European Journal of Agronomy 143 (2023) 126718

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Efficiency of assimilating leaf area index into a soybean model to assess within-field yield variability

Deborah V. Gaso [a,d,*], Allard de Wit [b], Sytze de Bruin [a], Laila A. Puntel [c], Andres G. Berger [d], Lammert Kooistra [a]

[a] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research, Wageningen 6708 PB, the Netherlands
[b] Wageningen Environmental Research, Wageningen 6708 PB, the Netherlands
[c] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Keim Hall, 1825 N 38th Street, Lincoln 68583-0915, NE, USA
[d] Instituto Nacional de Investigación Agropecuaria, Colonia CP 70006, Uruguay

ARTICLE INFO

Keywords:
Yield prediction
Assimilation efficiency
Crop models
Sentinel-2
Soybean

ABSTRACT

Methods for accurately estimating within-field yield are essential to improve site-specific crop management and resource use efficiencies, which would be a major step toward sustainable intensification of agricultural systems. We set out to assess the accuracy of within-field soybean yields predicted by two data assimilation methods and to assess these methods' assimilation efficiency (AE). Yields were estimated by assimilating remotely sensed leaf area index (LAI) data from Sentinel-2 into a soybean crop growth model on a pixel basis. The LAI data was integrated into the model by Ensemble Kalman Filtering (EnKF) or by recalibrating with the Subplex algorithm (recalibration-based). An open-loop setting which only integrates information on the soil layers was used as a baseline scenario for quantifying the AE. We assessed both data assimilation techniques on eight fields (3067 pixels) in the Corn Belt region (Nebraska, Kansas and Kentucky) in the United States. The data set encompassed substantial variation in crop growth conditions: three growing seasons (2018, 2019 and 2020), rainfed and irrigated fields, and early and late planting dates. Ground truth yield acquired from combine monitors was used to validate the yield estimations. Agreement between predicted and observed yield at pixel level was two times higher for both data assimilation methods compared to the open-loop. The root mean square error (RMSE) was 476 kg.ha^{-1} (RRMSE of 10 %) in the recalibration-based method and 573 kg.ha^{-1} (RRMSE of 12 %) in the EnKF-based method. For both data assimilation methods, assimilating the LAI improved predictions for 68 % of the pixels. For a further 12 % of pixels, there was no accuracy improvement. For the remaining 20 %, AE was positive for one of the two assimilation methods. The high proportion of pixels with positive AE indicates the potential for overcoming the limitations in applying crop models at high spatial resolution by integrating a crop growth indicator. Assimilating an in-season indicator of crop growth (LAI) into a soybean model made it possible to adjust the simulation pathway, thereby greatly improving the accuracy of the yield estimations at the pixel level. This study elucidates the practical applications of data assimilation strategies for fine-scale within-field crop yield mapping.

1. Introduction

Spatially explicit crop yield data provides valuable information for decision-makers in various sectors, such as farmers, processors, crop insurance companies, and food trading agencies (Deines et al., 2021). Crop yield data is essential for site-specific crop management to improve resource use efficiencies toward sustainable intensification of agricultural systems (Ittersum et al., 2013; Lobell, 2013; Maestrini and Basso, 2018). Yield information is publicly available for various administrative levels, such as provinces, states, and countries, but yield information at the individual field level is rarely available to researchers (Sykuta, 2016). Aggregated yield information is too coarse to reveal the within-field yield heterogeneity that needs to be known for site-specific management (Lobell, 2013). Although it used to be difficult and costly to collect high-resolution yield data, yield monitoring technology has made it feasible to measure fine-scale yield variability. However,

* Corresponding author at: Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research, Wageningen 6708 PB, the Netherlands.
E-mail addresses: deborah.gasomelgar@wur.nl, dgaso@inia.org.uy (D.V. Gaso).

https://doi.org/10.1016/j.eja.2022.126718
Received 7 March 2022; Received in revised form 17 October 2022; Accepted 5 December 2022
Available online 22 December 2022

because of concerns about privacy, yield maps are rarely publicly available for research (Deines et al., 2021; Sykuta, 2016). Moreover, sophisticated yield monitoring technology is not yet universally used, and farmers typically invest little effort in collecting high quality yield data. For these reasons, developing robust methods for within-field yield estimation for large-scale applications remains a challenging task for agricultural research.

Crop models are valuable because they describe the interaction between crop traits, management, growth, and environmental factors. They have the potential to improve crop management because they can take account of the effect of management decisions taken weeks or months earlier. Moreover, models estimate important crop variables (yield, total biomass, canopy N content, etc.) that are otherwise difficult to assess, and they can evaluate the impact of agro-management decisions and weather conditions on these variables. However, using crop models at high spatial resolution to estimate within-field yield variability remains challenging (Deines et al., 2021; Hunt et al., 2019; Kayad et al., 2019; Novelli and Vuolo, 2019) because of the lack of high-resolution input data (i.e., soil parameters) required to run a crop model (Kasampalis et al., 2018) and models' limited ability to simulate the impact of yield-reducing factors (weeds, pests, disease, etc.) that are often present in farmers' fields. Biophysical crop variables derived from remote sensing provide spatial information on crop growth conditions which could overcome this challenge (Dorigo et al., 2007). In recent decades, a growing number of studies have investigated whether and how the incorporation of remotely sensed data into crop models can provide spatially explicit yield estimates (De Wit et al., 2012; Huang et al., 2019; Ines et al., 2013; Jin et al., 2018; Kang and Özdoğan, 2019).

Leaf area index (LAI) is commonly used to link remote sensing and crop models as it has a clear connection with the crop model processes and directly affects simulated crop growth. Two types of approach have been used to link remote sensing observations with crop models: variational and sequential. Variational methods attempt to fit model simulations to the observations by optimizing uncertain model parameters or initial condition. Sequential methods directly modify the state variables of the model based on the uncertainties relating to weighting the observed values and to the model simulations. The assumption underlying this method is that updating the model with observations at a certain moment in time will nudge the model toward the correct simulation pathway, and consequently will result in better simulations.

The increasing accessibility of high-resolution remote sensing products has enabled observations of biophysical variables to be integrated into crop models to produce fine-scale yield maps (Gaso et al., 2019, 2021). However, due to the lack of spatially explicit ground truth data, model estimation of within-field yield variability has often been validated by aggregating fine-grained estimations to the field level for which yield data was available (Kang and Özdoğan, 2019; Sibley et al., 2014; Silvestro et al., 2017). When high-resolution yield maps are aggregated to field level, the yield estimates are usually more accurate than the pixel level assessments, as local errors partly cancel out (Deines et al., 2021; Peng et al., 2021). Therefore, reliable evaluation of the ability of data assimilation techniques to improve model estimation of within-field yield heterogeneity requires fine-scale (pixel-based) assessments.

Another shortcoming in addition to the lack of reliable validation of simulated within-field yield variability is that only a few crops have been tested. Wheat and corn have mostly been used in studies for within-field yield estimation (Huang et al., 2019); limited work has been done for soybean. A recent study successfully applied the integration of LAI derived from Sentinel-2 into a soybean model for predicting within-field yield through a variational method (Gaso et al., 2021). That study highlighted that uncertainty in LAI simulations caused by the assimilate partitioning logic in the model adversely affected the accuracy of yield estimations. It was concluded that a sequential method that directly updates LAI from the remote sensing-based estimates could adjust LAI more efficiently and offer an alternative to reduce LAI uncertainty, in order to be able to more accurately predict within-field soybean yield.

