for the model errors and can incorporate any probability density function allowing the propagation of non-Gaussian distributions through nonlinear models (van, Leeuwen, 2009).

Using DA methods, different types of remote sensing measurements have been assimilated into crop models, including soil moisture (Bolten et al., 2009; Zhuo et al., 2019), canopy cover (Silvestro et al., 2017; Jin et al., 2020), and more prominently, LAI (Li et al., 2014; Huang et al., 2015; Mokhtari et al., 2018; Li et al., 2018; Gilardelli et al., 2019). The reason behind LAI's popularity is due to its capability to effectively diagnose crop status, serving as an indicator of leaf abundance, phonological stage, and as a metric to assess diverse management practices. Several studies have subsequently proved the suitability of the Sentinel-2 (S2) multispectral imagery (10 m spatial resolution with a five-day revisit time) to assess crop status (Clevers et al., 2017), and to map within-field variability as a foundation for precision agriculture applications (Fieuzal et al., 2020). While these studies showed the emerging potential that satellite platforms have in advancing yield predictions, their spatial and temporal resolutions are still not high enough to properly resolve sub-field level variations that can be used to improve precision agriculture applications at smaller scales. For example, 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.

Recent advances in Earth Observation platforms such as component miniaturization and the use of reusable launch vehicles have enabled the development of microsatellites (referred to herein as CubeSats) that can provide unprecedented agricultural monitoring capabilities at a 3 m spatial resolution at daily revisit time (Aragon et al., 2021). Assimilating daily CubeSat products into crop models presents an opportunity to increase the predictive ability of the models and drive crop modeling advances, forecasting, and yield estimates (Ziliani, 2022). The increased spatiotemporal resolution of these products can resolve fine-scale site characteristics (e.g. soil, topography) that can affect crop growth and development, especially when combined with occurring stressors such as soil salinity or nutrient deficiency (Franz et al., 2020).

Here we present a novel approach for crop yield estimation by combining the daily and spatial explicit utility of CubeSat imagery with the predictive ability of crop models. Precisely, building on prior studies (Sibley et al., 2014; Lobell et al., 2015), we use the APSIM crop model to simulate realistic field-level yield and LAI data and then train a linear regression to translate CubeSat-derived LAI into 3 m yield maps. We further enhance the approach by using a particle filter to assimilate CubeSat LAI maps into the APSIM crop model. The assimilation of CubeSat imagery enables APSIM to forecast weeks of LAI maps which are used to predict yield - through the previously trained linear regression - without waiting for the satellite image to become available. Besides deriving the highest within-field resolution yield predictions to date, the proposed approach relies on daily satellite images that can potentially be assimilated into the crop model, relaxing the constrain imposed by satellite platforms with a longer revisit time. The proposed data assimilation scheme is then evaluated against *in situ* yield observations from a rainfed maize site located in the United States Corn belt. Our results indicate that the predicted yield and biophysical parameters (i.e. LAI) can be used in precision agricultural contexts where real-time forecasts of within-field conditions are required, avoiding the overhead time of acquiring and processing new imagery from current satellite platforms.

2. MATERIALS AND METHODS

2.1 The study site

The study area is part of the US corn belt and is actively managed by the Eastern Nebraska Research and Extension Center (ENREC) at the University of Nebraska-Lincoln (UNL). Figure 1 shows the site's location in relation to the United States of America (left) and the Intensive Management Zones (IMZ) where LAI samples were collected for validation (right) of the satellite product. Meteorological forcing data (air temperature, precipitation, and solar radiation) came from a weather station attached to an eddy covariance tower, part of the Amerifux network (ID: US-Ne3). The site's weather is characterized by cold winters, hot summers, and humid conditions (Baldocchi et al., 2001). The US-Ne3 site is under a no-till management policy, with fertilizer applied before each planting cycle; US-Ne3 is completely rainfed and alternates between maize (*Zea mays L.*) and soybeans (*Glycine max*). Typically, sowing occurs between late April and the middle of May, while harvesting happens at the end of October and early November. This study used data from 2001 to 2019, comprising 10 maize-growing seasons. Soil textures, layers, and other parameters were obtained from the US Web Soil Survey agency. A more detailed description of the site's characteristics can be found in Suyker and Verma (2008) and Verma et al. (2005). Samples of LAI (m^2/m^2) were collected throughout the season every 10 to 14 days at the six IMZs shown on the right of

Figure 1 (Nguy-Robertson et al., 2015). Yield at 10 m spatial resolution was collected at the end of each growth cycle via harvesting tractors equipped with GPS tracking technologies.

