

Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model

Yang Lu^{a,*}, Tendai P. Chibarabada^b, Matteo G. Ziliani^c, Jean-Marie Kileshye Onema^b, Matthew F. McCabe^c, Justin Sheffield^a

^a Geography and Environment, University of Southampton, Southampton, United Kingdom

^b Waternet, P.O. Box MP600, Mount Pleasant, Harare, Zimbabwe

^c Division of Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

ARTICLE INFO

Handling Editor - Dr Z Xiyang

Keywords:

AquaCrop
Calibration
Data assimilation
Canopy cover
Soil moisture
Yield prediction

ABSTRACT

Parameter calibration is normally required prior to crop model simulation, which can be a time-consuming and data-intensive task. Meanwhile, the growth stages of different hybrids/cultivars of the same crop often show some similarities, which implies that phenological parameters calibrated for one hybrid/cultivar may be useful for the simulation of another. In this study, a data assimilation framework is proposed to reduce the requirement for parameter calibration for maize simulation using AquaCrop. The phenological parameters were uniformly scaled from previous research performed in a different location for a different maize hybrid, and other parameters were taken from default settings in the model documentation. To constrain simulation uncertainties, soil moisture and canopy cover observations were assimilated both separately and jointly in order to update model states. The methodology was tested across a rain-fed field in Nebraska for 6 growing seasons. The results suggested that the under-calibrated model with uniformly scaled phenological parameters captured the temporal dynamics of crop growth, but may lead to large estimation bias. Data assimilation effectively improved model performance, and the joint assimilation outperformed single-variable assimilation. When soil moisture and canopy cover were jointly assimilated, the overall yield estimates (RMSE = 1.24 t/ha, nRMSE = 11.48%, R^2 = 0.695) were improved over the no-assimilation case (RMSE = 2.01 t/ha, nRMSE = 18.61%, R^2 = 0.338). Sensitivity analyses suggested that the improvement was still evident with temporally sparse soil moisture observations and a small ensemble size. Further testing using observations within 90 days after planting demonstrated that the method was able to predict yield around 3 months before harvest (RMSE = 1.7 t/ha, nRMSE = 15.74%). This study indicated that maize yield can be estimated and predicted accurately by monitoring the soil moisture and canopy status, which has potential for regional applications using remote sensing data.

1. Introduction

Crop modeling is a powerful tool in simulating crop growth, yield and the impacts of climate, pests and management on crop production (Boote et al., 2013; Araya et al., 2015; Jones et al., 2017). Such models can be mechanistic, empirical or a combination of both (Estes et al., 2013a,b). Crop growth is driven by water, solar radiation, carbon dioxide concentration and nutrients, all of which can be represented in different ways among crop models. As such, models can differ significantly in their complexity, structure and calculation processes (Bouman et al., 1996). Depending on the major driving factors, crop models are mainly categorized into carbon-driven, solar-driven and water-driven

models (Steduto, 2003).

Similar to other numerical simulation schemes, crop models are susceptible to uncertainties in forcing data, model structure and parameter specification (Dorigo et al., 2007; Liu and Gupta, 2007; Marin et al., 2017). To improve crop simulation performance, many studies have adopted data assimilation techniques to improve crop modeling performance. Data assimilation techniques combine complementary information from measurements and models into an optimal estimate of the state variable of interest (Reichle, 2008). The model is used to generate continuous and consistent simulation of state variables, and the independent observations are used to limit the drift from the 'truth' by considering the respective errors of model simulation and

* Corresponding author.

E-mail address: yang.lu@soton.ac.uk (Y. Lu).

<https://doi.org/10.1016/j.agwat.2021.106884>

Received 4 December 2020; Received in revised form 28 February 2021; Accepted 22 March 2021
0378-3774/© 2021 Elsevier B.V. All rights reserved.

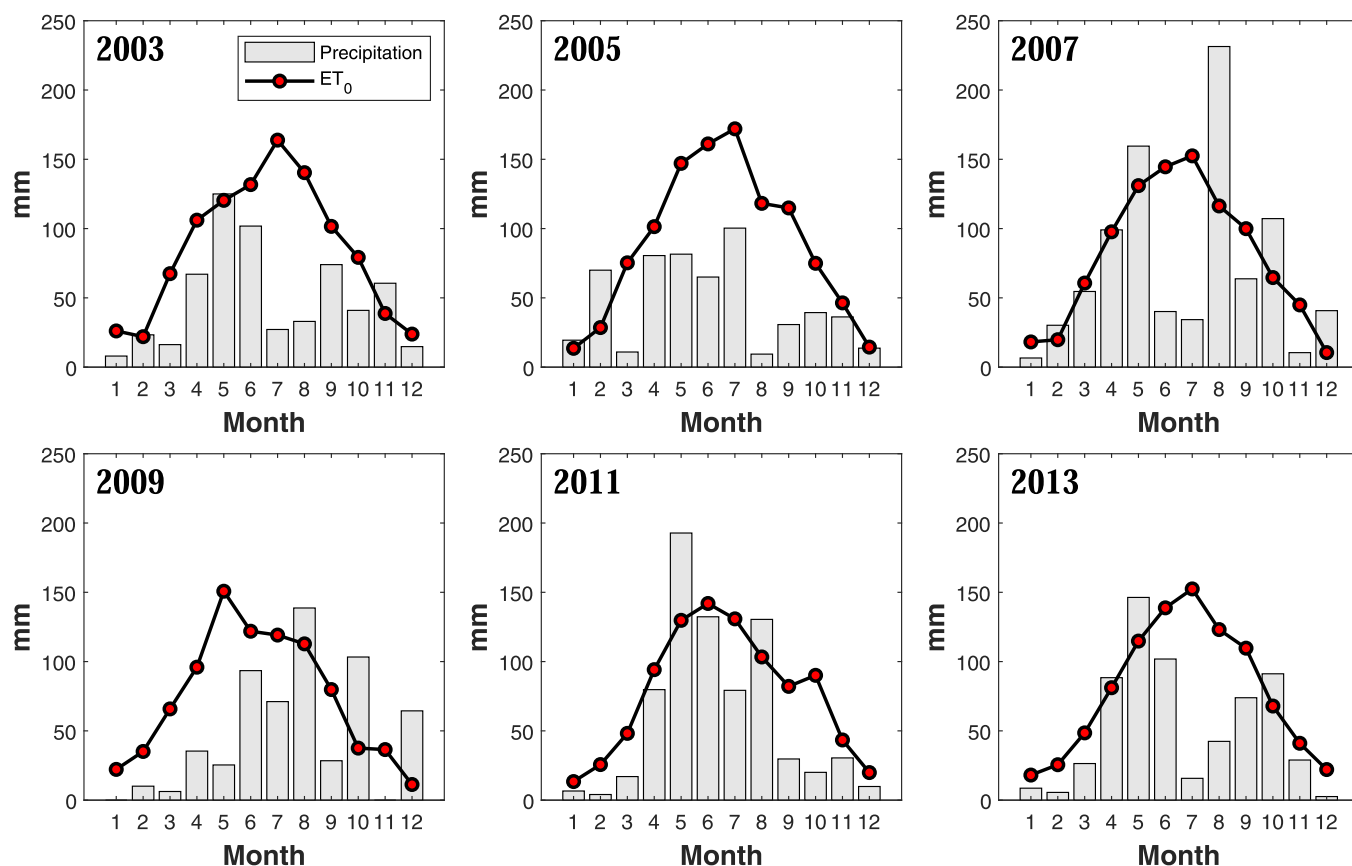


Fig. 1. Mean monthly precipitation and ET_0 in the 6 maize-growing years.