The study described here had two main objectives. The first was to assess the accuracy of predicting within-field soybean yield variability for several fields of soybean in the Corn Belt of the central US and to do so using Ensemble Kalman Filtering (EnKF) and a recalibration technique through the Subplex algorithm to assimilate remote sensing-based LAI into the soybean growth model as presented in Gaso et al. (2021). The second objective was to quantify the assimilation efficiency (AE) by comparing the accuracy of both data assimilation methods against the open-loop set-up. By so doing, we would provide insight into the efficiency and reduction of errors that can result from assimilating Sentinel-2 derived LAI observations into a soybean crop growth model. We also wanted to assess the efficiency of using either a sequential or variational approach.

2. Materials and methods

2.1. Study sites and yield information

This study involved eight soybean fields (total area 276 ha, 3067 pixels) located in the Corn Belt region (Nebraska, Kansas and Kentucky) in the United States. Field size ranged between 16 ha and 85 ha. Six fields were rainfed, the other two were irrigated (Table 1). All fields were planted with genetically modified soybean varieties which had crop cycle lengths ranging from maturity group (MG) III to IV. Fields were managed according to optimal agronomic practices for the region, to minimize the influence of biotic stresses (weeds, insects, and diseases) and nutrient availability. The planting dates were within the optimal window for the region and ranged from the end of April to the beginning of June. In all fields, rows were 0.7 m apart.

Yield data was collected using a yield monitoring system mounted on harvesting machines. The harvesting machines used varied, so a correction was applied by removing outliers on the basis of their frequency and on the basis of the minimum and maximum biological yield limits (Sun et al., 2013). We calculated yield from the grain flow, harvester width, and the distance and time between consecutive geo-referenced points. We then averaged the geo-referenced points within each 30 × 30 m pixel. The number of geo-referenced points within each pixel ranged from 45 to 92, depending on the type of combine harvester.

2.2. Crop growth model

We used the soybean crop growth model presented in Gaso et al. (2021) as the starting point. This model applies the water use efficiency concept driven by crop transpiration and uses elements from existing

Table 1
Location of the study fields and soybean management.

Field	Geographic coordinates of field center	Size	Growing	Planting	Maturity	Irrigation
		(ha)	**season**	**date**	**Group**	**(mm)**
1	41.029° N, 97.292° W	85	2020	Apr 20	3.5	210
2	41.013° N, 97.278° W	19	2020	Apr 20	3.5	210
3	40.119° N, 95.542° W	21	2020	May 11	3	0
4	39.975° N, 95.525° W	33	2020	Apr 23	3	0
5	40.982° N, 96.438° W	42	2020	May 6	3	0
6	40.012° N, 95.478° W	22	2019	Apr 25	4	0
7	37.434° N, 87.428° W	38	2018	Jun 9	4.5	0
8	37.443° N, 87.438° W	16	2018	May 26	4.5	0

models to describe the growth and development of soybean. The model has been found to perform well when describing soybean growth and yield under water-limiting conditions (Gaso et al., 2021). It requires specification of a limited number of parameters, which makes it attractive for data assimilation in large-area applications where information on crop management and genotype parametrization is insufficiently detailed.

We modified several important functions of the model, using the available information on soybean in the study area. Using information on soybean root growth velocity in the study area (Ordóñez et al., 2018), we modified the equation that described root growth to better represent soybean root exploration under the environmental conditions of the Corn Belt. To assess total root growth, we applied equal weights to in-row and central-row root elongation as obtained from the Ordóñez et al. (2018) equations. The root growth was set to cease at 828 °C days for the in-row position and slightly later for the central-row position (909 °C d), based on the information provided by Ordóñez et al. (2018).

Soybean was planted in rows 0.7 m apart, which produces a so-called "clumped canopy" early in the season. This clumped canopy has a large impact on both light interception and transpiration. Canopy light interception in the original model is simulated by the standard Beer's law equation, which produces overestimations in the heterogeneous canopy. We therefore decided to introduce a clumping factor to deal with this heterogeneous canopy early in the season. This factor plays a role when LAI is below four by reducing the light interception. For further details on the implementation of those modifications and on the parameter values used for the study region, see (https://github.com/dgaso/Soybean_CornBelt).

2.3. Spatial input data

Daily weather data required by the model was retrieved from two sources: Daily Surface Weather Data (DAYMET) on a 1 km grid for North America (Thornton et al., 2020) and NASA Prediction of Worldwide Energy Resources (NASA-POWER). DAYMET provides gridded estimates of daily weather with higher resolution (1 km grid) than NASA-POWER (grid of 0.5° × 0.5° latitude and longitude). We downloaded weather data for the central point of each field. The minimum and maximum temperatures, vapor pressure, global solar radiation, and precipitation were acquired from DAYMET, the windspeed was obtained from NASA-POWER. In the case of irrigated fields, the amount of water applied at each irrigation event was added to the precipitation variable.

Soil hydraulic properties were derived from the 30-meter probabilistic soil series map of the contiguous United States (Chaney et al., 2016), hereafter referred to as POLARIS. POLARIS is an innovative database that provides fine-scale soil information; the data is freely available to be downloaded in tiles of 1° × 1° latitude and longitude. For each field, we downloaded the median values of sand, clay, and organic matter for each soil layer. The information on these layers was then used to compute field capacity and wilting point following the pedotransfer functions presented by Saxton and Rawls (2006). Since hydraulic properties were considered homogeneous throughout the soil profile, to obtain a unique wilting point and field capacity for the entire soil profile we computed an average weighted with the layer thickness.

2.4. Remote sensing data

All cloudless images from the two Sentinel-2 satellites (2A and 2B) Level 2A were used to compute the red edge chlorophyll index ($CI_{red\ edge}$). The $CI_{red\ edge}$ was chosen because it is advantageous for LAI estimation as it does not saturate at high-density canopies (Gitelson et al., 2003). Using the publicly available Google Earth Engine (GEE) data archive, we retrieved $CI_{red\ edge}$ for the entire growing season of soybean in the Corn Belt (end of April to the beginning of October). Level 2 A information was not available for fields 7 and 8 (growing season 2018), so images were downloaded with the Sentinelsat API and the atmospheric corrections were carried out with the sen2cor tool. The $CI_{red\ edge}$ was computed for the time series of Sentinel-2 images using NIR band (Band7 in Sentinel-2, centered at 782 nm) and red edge band (Band5 in Sentinel-2, centered at 704 nm)(Eq. 1). These spectral bands of Sentinel-2 are equivalent to the spectral ranges used in the study of Nguy-Robertson et al. (2012) (red edge: 703.8–713 nm and NIR: 771.3–786.3 nm).

$$CI_{red\ edge} = (Band7 \text{ in Sentinel-2 / } Band5 \text{ in Sentinel-2}) - 1 \quad (1)$$

The number of cloudless images varies, depending on the field, as can be seen in Fig. 1. The $CI_{red\ edge}$ calculation and the downloading for the region of interest were done with the geemap package within the Python environment. The $CI_{red\ edge}$ was resampled to the spatial resolution of the digital soil map from POLARIS (30 m). Using the GDALWarp function from the GDAL library we resampled with the average method. The satellite-based estimates of LAI were derived from the $CI_{red\ edge}$ through a unified algorithm for soybean and maize (Nguy-Robertson et al., 2012). The algorithm used was developed on a system under maize–soybean rotation in Nebraska. Fig. 1 provides an overview of the number of Sentinel-2 images available throughout the soybean growing season. Green and red represent the average LAI for the pixels; the 10% highest and lowest yield whiskers indicate the range of LAI variability within these zones. The results reveal that in most fields, the lower yield zones have a lower average LAI. However, in many fields the difference is very small and the whiskers often overlap. Higher standard deviations in the lowest yielding zones indicate that in these pixels the crop growth was more heterogeneous. More importantly, the differences in the maximum LAI for each field indicate that the environmental conditions for crop growth varied greatly between fields.