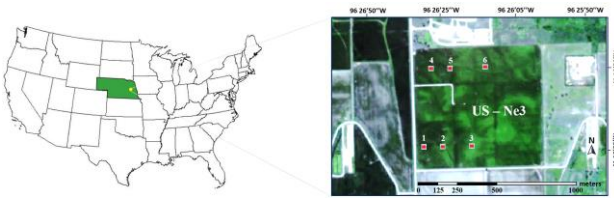

Figure 1. Location of the study site in the US and the Intensive Management Zones (IMZ) in the field US-Ne3 where LAI collection is performed during the crop growing season.

2.2 The APSIM model

APSIM is a system modeling framework that simulates the development of a crop through its phenological stages, leaf, biomass, and yield production, which are all evaluated on a daily time step (Wang et al., 2002). In this study, the maize model (APSIM-Maize version 7.9) included in the APSIM framework was used to simulate maize growth and development for the studied period in the area of interest (Section 2.1). For simplicity, from now on, the APSIM-Maize model will be referred to as the APSIM model. The model was constructed using as a base the CERES-Maize model (Keating et al., 1991), with a fundamental distinction from CERES-Maize being the routines that kill the crops if severe water stress is present during the early- to the mid-vegetative stage (Carberry and Abrecht, 1991).

Meteorological forcings are key components of the model to accurately drive these processes and simulations throughout the growing season. Maximum and minimum daily temperatures, daily rainfall, and solar radiation are needed to run the model. Management interventions generally include planting information such as crop type, planting date, depth, density, and nutrients applications (i.e. tillage, fertilizer, and irrigation).

APSIM partitions phenology into sub-phases. The length of each phase is affected by temperature, photoperiod, and nutrients deficiencies, such as nitrogen and water-limited conditions. When the latter is present, the expansion of the leaves diminishes and so is their capacity to capture solar radiation (Massignam et al., 2009). Grain yield is simulated as the product of grain size and grain number. The latter is estimated using the function developed by Edmeades and Daynard (1979) which considers the average daily growth rate per plant between tassel initiation and the start of grain filling and the potential grain number per ear. A complete overview of the APSIM model and its modules are provided in Keating et al. (2003).

2.3 LAI estimation using Planet CubeSat

The surface reflectance product used in this study comes from harmonizing and gap-filling CubeSat imagery from the Planet satellites constellation, each equipped with a four-band camera (blue, green, red, and near-infrared) in a 3U configuration. The CubeSat images were harmonized to Landsat 8 and Sentinel-2 surface reflectance (SR) data, which was computed using the FORCE methodology (Frantz, 2019), using the CubeSat Spatio-Temporal Enhancement Method (CESTEM) (Houborg and McCabe, 2018a). CESTEM is a machine learning approach that constructs relationships between CubeSat imagery and a high-quality radiometric reference (with a coarser spatiotemporal resolution), in this study that reference was the FORCE SR data. Additionally, CESTEM uses Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance (NBAR) data from the Moderate Resolution Imaging Spectroradiometer (MODIS) when there are no near-coincident reference images and for gap-filling purposes. Further details on the CESTEM methodology for radiometric normalization can be found in Houborg and McCabe (2018b).

The LAI imagery was generated by combining CESTEM with reference LAI data from the regularized canopy reflectance model (REGFLEC), a multi-step regularization approach that retrieves LAI from satellite reflectance data (Houborg and McCabe, 2016). REGFLEC used the FORCE SR data to produce LAI images at 30 m resolution, which served as the reference for the CESTEM process (i.e., CESTEM constructed relationships between the CubeSat imagery and the LAI from REGLEC) to derive 3 m daily images of LAI. A more detailed description of the CESTEM-LAI methodology is presented in Houborg and McCabe (2018c), and Houborg and McCabe (2018a).