measurements. The most frequently used variables in crop model assimilation include those that are relevant to crop canopy development and those that affect crop growth. For example, many studies have assimilated leaf area index (LAI) data (Wang et al., 2013; Chen et al., 2018; Kang and Özdoğan, 2019), various vegetation indices (Fang et al., 2008; Xie et al., 2017), or the fractional canopy cover (Linker and Ioslovich, 2017) to improve canopy development simulation. Other studies have sought to improve soil water simulation by assimilating soil moisture observations (Ines et al., 2013; Chakrabarti et al., 2014; Zhang et al., 2020) or soil water indices (De Wit and Van Diepen, 2007), or to improve biomass simulation by directly assimilating biomass observations (Linker and Ioslovich, 2017; Jin et al., 2020). The coupling of crop models with hydrological or radiative transfer models (Pauwels et al., 2007; Ma et al., 2013; Li et al., 2014; Hu et al., 2019), and/or the joint assimilation of multiple types of observations (Pauwels et al., 2007; Nearing et al., 2012; Ines et al., 2013; Zhang et al., 2020) has also been explored. Most studies reported an improved utility of yield estimation after assimilation, and dual-variable assimilation often outperformed single-variable assimilation as a result of the complementary information from different types of observations (Pauwels et al., 2007; Zhang et al., 2020).

Crop models need to be calibrated before they can be used for crop growth simulation (Huang et al., 2019). Model calibration is the process of adjusting model parameters to obtain accurate model representation (Hunt and Boote, 1998; Gupta et al., 2006; Boote et al., 2013). As crop performance is dependent on the climate, specific variety and field management, calibration is necessary when any of these change significantly (Wallach, 2011; Seidel et al., 2018; Boote, 2019). Calibration is typically performed in multiple stages to fit different observation data, such as phenology and crop yield (Seidel et al., 2018), and is deemed successful when the model is tested against independent data sets and minimal residuals are observed (Hunt and Boote, 1998; Gupta et al.,

2006).

Despite the importance of model calibration, it is often a data-demanding and time-consuming task, and requires expert knowledge and sophisticated data collection tools. For example, the median time devoted to crop model calibration is 25 days according to an online survey (Seidel et al., 2018). In addition, the parameter calibration is significantly constrained by data availability and data quality, and the calibrated parameters may not be unequivocal as a result of calibration method or even the calibration sequence (Palosuo et al., 2011; Seidel et al., 2018), which are among the major difficulties in the calibration of crop and other models (e.g., hydrological models) (Boote et al., 1996; Seidel et al., 2018). This can become more problematic when the model is applied using remote sensing data over a large area with notable spatial variability in the crop phenological parameters. Meanwhile, different hybrids/cultivars of the same crop often show some similarities in the relative duration of crop development stages, despite discrepancies in the length of the growing seasons. This implies that parameter values calibrated for one crop hybrid/cultivar could potentially be useful for the simulation of another hybrid/cultivar with some adjustment, which may reduce the burden of parameter calibration. However, this will inevitably introduce simulation uncertainties. As data assimilation has the potential to accommodate some uncertainties in model parameters by exerting constraints on simulated model states (e.g., canopy cover), the simulation uncertainties may be reduced if observations are assimilated to update key model states. This may be particularly meaningful for a remote sensing application, where pixel-level calibration is difficult or infeasible.

In this study, a data assimilation framework is proposed to improve maize simulation using the AquaCrop model without routine parameter calibration. Instead, the parameter values are either taken from default settings in the model documentation, or determined by uniformly scaling the phenological parameters calibrated for a different maize

Table 1

Meteorological and cultural conditions of each growing season. P and ET_0 are the sums between planting date and harvest date. GDD is the cumulative growing degree days between planting date and harvest date. Yield data are hand-harvested grain yield adjusted to 15.5% moisture content.

Year	P (mm)	ET ₀ (mm)	Cultivar	Planting date	Harvest date	GDD	Plants/ha	Yield (t/ha)
2003	307	668	Pioneer 33B51 BT	13/05	16/10	1885	57,559	7.63
2005	304	779	Pioneer 33G66/33G68 BT	27/04	18/10	2216	53,678	10.5
2007	636	705	Pioneer 33H26 HX	02/05	01/11	2276	55,758	12.45
2009	477	670	Pioneer 33T57	22/04	11/11	1828	60,530	12.81
2011	584	644	DeKalb 61-69VT3	02/05	18/10	2071	50,184	10.11
2013	427	663	DeKalb 62-98	13/05	21/10	2067	58,190	11.3

hybrid/cultivar in a different location. To constrain the simulation uncertainties, surface soil moisture and/or canopy cover observations are assimilated into the model, and the results are evaluated for 6 growing seasons, each with a different maize hybrid/cultivar. Sensitivity analyses are performed to assess the impacts of ensemble size and assimilation interval for potential applications with remote sensing data. The ability of the method to predict maize yield using data from early growth stages is also evaluated. Overall, the objectives of this study are to assess the utility of data assimilation to reduce the burden for model calibration and to improve maize simulation and prediction with uncertainties in the parameters. The study serves as a field-scale experiment of the proposed method for a potential remote sensing application at regional scale.