2.5. Data assimilation methods

The two assimilation techniques we assessed in this study were a sequential method and a variational method. Both have been amply used to integrate an observed quantity of a state variable into a crop model (Jin et al., 2018). We compared the assimilation techniques against a baseline scenario, defined as the open-loop set-up. For the variational method, we used a recalibration procedure that calibrates uncertain model parameters by means of the Subplex algorithm (Rowan, 1990). We hereafter refer to this method as recalibration-based. In the case of the sequential method, we implemented the EnKF, which is widely used (Evensen, 1992) to sequentially update state variables of a system when a new observation becomes available. We hereafter refer to this method as EnKF-based. Both methods were implemented within the Python Crop Simulation Environment (PCSE); the biophysical variable assimilated in the crop growth model was LAI.

The EnKF estimates the state of a system as the weighted average of the simulated state and the observed state, using weights derived from the uncertainties in the model and observations and expressing them through the Kalman gain. Uncertainties in the LAI observations were modeled in accordance with the error of the algorithm employed to predict LAI. This algorithm estimates the green LAI as a function of $CI_{red\ edge}$ with a root mean square error (RMSE) of 0.54 $m^2.m^2$, where maximum green LAI values are around 5.5 $m^2.m^2$ (Nguy-Robertson et al., 2012). Since the error in LAI estimates obtained from $CI_{red\ edge}$ increases with LAI, we decided to define the uncertainty in the observations as a proportion of the predicted LAI. We assumed the errors in LAI predictions to be non-systematic, and so the RMSE reported by Nguy-Robertson et al. (2012) was converted into a 10 % relative standard deviation. The leaf biomass state variable was also updated by the same proportion as the change in LAI, as both state variables are directly connected in the crop growth model through the specific leaf area (SLA). We decided to include the variables seed biomass and stem biomass, which are associated with LAI, in the state vector. This meant that the total aboveground production would then be updated as the sum of

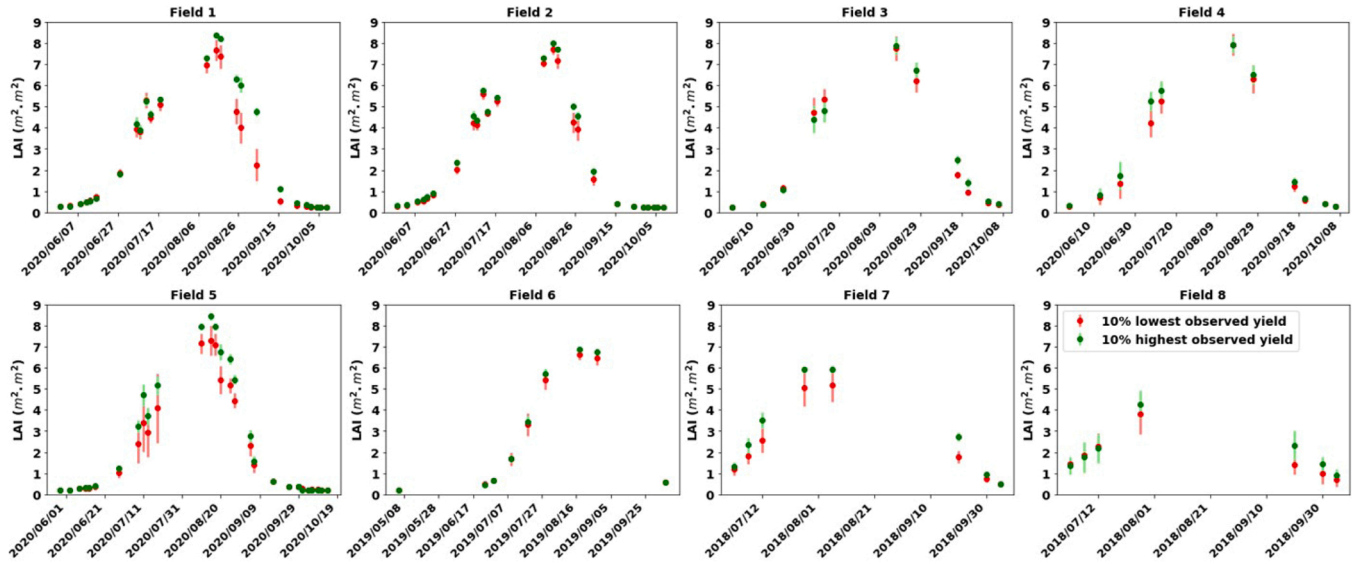

Fig. 1. Leaf area index (LAI) trajectory of the lowest and highest yielding zones of the fields as presented in Table 1: green dots indicate the 10% of pixels with the highest observed yield; red dots indicate the 10% pixels with the lowest observed yield. The number of available LAI values depends on availability of Sentinel-2 images. Error bars show the standard deviation of the LAI for each category of observed yield.

changes in leaf, stem, and seed biomass.

Our decisions on which parameters represent the uncertain part of the system were based on a sensitivity analysis performed with a variance-based method which assesses the response of the model output to the variations in the parameters. We used LAI as the target variable to assess model responses. The total variance in LAI was decomposed into fractions that can be attributed to a set of selected parameters. Four parameters were defined as the most uncertain part of the model (Table 2). These four parameters were modeled by Gaussian random variables with a mean equal to the default value and a standard deviation that is unraveled based on plausible values for each parameter. Table 2 specifies the mean and standard deviation for each uncertain parameter of the model. The parameter distribution, defined by the mean and standard deviation, is then used to draw a random value. Samples of uncertain observed LAI were also drawn from a Gaussian distribution centered on the LAI predicted from $CI_{\text{red edge}}$ with a 10 % standard deviation. A schematic overview of the methodology can be found in Curnel et al. (2011).

The number of ensemble members of the EnKF was set to 100, which has been reported as a good compromise between the minimization of the random component of the EnKF and the computational cost of the algorithm (Burgers et al., 1998). Since each pixel was considered as an independent observation, parallel processing is an obvious means to reduce the computational time. For this purpose, we used the multiprocessing package in Python.

The second data assimilation scheme was a recalibration method. This method performs the analysis through the minimization of a cost function, which implies that uncertain model parameters must be recalibrated. For the recalibration, we used the Subplex procedure equivalent to the framework described in Gaso et al. (2021) through the NLOpt library. Our cost function concerned the RMSE of LAI. For further details of the recalibration with the Subplex algorithm, see Steven (2020). The uncertain model parameters chosen for recalibration are listed in Table 2. The implementation of the Subplex algorithm in NLOpt explicitly supports bound constraints. We used plausible values for the upper and lower bounds of the parameter chosen for recalibration. See Table 2. As the same set of uncertain parameters is considered in both data assimilation methods, the methods can be compared on the same basis. The open-loop runs used the mean value of each uncertain parameter (Table 2), which corresponds to the default value.