2.4 Estimating crop yield using regression

Empirical models such as linear regressions can be rapidly applied to VIs derived from satellites to predict crop yield at the end of the season. The regression approach uses the APSIM crop model to calibrate a simple regression that correlates simulated yield to a simulated VI (Clevers, 1997; Sibley et al., 2014; Lobell et al., 2015). The trained regression can then be used to transform a VI from satellite into yield. The performed APSIM simulations spam a realistic range of meteorological, soil, and management practices in the region based on previous literature (Lobell et al., 2015; Jeffries et al., 2019), and information obtained from ENREC (Section 2.1). Since maize was considered for the analysis, simulations for each odd year from 2001 to 2017 were run (US-Ne3 grows maize on altering years), generating daily model outcomes including LAI and yield. A total of 7776 APSIM runs were performed based on multiple combinations of management inputs (fertilizer rates, sowing dates, sowing densities) and model parameters (i.e. cultivar thermal times). Daily meteorological data were used to initialize the model for each year, whose values were available from an *in situ* weather station, and include max and min temperatures, solar radiation, and precipitation.

The simulated yield and the simulated VI (i.e. LAI) can then be combined to train a simple regression model:

$$yield = a_t + b_t * VI_t + \varepsilon \ , \qquad (1)$$

where *a* and *b* represent the regression coefficients to be calibrated, and *VI* is the vegetation index map employed to predict yield at a specific date *t*. The proposed methodology does not rely on any field yield measurement but only on the crop model's ability to relate model variables (in this case LAI) to model outputs (yield).

A separate regression is created for each day of the season with the VI of that day as a predictor and the yield at the end of the season as a response. The training of the regression provides the coefficient estimates for each day and also information about the model fit (i.e. R^2) which is used to decide the day of the season (referred to as "regression date") that is more suited to predict yield because of the highest correlation between simulated VI and simulated yield. Once the date with the highest correlation is identified, the regression can be applied either to satellite-derived VI or - as our novel method proposes - to forecasted VI generated using a data assimilation method

(see Section 2.5 for more details). The regression can be applied on a pixel-by-pixel basis, which enables to obtain intra-field predictions of yield.
Differently from previous literature where the regression was created yearly considering meteorological data for that year (Sibley et al., 2014; Lobell et al., 2015; Jeffries et al., 2019), here we propose to use previous climate information (i.e. data from 2001-2017) to infer a regression that can be used to assess the reliability of the framework in later years (e.g. 2019). In real applications, weather is generally unknown for the full coming season, or if provided not trustworthy for a 6 months growth cycle.

2.5 Assimilating CubeSat LAI into APSIM using a particle filter

Particle filters (PFs) approximate the probability distribution of the model state using a number of particles. PFs are fully non-linear data assimilation techniques in which samples are created from the prior after which importance sampling is used to turn them into samples from the posterior where each sample is weighted with its likelihood value (Vetra-Carvalho et al., 2018). When an observation becomes available, its information content can be employed to compute the particle weights and a new weighted estimate of the model variable (i.e. LAI in our case) can be computed. This concept is referred to as sequential or Bayesian, importance sampling.
Let x_i^k be each LAI pixel value state of the system integrated forward in time and ***k*** be the time of the current observation, then the stochastic forward model for each particle ***i = 1….N*** is defined as:

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

where $\boldsymbol{M}$ is the deterministic crop model APSIM, β_i^k are random terms representing the Gaussian distributed model errors with mean zero and covariance matrix $\boldsymbol{Q}$. Now let x^m be the mean of $\boldsymbol{N}$ particles. The weight w_i^k for each LAI pixel value x_i^k simulated by the particle *i* at time ***k*** of the satellite acquisition is calculated 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 value (i.e. each pixel i of the satellite observation at time ***k***) x_o^k, with its standard deviation σ which quantifies the uncertainty of the observation, represents the discrete observation at a specific time step ***k***. Finally, all the weights $\mathrm{w_i^k}$ are normalized by the sum of the weights of all particles as:

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

The probability obtained with the weights allows to compute an expected weight E of the LAI values x_i^k called herein rs_{PF}, for the assimilation time-step k:

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

It should be noted that in Bayesian importance sampling, only the weights of the particles are updated but not the model state variables. The expected weights are assumed to be valid from time k until the time of a new observation where new weights are computed, keeping memory of the previous assimilation time steps.
An ensemble of particles was generated by running the APSIM model with different initial conditions. The latter were chosen based on previous literature (Lobell et al., 2015) and based on their influence on the model outputs (Ziliani, 2022). As for the generation of the ensemble in the regression step (see Section 2.4), input variables representing soil conditions and management practices were modified to generate the particles. In particular, 100 particles were generated by running the model with different soil characteristics, planting dates, and rates of fertilizer applied, whose ranges of variation were set based on typical farming operations at the ENREC facilities. Meteorological data for the year 2019 were used to run model simulations.