2. Materials and methods

2.1. AquaCrop model

The AquaCrop model is a simple, robust and user-oriented model developed by the Food and Agriculture Organization (FAO, [Steduto et al. \(2009\)](#)). In AquaCrop, the growth of different crops is simulated using a generic module that characterizes the water stress influence on crop transpiration, biomass accumulation and yield formation ([Raes et al., 2009](#)). The crop transpiration is first calculated from the reference evapotranspiration (ET_0), which is then translated into dry above-ground biomass (B) using a crop-specific water productivity, normalized for evaporative demand and CO_2 concentration ([Steduto et al., 2009](#)):

$$B = WP^* \cdot \sum \frac{Tr}{ET_0} \quad (1)$$

where Tr is crop transpiration, and WP^* is the normalized water productivity. The yield (Y) is derived as the product of crop biomass and the harvest index (HI):

$$Y = B \cdot HI \quad (2)$$

Different from most crop models that simulate the foliage development using the LAI such as APSIM ([Keating et al., 2003](#)) and STICS ([Brisson et al., 1998](#)), AquaCrop simulates the canopy cover instead, which introduces significant simplification in the model simulation ([Steduto et al., 2009](#)). The model simulates crop growth at daily time steps, and both thermal time (growing degree day (GDD)) and calendar time can be used. The forcing data required are precipitation, minimum and maximum temperature and ET_0 . In this study an open-source version based on the AquaCrop v6.0 was adopted ([Foster et al., 2017](#)).

2.2. Study site

The study site CSP3 (41.18° N, 96.44° W, alt. 363 m) is located at the University of Nebraska-Lincoln (UNL) Eastern Nebraska Research and Extension Center (ENREC) near Mead, United States. The site also col-locates with an AmeriFlux eddy covariance station (NE3) which has been operating since 2001 ([Baldocchi et al., 2001](#)). The region has a semi-arid continental climate with an annual average precipitation of 678 mm (averaged between 1982 and 2017), and the soil is mostly silt

loam and silty clay loam ([Foolad et al., 2017](#)). The field covers an area of 65.4 ha ([Verma et al., 2005](#)), and maize and soybean are rotationally grown under rain-fed conditions. The crop is typically planted in late April through early May and harvested in October. In this study, 6 maize-growing seasons (every other year from 2003 to 2013) were simulated. The monthly precipitation and ET_0 of each year are shown in [Fig. 1](#). For each maize-growing season, a different maize hybrid/cultivar was planted and the cultural information for each growing season is listed in [Table 1](#). Judging by the relative magnitude of growing season precipitation and ET_0 , 2003 and 2005 are relatively drier, and 2007 and 2011 are relatively wetter.

2.3. Data

The data used in this study include forcing data for AquaCrop, and the observations used for data assimilation and validation. The forcing data, including precipitation and daily minimum and maximum temperature were acquired from in situ observations, and the ET_0 data were calculated from weather data using the FAO Penman–Monteith (FAO-PM) method ([Allen et al., 1998](#)).

Data used for assimilation and validation include ET , soil moisture, crop canopy cover, dry biomass and grain yield. Latent heat flux (LE) was routinely measured on an hourly basis at the AmeriFlux site. LE was averaged to daily value and then corrected for energy balance closure ([Twine et al., 2000](#)) before converted to ET using the latent heat of vaporization. Soil moisture was monitored using Theta probes (Delta-T devices, Cambridge, UK) at 5 depths (10, 25, 50, 100, 175 cm) from 4 profiles in the field (soil moisture at 175 cm was not available for 2003) at the AmeriFlux site ([Suyker and Verma, 2008](#)). The hourly measurements were converted to daily values using simple averaging and averaged among the 4 profiles to generate the mean soil moisture at each depth.

During each growing season, crop samples were routinely collected at 6 Intensive Measurement Zones (IMZs) in the field every 10–14 days ([Nguy-Robertson et al., 2015](#)). The IMZs were established for detailed process-level studies such as crop growth and biomass partitioning ([Verma et al., 2005](#); [Hunt et al., 2014](#)), and each IMZ covered an area of 20 m × 20 m. Samples from 4 to 8 plants were collected from a 1 m sampling length within each IMZ. The samples were transported on ice to the laboratory and divided into green/dead leaves, stems and reproductive organs ([Nguy-Robertson et al., 2015](#)). The leaf area was then measured in the laboratory using an area meter (Model LI-3100, LI-Cor Lincoln, NE). The plant parts were then dried at 105 °C for dry biomass measurements ([Ajaere, 2012](#)). At the end of the growing season, maize was hand-harvested in each IMZ and the grain yield was adjusted to the 15.5% grain moisture content.

The LAI measurements were converted to canopy cover (CC) measurements using the empirical equation ([Heng et al., 2009](#); [Hsiao et al., 2009](#))

$$CC = 1.005 \times [1 - e^{(-0.6 \times LAI)}]^{1.2} \quad (3)$$

2.4. Parameters

As a different maize hybrid/cultivar was planted in each growing

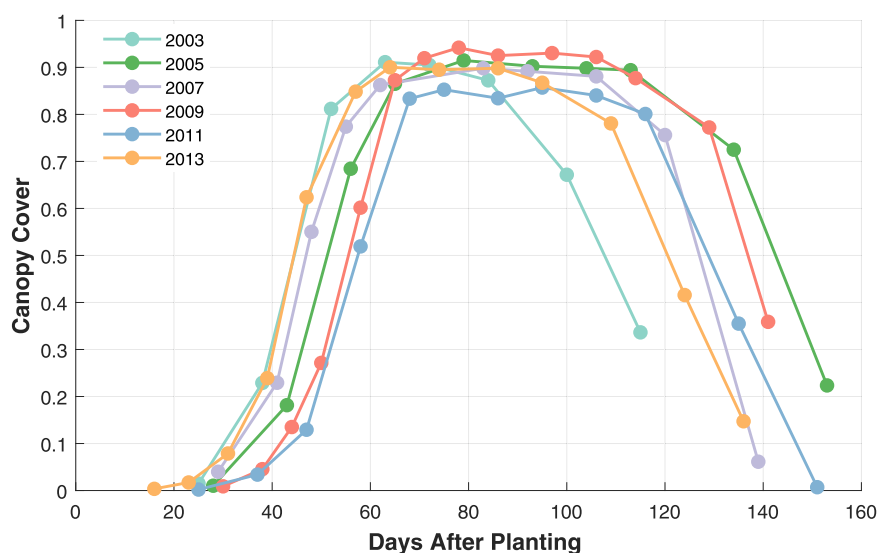


Fig. 2. Field-average canopy cover time series in the 6 growing seasons.

season, the duration of each growth stage can vary dramatically, despite the similarities in the growing cycle. The vast difference is demonstrated in the field-average canopy cover (converted from measured LAI) time series in the 6 growing seasons in Fig. 2. For example, maize was planted on the same date in 2003 and 2013 (Table 1), and the maximum canopy cover was reached within comparable time, but the canopy senescence started much sooner after reaching maximum canopy cover in 2003, leading to a significantly shorter growing season. In addition, the maximum canopy cover attainable also varies dramatically among different seasons. Consequently, the phenological parameters are not directly transferable between growing seasons, and need to be calibrated for each hybrid/cultivar separately.