Table 2
Selected uncertain model parameters and their mean and standard deviation values chosen as inputs for the EnKF, and the lower and upper bounds set for recalibration.

ID	Description	Unit	Upper bound	Lower bound	Mean	SD
LAIinit	Initial leaf area index	m^2. m^2	0.34	0.1	0.22	0.04
WUE	Water use efficiency	mbar	0.03	0.05	0.04	0.003
FNTR	Nitrogen translocated	%	50	25	33	5.5
RDMAX	Maximum root depth	m	1.5	0.9	1.2	0.1

2.6. Assimilation efficiency assessment

We first evaluated the accuracy of the yield estimates from the two assimilation methods and compared this to an open-loop approach by standard error metrics, including the mean error (ME), RMSE, and relative RMSE (RRMSE). We then assessed the added value of both assimilation techniques by means of the assimilation efficiency index (AE, Curnel et al. (2011). The AE was computed from the relative absolute error values estimated for the situation with (RAE_{DA}, Eq. 2) and without assimilation (RAE_{OL}, Eq. 3), as is formulated in Eq. 4. The relative absolute error (RAE) with assimilation uses simulations from EnKF-based or recalibration-based assimilation techniques (Eq. 2), while RAE without assimilation uses simulations from the open-loop situation (Eq. 3). The AE and RAE were computed for each pair of observed and simulated yield values, as is indicated with subscript j in Eqs. 2, 3, and 4.

$$RAE_{DA}(j) = \frac{|Y_{obs(j)} - Y_{DA(j)}|}{Y_{obs(j)}} \quad (2)$$

$$RAE_{OL}(j) = \frac{|Y_{obs(j)} - Y_{OL(j)}|}{Y_{obs(j)}} \quad (3)$$

$$AE(j) = 100 * \left(1 - \frac{RAE_{DA(j)}}{RAE_{OL(j)}}\right) \quad (4)$$

where $RAE_{DA}(j)$ and $RAE_{OL}(j)$ represent the relative absolute error with and without assimilation, respectively, for the interaction (pixel) j. Y_{obs}(j) is the observed yield, $Y_{OL}(j)$ is estimated yield with open-loop settings in the pixel j, $Y_{DA}(j)$ is estimated yield with data assimilation (EnKF-based or recalibration-based) in the pixel j, and AE(j) is assimilation efficiency index in the pixel j. For the EnKF-based method, the estimated yield is the mean of the ensemble yields.

For each of the study fields, we then computed the mean AE by using the Winsorized mean (Dixon and Yuen, 1974) of the total AE of the field. The Winsorized mean limits the effect of outliers, as it is the arithmetic mean after replacing the smallest and largest values. This method was chosen to avoid spurious outliers in the estimation of AE which could

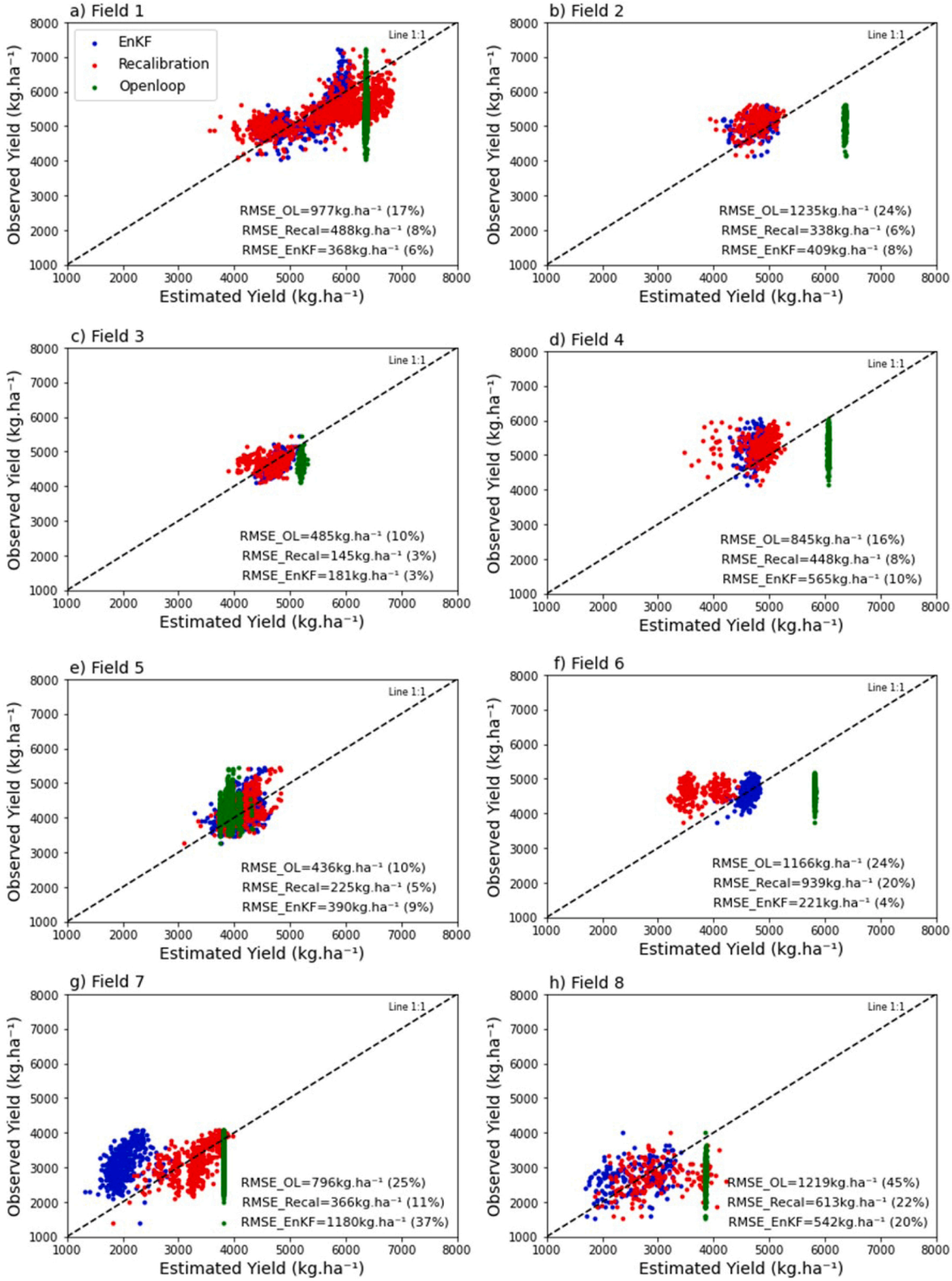

Fig. 2. Estimated yield compared with observed yield at pixel basis (aggregated to 30 m). RMSE_OL, RMSE_Recal, and RMSE_EnKF are the root mean square errors of the open-loop, recalibration, and EnKF, respectively. The numbers in parentheses are the relative root mean square errors expressed as percentages.

lead to unrealistic estimation of the mean AE values. The Winsorized mean was calculated with the Winsorize function from the Scipy library. The limits were set to the 0.1 and 0.9 quantiles.