3. RESULTS

3.1 Remotely sensed LAI

The efficacy of the particle filter in generating good estimates of the model's state variable LAI and end-of-season yield is highly dependent on the quality of the observation assimilated into the model. As such, obtaining high-quality input data from satellites is paramount for driving the assimilation step. Moreover, to achieve intra-field predictions of yield, high spatial resolution images are needed.
This study uses a gap-free LAI product generated from a combination of Planet CubeSat imagery and the CESTEM algorithm to achieve a high spatiotemporal dataset (daily and at 3 meters resolution; see Section 2.3 for the description of the approach) starting from day of year (DOY) 100 to DOY 320 2019 (from April 10 till November 16).
The accuracy of the CubeSat-derived LAI product was assessed by comparing them against the *in situ* LAI collected measurements at the IMZs (see Figure 1 for the locations of the IMZs). Figure 2 displays the daily CubeSat LAI at the same location (i.e. latitude and longitude) of the IMZ 1. Generally speaking, the CubeSat observations well reproduce the LAI trend with the majority of the values being within the range of the *in situ* LAI and in proximity to their average values. The quick development of the leaf area index observed from DOY 170 and 200 is well reproduced by the CubeSat LAI.

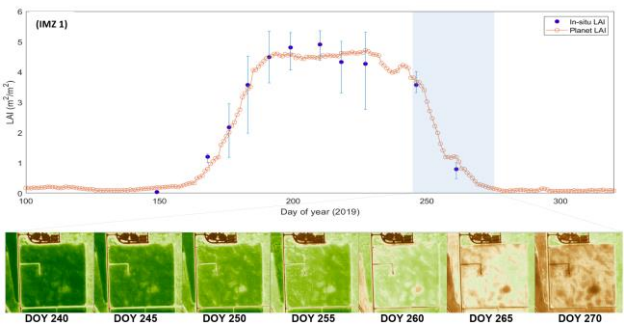

Figure 2. CubeSat-derived LAI values (orange dots) at the coordinate of the intensive management zone 1 and their comparison with the LAI *in situ* observations (blue bars). A sequence of CubeSat LAI maps at 3m is shown at the bottom.

During this rapid LAI growth phase, the plant stalk is characterized by substantial development, where leaf area and other plant components (i.e. canopy height) increase

dramatically (Ziliani et al., 2018). The first part of the reproductive stage (in this case from DOY 200 to 240), is characterized by an LAI response that remains more or less stable (with values around 4.5 m^2/m^2,). Following this flattened LAI period, the crop reaches physiological maturity by accumulating above-ground biomass while reducing leaf area.
Figure 2 (bottom) also shows an LAI sequence (every 5 days) where the 3 m spatial resolution of the CubeSat LAI product depicts the spatial heterogeneity within the US-Ne3 site; this time series corresponds to the LAI decreasing period from DOY 240 to 270. In this 30-day window, the field sees a rapid contraction in leaf area, which decreases from 4 m^2/m^2 to almost zero.
A recent study that assessed the performance of the CubeSat LAI product in the same area (Johansen et al., 2021) showed a correlation (R^2) of 0.92 between CubeSat LAI and *in situ* LAI with fairly low rRMSE and RMSE values (13 % and 0.4 m^2/m^2, respectively).

3.2 Estimating yield using a simple regression

The regression model (Eq. 1) was trained by running crop model simulations for the study years using daily weather data gathered from a local station and then applied to predict yield from CubeSat-derived LAI. From model simulations, it was identified when the LAI is the best predictor of yield, namely on DOY 212 (July 31), when a maximum R^2 of 0.72 was identified. The identified regression can then be applied on a pixel basis (i.e. to all pixels of the satellite map) to retrieve a yield map for the end of the season. Although this approach has shown some potential in predicting yield at large and smaller scales, it still relies on a satellite observation concident or close to the date of maximum correlation. Indeed, the regression can be applied to the satellite image only when this becomes available from the satellite overpass. To overcome this constraint, here we developed a procedure that uses a particle filter to assimilate CubeSat imagery into the model APSIM and then forecast the LAI map until the day of maximum correlation (DOY 212). Once the LAI forecast until DOY 212 is computed, the yield can be computed by applying the previously trained regression to the LAI forecast instead of the actual satellite LAI, permitting the prediction of yield in advance, even before the satellite image becomes available.