The major parameters used in the simulation were classified into three categories: (1) conservative parameters, (2) variant parameters and (3) phenological parameters. Conservative parameters are those with a fixed value generally applicable to the same crop according to the AquaCrop documentation (Raes et al., 2018), and the default values recommended in the documentation were used. Variant parameters include canopy growth coefficient, maximum canopy cover and maximum rooting depth, which could vary with maize hybrid/cultivar within a small range. The parameter ranges were relatively easy to determine based on model documentation, in situ observations and/or previous studies. The most important subset of parameters are the phenological parameters, which control the development stages of crop growth. Phenological parameters are hybrid/cultivar-specific, and are usually calibrated within physically reasonable ranges to reflect in situ observations. Here the phenological parameters calibrated by Sandhu and Irmak (2019b) for a different maize hybrid in a different field were used as the reference. The parameters were calibrated using in situ canopy cover observations, biomass and yield (Sandhu and Irmak, 2019a) for the maize hybrid Mycogen 2V732 in 2009 (Sandhu and Irmak, 2019b) in a field about 160 km southwest of our study site. To account for the phenological differences between maize hybrids/cultivars, a scaling factor is determined each year which uniformly scales the reference phenological parameters, so that the simulated canopy cover time series roughly fits the observations. For example, when a scaling factor of 0.9 is used, all the phenological parameters are assumed to be 90% of the corresponding reference values reported by Sandhu and Irmak (2019b). To determine the phenological scaling factor, a number of scaling factors were tested, and the one that led to the best agreement between simulated and observed canopy cover time series was selected. The scaling factor was then allowed to vary within an arbitrary ± 0.05 range in each year to account for the parameter

Table 2

Major parameter values and scaling factors used in the maize growth simulation.

Conservative parameters	Value	Reference
base temperature (°C)	8	Raes et al. (2018)
cut-off temperature (°C)	30	Raes et al. (2018)
canopy cover at 90% emergence (cm ² /plant)	6.5	Raes et al. (2018)
canopy decline coefficient (%/GDD)	0.01	Raes et al. (2018)
crop basal coefficient prior to senescence (Kcb)	1.05	Raes et al. (2018)
normalized water productivity (g/m ²)	33.7	Raes et al. (2018)
canopy expansion stress threshold (upper)	0.14	Raes et al. (2018)
canopy expansion stress threshold (lower)	0.72	Raes et al. (2018)
shape factor for canopy expansion	2.9	Raes et al. (2018)
stomatal conductance stress threshold (upper)	0.69	Raes et al. (2018)
shape factor for stomatal control	6	Raes et al. (2018)
canopy senescence stress threshold (upper)	0.69	Raes et al. (2018)
shape factor for canopy senescence	2.7	Raes et al. (2018)
Variant parameters	Value	Reference
canopy growth coefficient (%/GDD)	0.012–0.013	Raes et al. (2018)
maximum fractional canopy cover	0.85–0.98	observations
maximum effective rooting depth (m)	1–2	Sandhu and Irmak (2019b)
Reference phenological parameters	Value	Reference
time from sowing to emergence (GDD)	70	Sandhu and Irmak (2019b)
time from sowing to maximum root depth (GDD)	1592	Sandhu and Irmak (2019b)
time from sowing to senescence (GDD)	1603	Sandhu and Irmak (2019b)
time from sowing to maturity (GDD)	1750	Sandhu and Irmak (2019b)
time from sowing to flowering (GDD)	885	Sandhu and Irmak (2019b)
duration of flowering (GDD)	190	Sandhu and Irmak (2019b)
Phenological scaling factors	Minimum	Maximum
2003	0.75	0.85
2005	1.05	1.15
2007	1	1.1
2009	0.8	0.9
2011	0.95	1.05
2013	0.95	1.05