3. Results

3.1. Evaluation of the yield estimates with the EnKF-based and recalibration-based techniques

Our dataset of ground yield observations comprised a diversity of environments as a result of combinations of weather, soil, crop, and management. The open-loop simulations (scenario without assimilation) captured the yield potential of the environment and ranged from ~4000 kg.ha^{-1} to ~ 6000 kg.ha^{-1} (Fig. 2). The effect of a late planting date on the crop yield is apparent from the lower yields for fields 7 and 8 as predicted by the soybean model (Fig. 2). Model overestimations with the open-loop simulations were consistent: the ME ranged from 513 to 1240 kg.ha^{-1} in seven of the eight fields. The exception was field 5, where ME was − 326 kg.ha^{-1}.

Fig. 2 shows a persistent disagreement between observed yields and those obtained by the open-loop simulations that only take account of the soil data. The spatial yield variability was insufficiently represented by the open-loop simulations, as the within-field variability in the soil hydrological parameter values derived from POLARIS (see Figs. S1 to S8) do not match the spatial variability in crop yields. Across all fields except field 5, the open-loop runs presented a narrow range of variability, inconsistent with the within-field observed yields (Figs. 2 and 3). The overall ME, RMSE, and RRMSE of the simulated yields in the open-loop simulations were 704 kg.ha^{-1}, 901 kg.ha^{-1}, and 19 %, respectively.

The EnKF-based and recalibration-based LAI data assimilations were able to reduce the RRMSE of predicted yields at pixel level for most fields, with one exception: the EnKF-based predictions in field 7 (Fig. 2). Across all datasets, the overall ME, RMSE, and RRMSE of the simulated yield in the EnKF-based assimilation method were − 213 kg.ha^{-1}, 573 kg.ha^{-1}, and 12 %, respectively. The recalibration-based assimilation method performed slightly better, with overall ME = −63 kg.ha^{-1}, RMSE = 476 kg.ha^{-1}, and RRMSE = 10 %. The data assimilation methods performed less accurately in fields 6, 7, and 8, where due to the lack of cloud-free images less information was available for assessing the dynamics of crop senescence (fields 6, 7, and 8 in Fig. 1).

Fig. 3 shows the simulated yield maps for field 1 (one of the irrigated fields). The potential yield obtained using the open-loop method was substantially higher than the observed yield. The simulated yield maps generated by both assimilation methods (Fig. 3a and b) partly reproduced the yield variation within the central pivot irrigated field, as can be seen by comparing the patterns of dark blue (low yields) and yellow (high yields) pixels in maps a and b with the observed yields in map d. The assimilation of LAI amplified the yield variation within the pivot.

3.2. Assimilation efficiency assessment

The AE index is an estimate of the average amount by which errors in estimated yield have been reduced. A positive AE indicates that the assimilation method enables improvement in yield estimates, while a negative value means that the errors in the estimated yield are higher than the situation without assimilation (open-loop simulations). In all but two cases, the mean AE value was positive across the study fields. The exceptions were for the EnKF-based method in fields 5 and 7 (Fig. 4b). In field 5, open-loop simulations mostly overlap with the

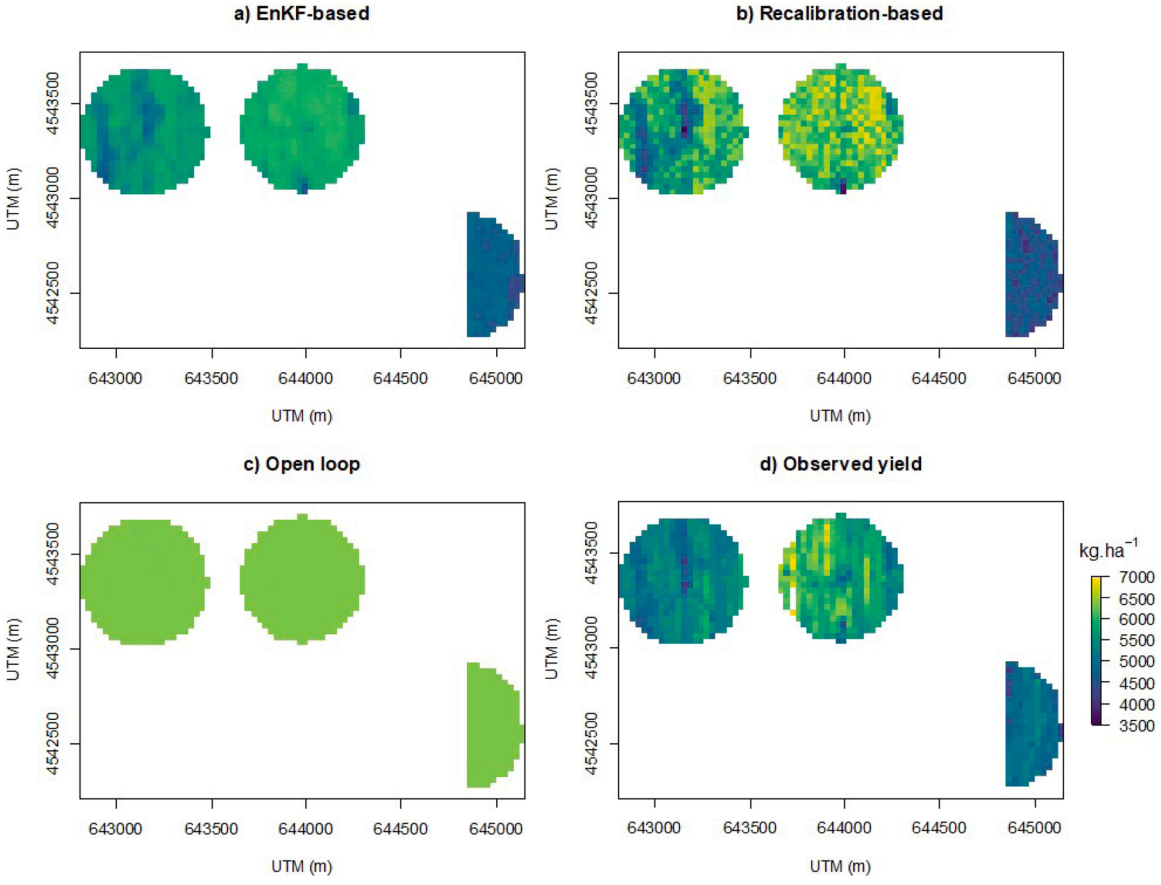

Fig. 3. Maps of observed and simulated soybean yield obtained using the EnKF-based, recalibration-based, and open-loop runs for field 1 with 944 pixels at 30 m spatial resolution.