3.3 Assimilating LAI into APSIM with a particle filter

To assess the ability of the particle filter in predicting the LAI maps, different assimilation dates were evaluated to understand how long in advance is it possible to generate a reliable LAI map that can be used to predict the yield. Here we assume three different satellite acquisition dates: one, two, and three weeks prior to the regression date. Figure 3 (top) shows the assimilation results for the three weeks assimilation period.
The acquisition day of the CubeSat image (which corresponds to the assimilation day) is shown as a blue vertical line, while the regression date is depicted as a red vertical line. The results show that all the weighted mean LAI predictions (blue dots) clearly shift towards the mean of the observations (orange dots) after assimilation. Assimilating up to three weeks before the regression date ensures accurate LAI predictions throughout the forecast period. The root mean square error of the weighted mean against the observation is maintained below 0.4 m^2/m^2 if the assimilation is performed up to three weeks before the regression date.

The particle filter DA approach is carried out on a pixel-by-pixel basis allowing the forecast of spatially varying LAI maps for each day in which the data assimilation forecast step is performed. As such, an accurate forecast of LAI maps is key to properly estimating intra-field yield. Figure 3 (bottom) displays 21 days of LAI forecast (with a 3 days cadence) generated after the assimilation of one CubeSat image at DOY 191 (three weeks before the regression date) and compared against the corresponding CubeSat LAI observations. Overall, the LAI predictions show a high level of agreement when compared to the CubeSat LAI for almost all the forecast period, with a slightly less defined pattern visible only at the end of the forecast period (i.e. DOY 212).

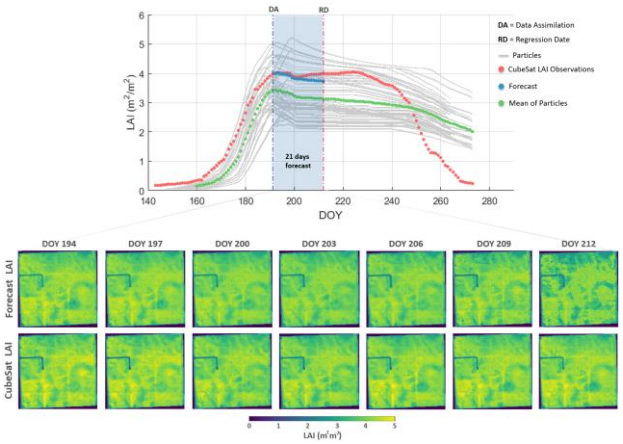

Figure 3. LAI particles (grey lines) produced by APSIM whose spread includes the CubeSat LAI observations (orange dots). The green dots depict the particles' standard mean, while the blue dots stand for the daily LAI forecast from the day of assimilation (DA) until the regression date (RD).

3.4 Assimilating CubeSat LAI accurately maps within-field yield

The results show that particle filter and regression-based yield estimates are accurate in reproducing the spatial variability of yield, with correlation values that range from 0.68 (results with the assimilating three weeks prior to the regression date) to 0.73 (results with the assimilating up to one week before the regression date). The correlation was computed on a pixel-by-pixel basis between the *in situ* yield and the predicted ones. The latter, at 3m resolution, were upscaled to the resolution of the measured yield (10m) to allow a consistent datasets comparison which comprises a total of 74,646 pixels. The predicted yield shown in Figure 4 demonstrates that the proposed approach reproduced fairly well the sub-field yield variations, suggesting that the estimates are accurate at the sub-field scale. Assimilating closer in time to the regression date (Figure 4, forecast of 7 days) produces more accurate sub-level yield predictions compared to earlier assimilations (2-3 weeks before the regression date), both in terms of captured spatial variations and in terms of actual yield values. Assimilating two and three weeks prior to the regression date still allows the prediction of reliable yield maps with sub-field variability that is similar to those of the measured yield, albeit with an R^2 decreasing to 0.7 (RMSE = 1.71) and 0.68 (RMSE = 1.82), respectively.
It can be noted that although the spatial variation of yield is fairly well replicated, all the predicted yields are unable to reproduce the lowest and highest values that are present *in situ* yield (such as the darker blue and red patches of the observed

yield in Figure 4). A possible explanation is that the assimilation is performed using CubeSat-derived LAI observations, which might not be the best predictor for end-of-season yield. Indeed, it appears that the index saturates (to a certain extent) the yield values within the field, and cannot reproduce the highest and lowest values.