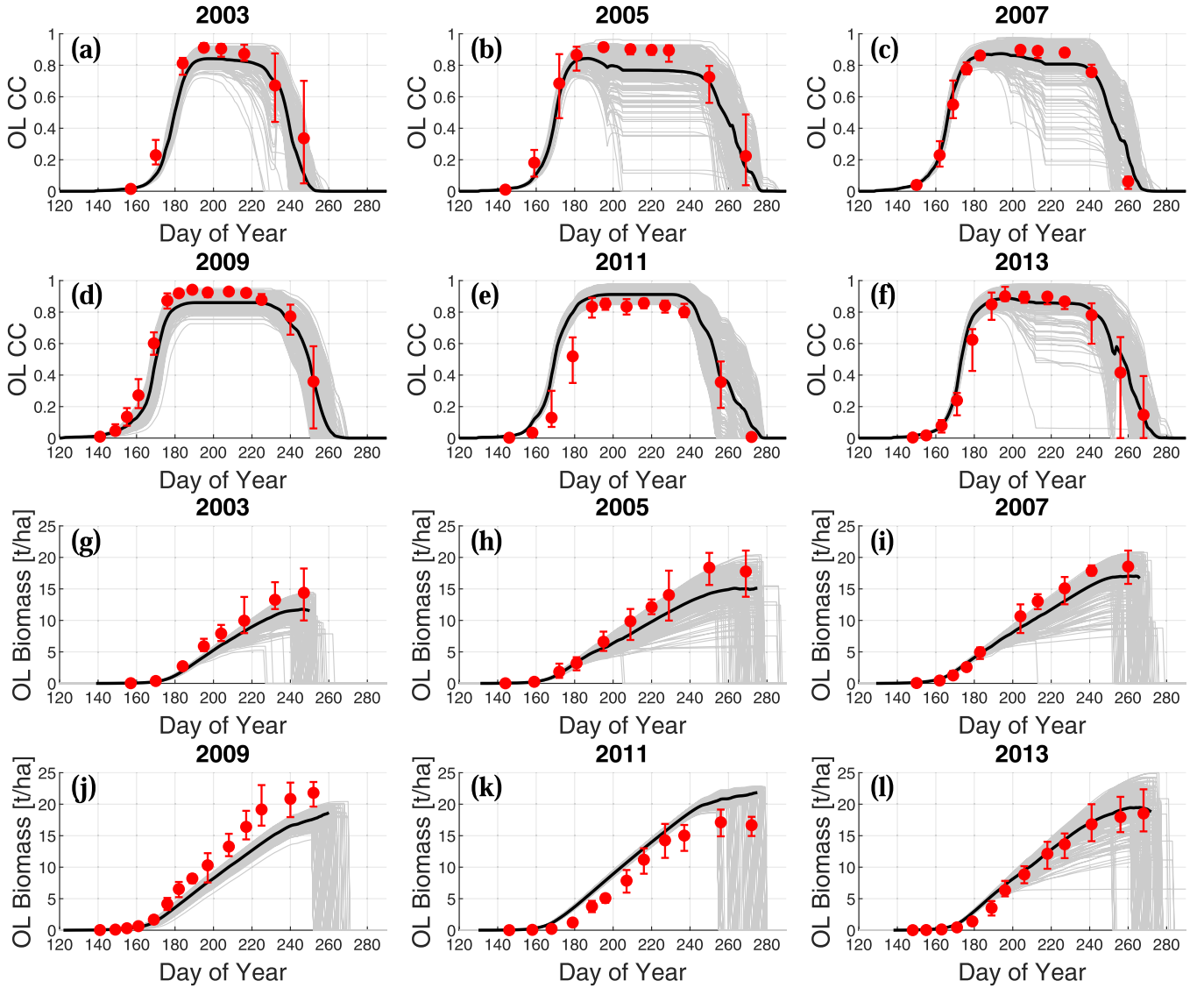


Fig. 3. Canopy cover and biomass estimates from open-loop (OL) simulations in the 6 growing seasons. The gray lines represent the 200 OL ensemble realizations, and the thick black lines represent the mean OL estimates. The decrease of biomass simulation to zero indicates the end of growing season for each ensemble realization. The red circles are the field average measurements, and the error bars indicate the variation range of the 6 IMZ samplings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

uncertainties. The parameter values and scaling factors are summarized in Table 2.

2.5. Data assimilation framework

The ensemble Kalman filter (EnKF) is adopted in this study. The EnKF is a Monte-Carlo variant of the Kalman filter (Kalman, 1960) which sequentially synthesizes complementary information from the model simulation (the ‘forecast’) and available observations to generate an optimal estimate (the ‘analysis’) of the model states (Evensen, 2003). The original Kalman filter works well for linear models, while the EnKF is more feasible for non-linear and high-dimensional models. In the EnKF, the forward model is run for a number of realizations with perturbed forcing data and/or model parameters to generate an ensemble of model forecasts. When an observation is available, each realization is updated by analyzing the ensemble covariance and the observation error distribution. The ensemble mean of all the analyses is taken as the updated estimate of EnKF.

The evolution of crop model state variables in time can be described by

$$\mathbf{x}_{i,t}^f = M(\mathbf{x}_{i,t-1}^f, \mathbf{u}_{i,t}, \mathbf{b}_{i,t}) + \mathbf{w}_t \quad (4)$$

where M is the forward model (AquaCrop in this case), $\mathbf{x}_{i,t}^f$ is the state vector of the model forecast (canopy cover or soil moisture) of the i th realization at time step t , $\mathbf{u}_{i,t}$ is the forcing data, $\mathbf{b}_{i,t}$ is the model parameter vector, and \mathbf{w}_t represents model error.

The observation is related to the true state by

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \quad (5)$$

where \mathbf{y}_t is the observation, \mathbf{H} is the observation operator that translates the state variables to the observation, \mathbf{x}_t is the true state, and \mathbf{v}_t is the observation error.

When an observation is available, the EnKF updates the state variables in each model realization by

$$\mathbf{x}_{i,t}^a = \mathbf{x}_{i,t}^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{i,t}^f) \quad (6)$$

where $\mathbf{x}_{i,t}^a$ is the model analysis after assimilation. \mathbf{K}_t is the Kalman gain, which is derived by

$$\mathbf{K}_t = \mathbf{P}_t^f \mathbf{H}^T (\mathbf{H} \mathbf{P}_t^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (7)$$

Here \mathbf{P}_t^f and \mathbf{R} are the covariance of the ensemble model forecast and the observations, respectively. When only one observation is available, \mathbf{R} equals the observation variance. In the EnKF, the model and observation errors are assumed to follow Gaussian distribution with zero mean.

The determination of observation error is crucial for the EnKF performance. To determine the soil moisture observation error, the daily standard deviation of the observations from the 4 soil moisture profiles was first calculated, and the multi-year average value ($0.04 \text{ m}^3/\text{m}^3$) was taken as the observation error. The canopy cover observation error varies dramatically during the growing season. In the exponential growth stage and the decay period canopy cover demonstrated significant variability among the IMZs (standard deviation as large as 0.297), while the variability was only marginal when the canopy was near maximum (shown in Fig. 3). Hence the canopy cover observation error was assumed dynamic, and the standard deviation of the samplings from the 6 IMZs on each sampling day was used separately.

2.6. Experiment set-up

In this study, 6 growing seasons were simulated (2003, 2005, 2007, 2009, 2011, 2013) for different maize hybrids/cultivars (Table 1). A soil column of 2.15 m was divided into 5 layers (0–15, 15–35, 35–65, 65–135, 135–215 cm) to fit the soil moisture sensor depth. For each year, the soil moisture profile was initialized on January 1 using multi-year average soil moisture profile. Soil texture variation range (sand (5–15%), clay (20–30%)) was first derived based on Web Soil Survey Data (<https://websoilsurvey.sc.egov.usda.gov>) as presented in Foolad et al. (2017). The soil texture of each ensemble member was then randomly sampled within the range to account for texture spatial heterogeneity, and soil hydraulic properties (e.g., field capacity, hydraulic conductivity at saturation) were estimated using a pedotransfer model (Saxton and Rawls, 2006) from soil texture data to keep the properties consistent. An ensemble size of 200 was used in the analysis, which proved sufficient following a sensitivity test (impact of ensemble size is evaluated in Section 3.6 in this paper). For each model realization, the variant parameters and the uniform phenological scaling factors were randomly sampled within the given ranges. Model estimates from four strategies were compared: (1) open-loop (OL: model simulations using randomly sampled parameters without performing assimilation); (2) soil moisture assimilation (DA_{SM} : model simulations with only soil moisture assimilation); (3) canopy cover assimilation (DA_{CC} : model simulations with only canopy cover assimilation); (4) soil moisture and canopy cover assimilation (DA_{SM+CC} : model simulations with both soil moisture and canopy cover assimilation). Here only the surface layer soil moisture (10 cm) observations were assimilated to update the soil moisture profile to mimic a remote sensing application.

2.7. Result evaluation

The model estimates of canopy cover, ET , biomass and yield from the four strategies were evaluated against in situ observations in the 6 growing seasons. The canopy cover, biomass and yield estimates were compared against field data, and the ET estimates were evaluated using AmeriFlux data. Five statistical metrics were used to evaluate the goodness of fit for each simulation, including the coefficient of determination (R^2), the root mean square error (RMSE), the normalized RMSE (nRMSE), the bias and the Willmott's index of agreement (d) (Willmott, 1981).