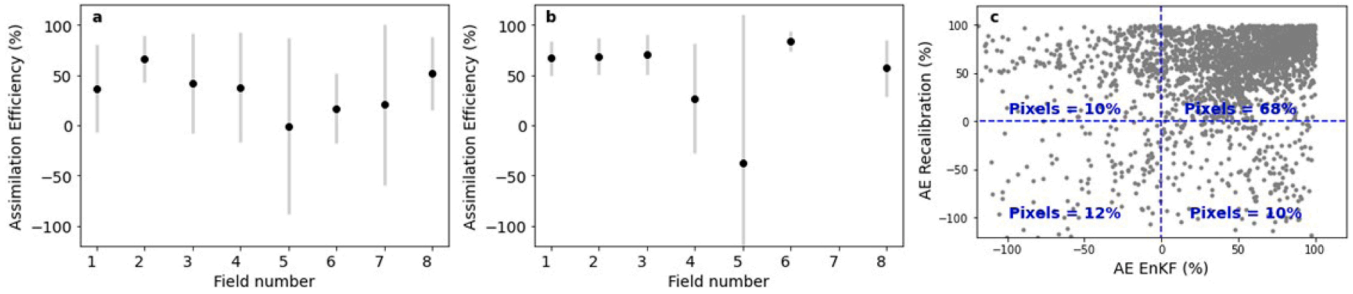

Fig. 4. Means (black dots) and standard deviations (vertical bars) of the assimilation efficiency index (AE) of (a) the recalibration-based method, (b) the EnKF-based method, for eight soybean fields. Field 7 was excluded from panel b for visualization (mean and standard deviation of AE of field 7 in EnKF were −393 % and 675 %, respectively). Panel c represents the AE value of each method per pixel. Values within panel c are the proportion of pixels in each quadrant.

simulations from LAI assimilation, which lead to extreme values for AE. It should be noted from the AE formulation that when RAE_{DA} and RAE_{OL} are low (as in field 5), AE index becomes extremely low or high, and consequently, the standard deviation is higher than in other fields (Fig. 4a and b). For 67 % of the pixels in field 7, the RAE_{OL} (denominator in Eq. 4) is lower than RAE_{DA} in the EnKF-based assimilation, resulting in an extremely low average AE value for this field: − 393 %.

Assimilating LAI data resulted in an appreciable improvement in the accuracy of yield estimations: in six of the eight fields, the average AE exceeded 20 % (Fig. 4a and b). Overall, 68 % of the points (2071 pixels) had positive AE value in both data assimilation methods (Fig. 4c), while only 12 % of the points (362 pixels) had negative AE in both data assimilation methods. In the remaining 20% of the points, AE was only positive in one of the assimilation methods: see Fig. 4c.

4. Discussion

Using the recalibration-based and EnKF-based methods to assimilate LAI into a soybean growth model we were able to estimate within-field spatial variability in soybean yield with an average error of − 63 kg.ha^{-1} (SD 547 kg.ha^{-1}) for the recalibration-based method and − 213 kg.ha^{-1} (SD 532 kg.ha^{-1}) for the EnKF-based method. The results demonstrate that for all eight fields in our study, the LAI assimilation was able to reduce bias in the estimated yield compared to the open-loop scenario. Moreover, assimilating the LAI also improved the estimation of within-field variability of crop yield for both methods, the AE was positive in 68 % of the pixels but was negative in only 12 % of pixels. No improvement was apparent in field 5: here, the open-loop simulations are in the same range as the EnKF-based and recalibration-based simulations. A plausible explanation is that in this field, the water stress (crop transpiration - potential transpiration) during the reproductive stage was substantially greater than in the other fields (see Figs. S9 to S16), which may have contributed to the limited success of LAI assimilation in nudging the model toward a more accurate simulation pathway. In the other seven fields, the success of the assimilation methods suggests that the impact of reducing and limiting factors not included in the simulation model could be mitigated by LAI assimilation. Both data assimilation techniques were able to improve the accuracy of yield estimates, but the recalibration-based strategy slightly outperformed the EnKF-based one, as the overall RRMSE was 10 % for the recalibration, 12 % for the EnKF, and 19 % for the open loop.

The improvements in yield estimation obtained highlight the ability of the LAI assimilation to overcome the lack of spatial variation in soil properties and to capture the impact of yield-reducing factors not included in the model. This is supported by the observations that the impact of the LAI assimilation decreased when the model deviated from the LAI trajectory because there were insufficient images for the senescence period (Fig. 1). For instance, the differences in the senescence rate of field 1 between zones in which yields varied (Fig. 1) accounted for the spatial yield variation in both data assimilation methods, which was in accordance with the observed yield variability (Fig. 2a). The lack of LAI observations during the senescence phase in fields 6, 7, and 8 (Fig. 1) was associated with lower accuracy in both data assimilation methods (Fig. 2f, g, and h). In the latter cases, the EnKF-based strategy was unable to directly adjust the model state during the senescence phase because there were too few observations, while the recalibration strategy could not accurately estimate the nitrogen translocation rate (FNTR, Table 2).

These results agree with the findings from Silvestro et al. (2021), who demonstrated that the EnKF-based method was suitable for assimilating LAI into the SAFYswb model as it performed well when adjusting maize yield for field conditions. The results from Dhakar et al. (2022) and Kang and Özdoğan (2019) also indicate that assimilation of satellite-derived LAI estimates can be used to resolve yield variability within and between fields. However, these two studies also indicate that for the EnKF to perform well, it is crucial to have a carefully calibrated model. In particular, a bias in phenology simulations leading to a "phenological shift" (Curnel et al., 2011) will adversely affect the performance of the EnKF. A similar concern applies to our study, as the yield formation by our model is simulated through partitioning functions that are highly dependent on a correct simulation of crop phenology (Gaso et al., 2021).

In this study, we demonstrated that the EnKF-based and recalibration-based strategies have similar performance in reducing the error of the simulated yield compared to the open-loop simulation. However, there are important differences between the two strategies. The EnKF directly adjusts the model state variables in order to improve the simulated yield. Although this strategy is effective and computationally efficient, it provides little insight into the drivers behind the within-field variability. The recalibration strategy has the advantage that it provides maps of adjusted parameter values which have a biophysical interpretation. For example, the LAIinit parameter (Table 2) can be interpreted as a low planting density that impacts the yield of soybean growing under low-yield conditions (Carciochi et al., 2019). Similarly, variations in water use efficiency (WUE, Table 2) and the percentage of translocatable nitrogen (FNTR, Table 2) point at deficiencies in crop growth that could probably be tackled by adjusting crop management.

An advantage of the EnKF over the current recalibration strategy is that the EnKF lends itself better to real-time inferences of crop status and yield estimation (Huang et al., 2019). Our recalibration strategy instead requires a full crop cycle to estimate the model parameters, so the strategy is only useful for retrospective analysis. However, given that the recalibration strategy has some advantages over the EnKF, we argue that extending the objective function with prior information on parameters would make the recalibration strategy more suitable for real-time inferences too. It essentially becomes a 2DVAR algorithm, like the one used by Zhuo et al. (2022) to assimilate LAI for within-season estimation of winter wheat yields.

Spatially explicit crop yield data is also an essential requirement for

quantifying the within-field yield gap and targeting crop management actions to reduce that gap. Our results demonstrate that the assimilation of LAI estimates from Sentinel-2 observations can be used to estimate a spatially explicit yield gap, and that therefore this strategy has potential to assist in improving soybean crop management. The feasibility of this approach still needs to be tested over more years and regions, and the results need to be discussed with farmers to assess if they recognize the spatial variability in parameters and whether it can be related to management actions.

One of the parameters in the model defines the percentage of translocated nitrogen that is clearly related to soybean nitrogen content, which in turn is related to soybean protein content (Cafaro La Menza et al., 2017; Ciampitti and Salvagiotti, 2018). It still needs to be demonstrated that such parameter maps can predict the spatial variability in soybean protein content. However, if the feasibility is demonstrated, a potential application is to use the parameter maps to plan differential harvesting of soybean (Kravchenko and Bullock, 2002). The study by Ciampitti et al. (2021) showed that high-yielding environments rely more on soil N supply, and that in this situation within-field variability in protein content is more likely. Similarly, the meta-analysis by Assefa et al. (2019) found high variability in the protein content of soybean, which is an important issue for growers and industry. Thus, differential harvesting from the areas in the fields where soybeans have higher protein or oil concentrations would be beneficial for producers and buyers as it would allow produce to be graded more easily (William et al., 2020).