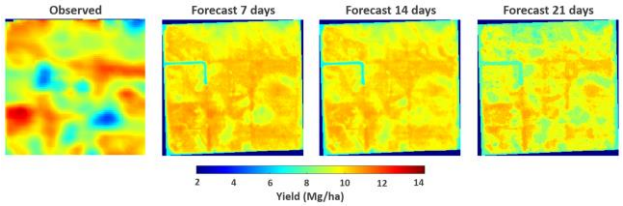

Figure 4. Observed against predicted yield after the assimilation of one satellite image 7, 14, and 21 days before the regression date.

4. DISCUSSION

Our results demonstrate how CubeSat data can be integrated into the APSIM model to improve yield prediction estimates. Due to the spatial resolution of the CubeSat data (3 m), the within-field variability of the maize field was well represented. Additionally, the high temporal resolution of the CubeSat constellation allows for daily and cloud-free data that can be readily integrated into APSIM or other crop models for yield prediction. The novelty of our approach lies in the assimilation using a particle filter, where only the particle weights are updated with additional observations, resulting in a new weighted estimate of LAI after one assimilation step. Figure 3 shows that the particles generated during the APSIM runs showed that the LAI dynamics were mostly parallel during the season. This behavior is a result of limiting the perturbations to the initial conditions (i.e., planting dates, soils, fertilizer amounts, and cultivar traits) and to the driving model parameters (i.e., climate conditions). Our results demonstrate that assimilating a single observation up to three weeks before the regression date led to accurate intra-field yield predictions. While only LAI was assimilated, it is important to note that integrating other vegetation indices or data from other remote sensing platforms (e.g., other satellites, remotely piloted aircraft) could strengthen the predictive power and provide further insights into the crop's development (Franz et al., 2020). Finally, even though the results of this research showed how to employ high spatiotemporal resolution CubeSat imagery for yield prediction without the need for ground calibration, further research is necessary to test the scalability of the proposed framework. Indeed, changing the environment or management practices could lead to variations in crop canopy, and consequently crop yield, which could limit the applicability of yield prediction approaches (Prasad et al., 2006).

5. CONCLUSION

This paper discussed the utility of using a particle filter as a tool to integrate high spatiotemporal resolution CubeSat LAI maps into the APSIM model for corn yield prediction. APSIM was used to derive a regression model that was subsequently applied to an LAI forecast after a single assimilation step. The results of this study are in good agreement with the measured yield. More importantly, the presented framework forgoes the need to employ on-site data to calibrate the APSIM-based regression model while still delivering accurate yield prediction results even with a single assimilation step. Future research should focus on further improving yield estimates with other remote sensing data, such as evapotranspiration. Furthermore, the influence of the LAI assimilation window on yield estimation should be evaluated.

REFERENCES

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 Sensing*, *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 3 m resolutions. *Scientific Reports*, *11*(1), 1-12.

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. *Bulletin of the American Meteorological Society*, *82*(11), 2415-2434.

Basso, B., Ritchie, J., Pierce, F., Braga, R., & Jones, J. (2001). Spatial validation of crop models for precision agriculture. *Agricultural Systems*, *68*(2), 97-112.

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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(1), 57-66.

Carberry, P., & Abrecht, D. (1991). Tailoring crop models to the semiarid tropics.

Carpenter, J., Clifford, P., & Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation*, *146*(1), 2-7.

Clevers, J. (1997). A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sensing of Environment*, *61*(2), 221-228.

Clevers, J. G., Kooistra, L., & Van den Brande, M. M. (2017). Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing*, *9*(5), 405.

Cordell, D., Drangert, J.-O., & White, S. (2009). The story of phosphorus: global food security and food for thought. *Global environmental change*, *19*(2), 292-305.