The metrics are calculated by

$$R^2 = \left[\frac{1}{n} * \frac{\sum_{i=1}^n (S_i - \bar{S})(O_i - \bar{O})^2}{\sigma_S * \sigma_O} \right]^2 \quad (8)$$

Table 3

Goodness of fit metrics for canopy cover estimates in the 6 growing seasons under different simulation strategies.

Year	Strategy	R^2	RMSE	nRMSE (%)	Bias	d
2003	OL	0.975	0.08	13.64	−0.06	0.985
	DA_{SM}	0.959	0.07	12.29	−0.02	0.988
	DA_{CC}	0.976	0.07	11.62	−0.03	0.990
	DA_{SM+CC}	0.956	0.08	14.14	−0.02	0.985
2005	OL	0.979	0.09	13.81	−0.06	0.981
	DA_{SM}	0.979	0.05	7.78	0.00	0.995
	DA_{CC}	0.993	0.04	6.03	−0.02	0.997
	DA_{SM+CC}	0.989	0.04	6.19	−0.01	0.997
2007	OL	0.944	0.09	15.31	0.02	0.978
	DA_{SM}	0.941	0.11	17.83	0.07	0.972
	DA_{CC}	0.997	0.02	3.53	0.01	0.999
	DA_{SM+CC}	0.994	0.03	5.38	0.02	0.998
2009	OL	0.978	0.08	13.05	−0.06	0.987
	DA_{SM}	0.991	0.05	7.50	−0.03	0.996
	DA_{CC}	0.992	0.04	6.04	−0.01	0.997
	DA_{SM+CC}	0.996	0.03	4.24	−0.01	0.999
2011	OL	0.946	0.14	27.50	0.11	0.961
	DA_{SM}	0.941	0.14	28.09	0.11	0.959
	DA_{CC}	0.998	0.02	4.35	0.02	0.999
	DA_{SM+CC}	0.998	0.02	4.75	0.02	0.999
2013	OL	0.969	0.08	14.52	0.04	0.989
	DA_{SM}	0.961	0.08	15.49	0.03	0.987
	DA_{CC}	0.986	0.04	8.52	0.00	0.996
	DA_{SM+CC}	0.989	0.04	7.74	−0.01	0.997

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_i - O_i)^2}{n}} \quad (9)$$

$$nRMSE = \frac{RMSE}{\bar{O}} * 100\% \quad (10)$$

$$bias = \frac{1}{n} \sum_{i=1}^n (S_i - O_i) \quad (11)$$

$$d = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (|S_i - \bar{S}| + |O_i - \bar{O}|)^2} \quad (12)$$

where O_i stands for the i th observation, S_i is the corresponding model simulation, n is the number of observations available, σ_S and σ_O are the standard deviations of model simulations and observations, and \bar{S} and \bar{O} are the mean of model simulations and observations, respectively. The selected metrics were used to assess the goodness of fit from different aspects, in which R^2 depicts the proportion of the observation variance that is predictable from model simulation, the RMSE and nRMSE illustrate the absolute and relative magnitude of the difference between simulation and observation, the bias refers to the tendency of the simulation to over- or under-estimate the model state, and d represents the ratio of the mean square error and the potential error (1 indicates a perfect match, and 0 indicates no agreement).

3. Results

3.1. Open-loop simulations

It is crucial for an assimilation application that the OL simulation captures the dynamics of key model states and that the ensemble spread reflects the model state variability. Here the OL estimates for canopy cover and biomass are examined in Fig. 3. Overall, the model simulation with uniformly scaled phenological parameters and the ± 0.05 uncertainties successfully characterized the temporal dynamics of canopy cover development (Fig. 3a–f) in each growing season, and the spatial variation range of canopy cover in the field generally falls within the ensemble spread. This suggests that the assumption of similarities in the relative durations of growth stages of different maize hybrids/cultivars