Another potential application is to use the parameter maps to estimate the spatial variability in the amount of N extracted from the system. Soybean contributes nearly to 25 % of the total fixed N in agricultural systems (Herridge et al., 2008) where, on average, biological N_2 fixation represents 50–60 % of the crop N demands and the partial N balance (fixed N in aboveground biomass – N seeds) is negative in 80% of the cases (Ciampitti and Salvagiotti, 2018). Such partial N balance plays a major role in the sustainability of the agricultural systems. Estimates of the partial N balance could assist the development of variable rate fertilizer strategies that replenish nutrients in the agricultural systems instead of fertilizing each individual crop.

A prerequisite for successfully applying data assimilation is a model that is well calibrated, particularly for the phenological development of soybean. The phenology model used for this study (Setiyono et al., 2007) is a generic model based on the MG classification of the soybean variety. Although MG ratings of soybean varieties are not always well standardized, such models have demonstrated enough predictive ability to simulate phenology stages, as their accuracy was equivalent to the simulations based upon genotype-specific parameters (Salmerón and Purcell, 2016). Another study (Archontoulis et al., 2014) also showed the feasibility of obtaining accurate estimations of flowering and maturity stage by using generic parameters that depend on MG. Thus, phenological models that account for temperature and photoperiod interactions throughout the crop cycle and that rely on the MG classification and avoid dependency on genotype-specific parameters can predict soybean stages sufficiently accurately. Moreover, they are necessary to broaden the applicability of the crop simulation models.

In addition to a calibrated model, satellite observations are required to capture the within-field spatial variability of soybean LAI. For this study the Sentinel-2 observations were resampled from 20 to 30 m to match the POLARIS digital soil map. However, Skakun et al. (2021) caution that a moderate spatial resolution (20–30 m) does not fully capture the within-field yield variability. They concluded that moving the spatial resolution from 3 to 30 m results in an important reduction of yield variance. Similarly, Yang (2020) concluded that a 30 m spatial resolution can be used for variable rate application task maps, but that 5 m resolution imagery would be more appropriate for variable rate application in precision agriculture. Testing the technology developed with this study at such resolutions will make it even more useful for site-specific crop management. However, running a coupled crop model–data assimilation framework at such high spatial resolution is computationally daunting and will probably require techniques from software engineering and machine learning for deriving efficient meta models.

5. Conclusions

We have demonstrated that the assimilation of LAI derived from satellite observations into a soybean growth model led to improvements in the accuracy of yield estimations at pixel level. The recalibration-based approach slightly outperformed the EnKF-based one (RRMSE 10% vs 12 %). Compared to the baseline scenario (open-loop settings), both data assimilation methods reduced the RMSE of yield estimates by an average of 42 %. Our study has demonstrated the impact of assimilating LAI to compensate for the lack of spatially explicit input data and for growth-reducing factors not being accounted for by the model. Moreover, the quantitative assessment of the efficiency of assimilating LAI revealed an appreciable improvement in accuracy, which highlights the relevance of integrating a proxy for crop growth to adjust model simulations toward a correct pathway. Therefore, these results provide important insights to assist practical applications of the data assimilation methods for obtaining high-resolution within-field yield maps for soybean.

CRediT authorship contribution statement

Deborah V Gaso: Conceptualization, Formal analysis, Writing – review & editing. **Allard de Wit**: Conceptualization, Writing – review & editing. **Sytze de Bruin**: Conceptualization, review & editing. **Laila A. Puntel**: Conceptualization, Writing - review & editing. **Andres G. Berger:** Review & editing **Lammert Kooistra**: Conceptualization, review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This research was funded by the Instituto Nacional de Investigación Agropecuaria de Uruguay (INIA) and a Ph.D. fellowship provided by Agencia Nacional de Investigación e Innovación (ANII, scholarship code: POS_EXT_2017_1_147121). We would like to thank Laura J. Thompson and Joe Luck from the Department of Agronomy and Horticulture of the University of Nebraska-Lincoln for sharing the field data and providing support on the analysis. Joy Burrough language edited a near-final draft of the paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.eja.2022.126718.

References

Archontoulis, S.V., Miguez, F.E., Moore, K.J., 2014. A methodology and an optimization tool to calibrate phenology of short-day species included in the APSIM PLANT model: application to soybean. Environ. Model. Softw. 62, 465–477. https://doi.org/10.1016/j.envsoft.2014.04.009.

Assefa, Y., Purcell, L.C., Salmeron, M., Naeve, S., Casteel, S.N., Kovács, P., Archontoulis, S., Licht, M., Below, F., Kandel, H., Lindsey, L.E., Gaska, J., Conley, S.,

Shapiro, C., Orlowski, J.M., Golden, B.R., Kaur, G., Singh, M., Thelen, K., Laurenz, R., Davidson, D., Ciampitti, I.A., 2019. Assessing variation in us soybean seed composition (protein and oil). Front. Plant Sci. 10 https://doi.org/10.3389/fpls.2019.00298.

Burgers, G., Van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126, 1719–1724. https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Cafaro La Menza, N., Monzon, J.P., Specht, J.E., Grassini, P., 2017. Is soybean yield limited by nitrogen supply? Field Crop. Res. 213, 204–212. https://doi.org/10.1016/j.fcr.2017.08.009.

Carciochi, W.D., Schwalbert, R., Andrade, F.H., Corassa, G.M., Carter, P., Gaspar, A.P., Schmidt, J., Ciampitti, I.A., 2019. Soybean seed yield response to plant density by yield environment in north america. Agron. J. 111, 1923–1932. https://doi.org/10.2134/agronj2018.10.0635.

Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C. W., Odgers, N.P., 2016. POLARIS: a 30-meter probabilistic soil series map of the contiguous United States. Geoderma 274, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025.

Ciampitti, I.A., Salvagiotti, F., 2018. New insights into soybean biological nitrogen fixation. Agron. J. 110, 1185–1196. https://doi.org/10.2134/agronj2017.06.0348.

Ciampitti, I.A., de Borja Reis, A.F., Córdova, S.C., Castellano, M.J., Archontoulis, S.V., Correndo, A.A., Antunes De Almeida, L.F., Moro Rosso, L.H., 2021. Revisiting biological nitrogen fixation dynamics in soybeans. Front. Plant Sci. 12, 1–11. https://doi.org/10.3389/fpls.2021.727021.

Curnel, Y., de Wit, A.J.W., Duveiller, G., Defourny, P., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. Agric. For. Meteorol. 151, 1843–1855. https://doi.org/10.1016/j.agrformet.2011.08.002.

De Wit, A., Duveiller, G., Defourny, P., 2012. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. Agric. For. Meteorol. 164, 39–52. https://doi.org/10.1016/j.agrformet.2012.04.011.

Deines, J.M., Patel, R., Liang, S.Z., Dado, W., Lobell, D.B., 2021. A million kernels of truth: insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US corn belt. Remote Sens. Environ. 253, 112174 https://doi.org/10.1016/j.rse.2020.112174.

Dhakar, R., Sehgal, V.K., Chakraborty, D., Sahoo, R.N., Mukherjee, J., Ines, A.V.M., Kumar, S.N., Shirsath, P.B., Roy, S.B., 2022. Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing. Agric. Syst. 195, 103299 https://doi.org/10.1016/j.agsy.2021.103299.