Edmeades, G., & Daynard, T. (1979). The relationship between final yield and photosynthesis at flowering in individual maize plants. *Canadian Journal of Plant Science*, *59*(3), 585-601.
Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, *53*(4), 343-367.

Fieuzal, R., Bustillo, V., Collado, D., & Dedieu, G. (2020). Combined use of multi-temporal Landsat-8 and sentinel-2 images for wheat yield estimates at the intra-plot spatial scale. *Agronomy*, *10*(3), 327.

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 Sensing*, *11*(9). doi: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 Research*, *252*, 107788.

Gilardelli, C., Stella, T., Confalonieri, R., Ranghetti, L., Campos-Taberner, M., García-Haro, F. J., & Boschetti, M. (2019). Downscaling rice yield simulation at sub-field scale using remotely sensed LAI data. *European journal of agronomy*, *103*, 108-116.

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. *Environmental Modelling & Software*, *62*, 327-350.

Hoogenboom, G. (2000). Contribution of agrometeorology to the simulation of crop production and its applications. *Agricultural and Forest Meteorology*, *103*(1-2), 137-157.

Houborg, R., & McCabe, M. (2018a). Daily Retrieval of NDVI and LAI at 3 m Resolution via the Fusion of CubeSat, Landsat, and MODIS Data. *Remote Sensing*, *10*(6), 890. doi:10.3390/rs10060890

Houborg, R., & McCabe, M. F. (2016). Adapting a regularized canopy reflectance model (REGFLEC) for the retrieval challenges of dryland agricultural systems. *Remote Sensing of Environment*, *186*, 105-120. doi:10.1016/j.rse.2016.08.017

Houborg, R., & McCabe, M. F. (2018b). A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sensing of Environment*, *209*, 211-226. doi:10.1016/j.rse.2018.02.067

Houborg, R., & McCabe, M. F. (2018c). A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, 173-188. doi:10.1016/j.isprsjprs.2017.10.004

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. *Agricultural and Forest Meteorology*, *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. *Agricultural and forest meteorology*, *204*, 106-121.

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. *Precision Agriculture*, 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. *European Journal of Agronomy*, *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. *Agricultural Water Management*, *227*, 105846.

Johansen, K., Ziliani, M. G., Houborg, R., Franz, T. E., & McCabe, M. F. (2021). Cubesat Constellations Provide Enhanced Crop Phenology And Digital Agricultural Insights Using Daily Leaf Area Index Retrievals.

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. *European Journal of Agronomy*, *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 Internatinal, 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. *European Journal of Agronomy*, *18*(3-4), 267-288.

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. *Agricultural and forest meteorology*, *256*, 445-457.

Li, Y., Zhou, Q., Zhou, J., Zhang, G., Chen, C., & Wang, J. (2014). Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. *Ecological modelling*, *291*, 15-27.

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, *143*, 56-64.

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, *164*, 324-333.

Massignam, A., Chapman, S., Hammer, G., & Fukai, S. (2009). Physiological determinants of maize and sunflower grain yield as affected by nitrogen supply. *Field Crops Research*, *113*(3), 256-267.

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. *Agricultural and forest meteorology*, *250*, 159-170.

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. *Agricultural and Forest Meteorology*, *213*, 160-172.

Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation, 8*(1), 26-33.

Resnick, D. (2020). Political economy of food system reform. *Nature Food, 1*(3), 154-154.

Ritchie, S. W., & Hanway, J. J. (1989). *How a corn plant develops*. Retrieved from

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. *Agronomy Journal, 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 Sensing, 9*(5), 509.

Suyker, A. E., & Verma, S. B. (2008). Interannual water vapor and energy exchange in an irrigated maize-based agroecosystem. *Agricultural and Forest Meteorology, 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.

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. *Agricultural and Forest Meteorology, 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: Dynamic Meteorology and Oceanography, 70*(1), 1-43.

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. *European Journal of Agronomy, 18*(1-2), 121-140.

Yu, Q., Wu, W., Yang, P., Li, Z., Xiong, W., & Tang, H. (2012). Proposing an interdisciplinary and cross-scale framework for global change and food security researches. *Agriculture, Ecosystems & Environment, 156*, 57-71.

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 Sensing, 11*(13), 1618.

Ziliani, M. G. (2022). *Predicting Crop Yield Using Crop Models and High-Resolution Remote Sensing Technologies.*

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 Sensing, 10*(12), 2007.