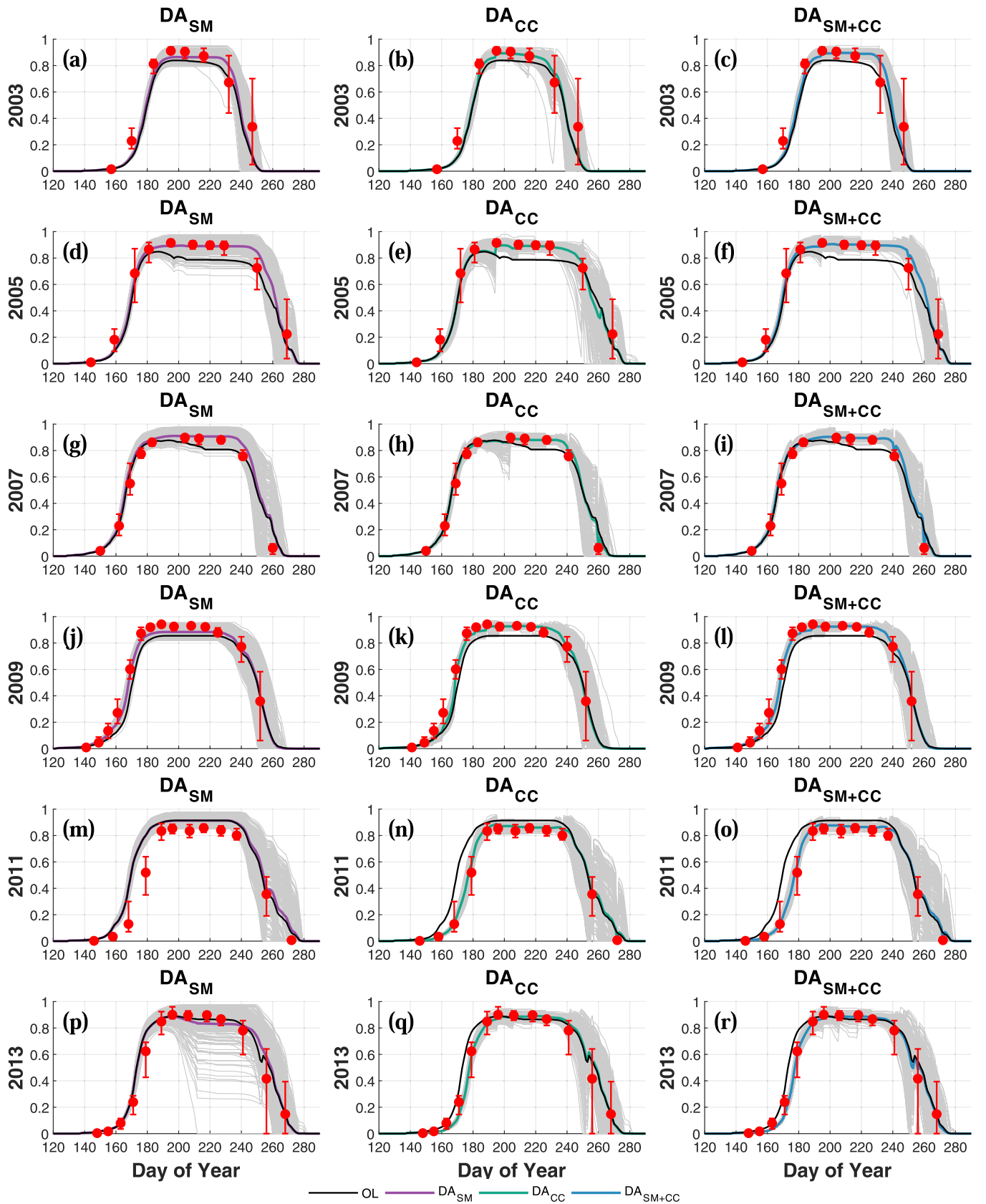


Fig. 4. Canopy cover estimates from the three data assimilation strategies. The gray lines represent the 200 assimilation ensemble realizations, and the thick line represent the mean estimates from different strategies. The OL mean is also plotted in black lines as a reference. The red circles are the field average measurements, and the error bars indicate the variation range of the 6 IMZ samplings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

Goodness of fit metrics for daily *ET* estimates in the 6 growing seasons under different simulation strategies.

Year	Strategy	R ²	RMSE (mm/day)	nRMSE (%)	Bias (mm/day)	d
2003	OL	0.594	1.19	29.42	-0.35	0.866
	DA _{SM}	0.503	1.28	31.80	0.26	0.833
	DA _{CC}	0.567	1.24	30.76	-0.39	0.853
	DA _{SM+CC}	0.484	1.31	32.44	0.28	0.825
2005	OL	0.490	1.37	36.54	-0.55	0.807
	DA _{SM}	0.603	1.09	29.04	0.10	0.876
	DA _{CC}	0.494	1.37	36.46	-0.55	0.811
	DA _{SM+CC}	0.603	1.09	28.99	0.05	0.877
2007	OL	0.512	1.40	34.52	-0.70	0.789
	DA _{SM}	0.468	1.31	32.47	-0.23	0.815
	DA _{CC}	0.516	1.40	34.47	-0.70	0.789
	DA _{SM+CC}	0.469	1.32	32.52	-0.26	0.813
2009	OL	0.823	1.28	33.37	-0.97	0.875
	DA _{SM}	0.663	1.32	34.33	-0.65	0.840
	DA _{CC}	0.805	1.32	34.62	-1.00	0.868
	DA _{SM+CC}	0.663	1.34	34.99	-0.70	0.837
2011	OL	0.561	1.43	39.19	-0.59	0.833
	DA _{SM}	0.539	1.44	39.52	-0.54	0.827
	DA _{CC}	0.569	1.44	39.38	-0.64	0.833
	DA _{SM+CC}	0.560	1.43	39.24	-0.60	0.831
2013	OL	0.484	1.21	32.76	-0.30	0.824
	DA _{SM}	0.512	1.14	31.00	-0.20	0.840
	DA _{CC}	0.558	1.11	29.94	-0.33	0.850
	DA _{SM+CC}	0.538	1.13	30.59	-0.28	0.846

is valid despite the differences in the total season length, which provides a good basis for updating canopy cover using in situ observations. The goodness of fit metrics of OL canopy cover estimates are included in Table 3. In most growing seasons, the canopy cover estimates agreed well with observations ($R^2 > 0.95$, $RMSE < 0.1$). In 2003 and 2009, canopy cover could still reach above 0.9 despite the low precipitation. While in 2005, with the lowest precipitation and the highest ET_0 (Table 1), the water stress was the highest among the 6 years, which triggered early senescence in some ensemble realizations that led to low canopy cover estimates. In 2011, OL over-estimated canopy cover over the entire growing season, which relates to the abundant precipitation, particularly in the exponential growth period in May and June (Fig. 1).

The statistical metrics of OL biomass estimates (Fig. 3g–l) are included in Table 5. The ensemble realizations captured the overall trends of biomass accumulation, but the estimates may deviate a lot from observations, which is closely related to canopy cover simulation. For example, biomass was under-estimated in 2003, 2005, 2007 and 2009, and over-estimated in 2011, which is in line with canopy cover simulations. In particular, 2009 was the only year with the final biomass over 20 t/ha, which was also significantly higher compared to other years.

3.2. Canopy cover estimates

The canopy cover estimates from different assimilation strategies are shown in Fig. 4, and the goodness of fit metrics are included in Table 3. Compared with OL estimates (Fig. 3), DA_{SM} improved canopy cover estimates in the two drier growing seasons (2003 and 2005), while the performance was slightly worse in the two wetter growing seasons (2007 and 2011). In dry environments, soil moisture is the major limiting factor in canopy development, and the water stress coefficient for leaf expansion is more sensitive to soil moisture variation than under wet conditions (Raes et al., 2012). When soil moisture was assimilated, the water stress was better depicted. As a result, the canopy cover dynamics were simulated more accurately, and the early senescence in some ensemble realizations in 2005 was corrected. In wet environments, soil moisture is not a strong constraint for canopy cover as the water stress is low, evidenced by the positive bias in OL estimates. Consequently, the improvement in canopy cover estimates from DA_{SM} was limited. In

DA_{CC}, canopy cover was directly updated using in situ observations, therefore the improvement was evident in all growing seasons. In particular, the OL estimates in 2011 gave the largest RMSE (0.14) among the 6 years, which was largely caused by a mismatch in the relative lengths of phenological stages. After canopy cover assimilation, the mismatch was to a large extent reduced ($RMSE = 0.02$, $d = 0.999$). DA_{SM+CC} combined the strengths of DA_{SM} and DA_{CC}, and gave the best overall performance among all years.

3.3. ET estimates

The goodness of fit metrics for daily *ET* estimates are shown in Table 4. The OL *ET* estimates generally agree with the observations, and the performance is better in the drier growing seasons (2003 and 2005) than in the wetter growing seasons (2007 and 2011). This may relate to the new feature in water stress determination in AquaCrop v6.0. The simulation in very dry conditions was improved in AquaCrop v6.0, in which the soil water stress was determined by the wetter soil between the total root zone and the top soil with a default depth of 0.1 m. This improvement essentially reduces the soil water stress on crop growth, thus the model becomes more robust to soil moisture simulation uncertainty. *ET* estimation RMSEs from different simulation strategies do not differ too much in most years, but assimilating soil moisture tends to reduce *ET* bias.