Dixon, W.J., Yuen, K.K., 1974. Trimming and winsorization: a review. Stat. Hefte 15, 157–170. https://doi.org/10.1007/BF02922904.

Dorigo, W.A., Zurita-milla, R., Wit, A.J.W. De, Brazile, J., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling 9, 165–193. https://doi.org/10.1016/j.jag.2006.05.003.

Evensen, G., 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. J. Geophys. Res. 97 https://doi.org/10.1029/92jc01972.

Gaso, D., Berger, A., Ciganda, V., 2019. Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. Comput. Electron. Agric. 159, 75–83. https://doi.org/10.1016/j.compag.2019.02.026.

Gaso, D.V., de Wit, A., Berger, A.G., Kooistra, L., 2021. Predicting within-field soybean yield variability by coupling Sentinel-2 leaf area index with a crop growth model. Agric. For. Meteorol. 308–309, 108553 https://doi.org/10.1016/j.agrformet.2021.108553.

Gitelson, A.A., Vina, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30, 4–7. https://doi.org/10.1029/2002gl016450.

Herridge, D.F., Peoples, M.B., Boddey, R.M., 2008. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311, 1–18. https://doi.org/10.1007/s11104-008-9668-3.

Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.H., Wu, Y., Zhao, F., Wang, J., Xie, X., 2019. Assimilation of remote sensing into crop growth models: current status and perspectives. Agric. For. Meteorol. 276–277, 107609 https://doi.org/10.1016/j.agrformet.2019.06.008.

Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using sentinel-2. Remote Sens. Environ. 233, 111410 https://doi.org/10.1016/j.rse.2019.111410.

Ines, A.V.M., 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. https://doi.org/10.1016/j.rse.2013.07.018.

Ittersum, M.K.Van, Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Field crops research yield gap analysis with local to global relevance – a review. Field Crop. Res 143, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009.

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. https://doi.org/10.1016/j.eja.2017.11.002.

Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. Remote Sens. Environ. 228, 144–163. https://doi.org/10.1016/j.rse.2019.04.005.

Kasampalis, D.A., Alexandridis, T.K., Deva, C., Challinor, A., Moshou, D., Zalidis, G., 2018. Contribution of remote sensing on crop models: a review. J. Imaging 4. https://doi.org/10.3390/jimaging4040052.

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., Pirotti, F., 2019. Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques. Remote Sens. 11 https://doi.org/10.3390/rs11232873.

Kravchenko, A.N., Bullock, D.G., 2002. Spatial variability of soybean quality data as a function of field topography: I. Spatial data analysis. Crop Sci. 42, 804–815. https://doi.org/10.2135/cropsci2002.8040.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. Field Crop. Res. 143, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008.

Maestrini, B., Basso, B., 2018. Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. Sci. Rep. 8, 1–9. https://doi.org/10.1038/s41598-018-32779-3.

Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., Rundquist, D., 2012. Green leaf area index estimation in maize and soybean: combining vegetation indices to achieve maximal sensitivity. Agron. J. 104, 1336–1347. https://doi.org/10.2134/agronj2012.0065.

Novelli, F., Vuolo, F., 2019. Assimilation of sentinel-2 leaf area index data into a physically-based crop growth model for yield estimation. Agronomy 9. https://doi.org/10.3390/agronomy9050255.

Ordóñez, R.A., Castellano, M.J., Hatfield, J.L., Helmers, M.J., Licht, M.A., Liebman, M., Dietzel, R., Martinez-Feria, R., Iqbal, J., Puntel, L.A., Córdova, S.C., Togliatti, K., Wright, E.E., Archontoulis, S.V., 2018. Maize and soybean root front velocity and maximum depth in Iowa, USA. Field Crop. Res. 215, 122–131. https://doi.org/10.1016/j.fcr.2017.09.003.

Peng, X., Han, W., Ao, J., Wang, Y., 2021. Assimilation of lai derived from UAV multispectral data into the safy model to estimate maize yield. Remote Sens. 13, 1–17. https://doi.org/10.3390/rs13061094.

Rowan, T.H., 1990. Functional stability analysis of numerical algorithms. The University of Texas at Austin. Thesis Diss. 218.

Salmerón, M., Purcell, L.C., 2016. Simplifying the prediction of phenology with the DSSAT-CROPGRO-soybean model based on relative maturity group and determinacy. Agric. Syst. 148, 178–187. https://doi.org/10.1016/j.agsy.2016.07.016.

Saxton, K.E., Rawls, W.J., 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. Soil Sci. Soc. Am. J. 70, 1569–1578. https://doi.org/10.2136/sssaj2005.0117.

Setiyono, T.D., Bastidas, A.M., Weiss, A., Cassman, K.G., Specht, J.E., Dobermann, A., 2007. Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield co. Field Crop. Res. 100, 257–271. https://doi.org/10.1016/j.fcr.2006.07.011.

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, 24–32. https://doi.org/10.2134/agronj2013.0314.

Silvestro, P.C., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., 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. 9, 1–24. https://doi.org/10.3390/rs9050509.

Silvestro, P.C., Casa, R., Hanuš, J., Koetz, B., Rascher, U., Schuettemeyer, D., Siegmann, B., Skokovic, D., Sobrino, J., Tudoroiu, M., 2021. Synergistic use of multispectral data and crop growth modelling for spatial and temporal evapotranspiration estimations. Remote Sens. 13, 1–25. https://doi.org/10.3390/rs13112138.

Skakun, S., Kalecinski, N.I., Brown, M.G.L., Johnson, D.M., Vermote, E.F., Roger, J.C., Franch, B., 2021. Assessing within-field corn and soybean yield variability from worldview-3, planet, sentinel-2, and landsat 8 satellite imagery. Remote Sens. 13, 1–18. https://doi.org/10.3390/rs13050872.

Steven, G.J., 2020. The NLopt nonlinear-optimization package, ⟨http://github.com/stevengj/nlopt⟩.

Sun, W., Whelan, B., McBratney, A.B., Minasny, B., 2013. An integrated framework for software to provide yield data cleaning and estimation of an opportunity index for site-specific crop management. Precis. Agric. 14, 376–391. https://doi.org/10.1007/s11119-012-9300-7.

Sykuta, M.E., 2016. Big data in agriculture: property rights, privacy and competition in Ag data services. Int. Food Agribus. Manag. Rev. 19, 57–74. https://doi.org/10.22004/ag.econ.240696.

Thornton, M.M., Shrestha, R., Wei, Y., Thornton, P.E., Kao, S., Wilson, B.E., 2020. Daymet: Daily surface weather data on a 1-km grid for North America. Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/doi.org/10.3334/ORNLDAAC/1840.

William, W., Dahl, B., Hertsgaard, D., 2020. Soybean quality differentials, blending, testing and spatial arbitrage. J. Commod. Mark. 18, 100095 https://doi.org/10.1016/j.jcomm.2019.100095.

Yang, C., 2020. Remote sensing and precision agriculture technologies for crop disease detection and management with a practical application example. Engineering 6, 528–532. https://doi.org/10.1016/j.eng.2019.10.015.

Zhuo, W., Fang, S., Gao, X., Wang, L., Wu, D., Fu, S., Wu, Q., Huang, J., 2022. Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: a case study for winter wheat in Hebei, China during 2009–2013. Int. J. Appl. Earth Obs. Geoinf. 106, 102668 https://doi.org/10.1016/j.jag.2021.102668.