The daily ratio of Tr/ET is plotted in Fig. 5. As no *Tr* observations were available, the estimates from data assimilation were only compared to OL simulations. The Tr/ET ratio dynamics closely resemble the canopy cover development, as a denser canopy increases the crop transpiration coefficient (K_{cTr}) and decreases the soil evaporation coefficient (Ke), which leads to a larger Tr/ET ratio, and vice versa. In 2011, the *ET* estimates were comparable from different strategies, but the Tr/ET ratio in the exponential growth stage was significantly reduced after canopy cover assimilation, which would largely reduce the *Tr* and hence the biomass production. Fig. 5 demonstrated that despite the similar *ET* estimates from different strategies, the *Tr* estimates can differ a lot through data assimilation, which would have a large impact on biomass estimation.

3.4. Biomass estimates

The biomass estimates from the three data assimilation strategies are shown in Fig. 6 and the statistical metrics are listed in Table 5. Assimilating soil moisture or canopy cover generally improved biomass estimation compared to OL results. DA_{SM} outperformed DA_{CC} in 2003, 2005, 2007 and 2009, while DA_{CC} gave better estimates in 2011 and 2013. In 2003, 2005, 2007 and 2009, biomass was under-estimated by OL as a result of the smaller *Tr* estimates caused by the under-estimated canopy cover (Fig. 3). The assimilation of soil moisture resulted in higher *ET* estimates. Consequently, the negative bias in biomass estimates was largely reduced. In 2011 and 2013, soil moisture was not a major limiting factor in crop growth, and OL over-estimated biomass as a result of the high positive canopy cover estimation bias. When canopy cover was assimilated, the canopy cover estimation bias was reduced, which led to lower *Tr* estimates. As a result, the positive estimation bias decreased significantly. Overall, DA_{SM+CC} performed best in all the 6 growing seasons. In 2009, the final biomass was significantly higher compared to other years, despite the low precipitation. This may be attributed to the unique aridity-resistant features of the specific hybrid/cultivar that were not accounted for in the parametrization, and may require more detailed information to further improve the model performance. Aside from 2009, the DA_{SM+CC} biomass estimation RMSE was basically smaller than 1 t/ha, and the nRMSE ranged between 5.48% and 13.23%. The overall evaluation of biomass estimates from the 6 years is shown in Fig. 7. DA_{SM} and DA_{CC} performed similarly over the 6 years ($RMSE \approx 1.6$ t/ha, $nRMSE \approx 20\%$). When soil moisture and canopy cover were assimilated together, the best biomass estimates were

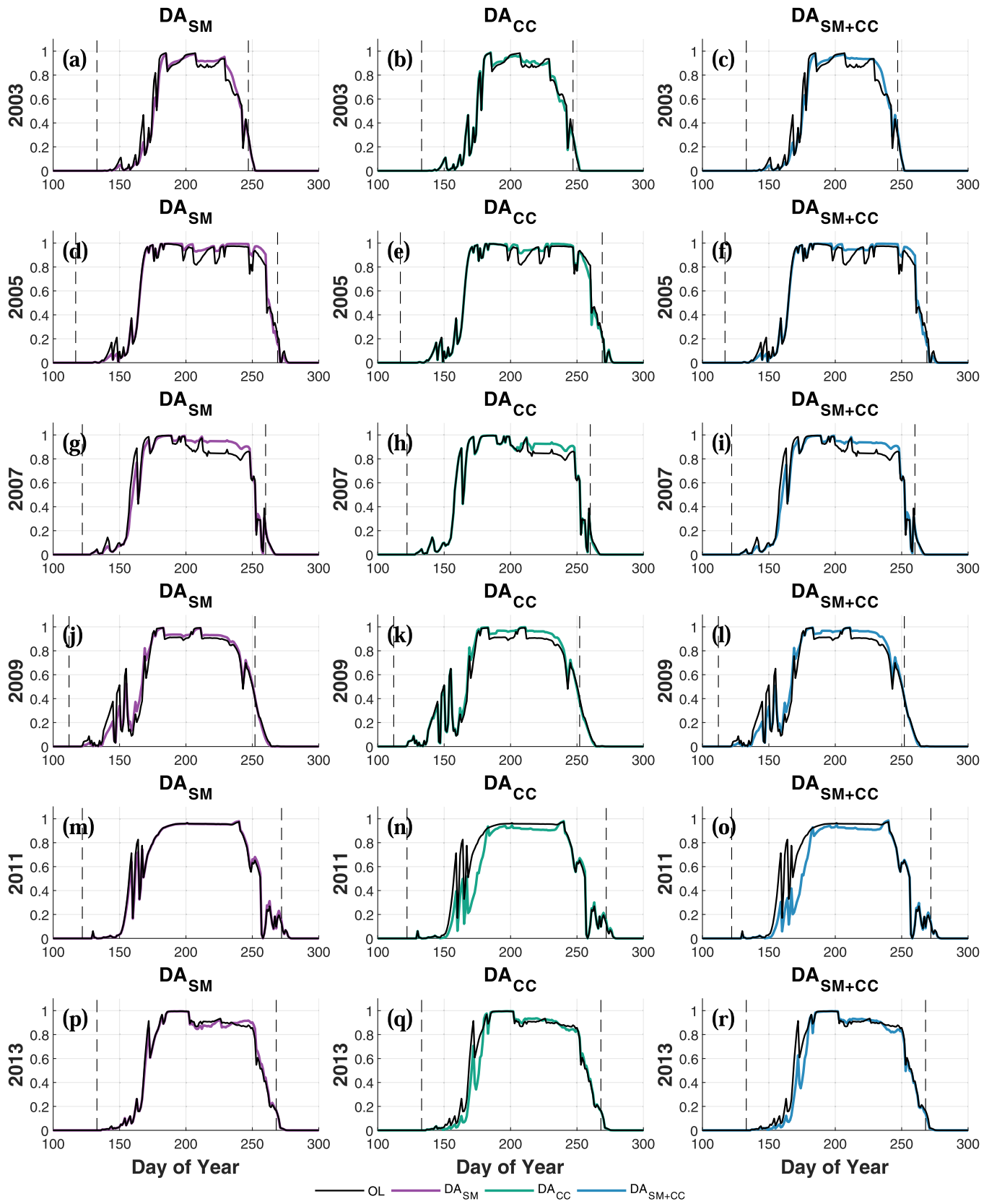


Fig. 5. Daily Tr/ET ratio in the 6 growing seasons under different simulation strategies. The OL estimates are also plotted in black lines as a reference. The dashed lines indicate the planting date and the last field sampling date.

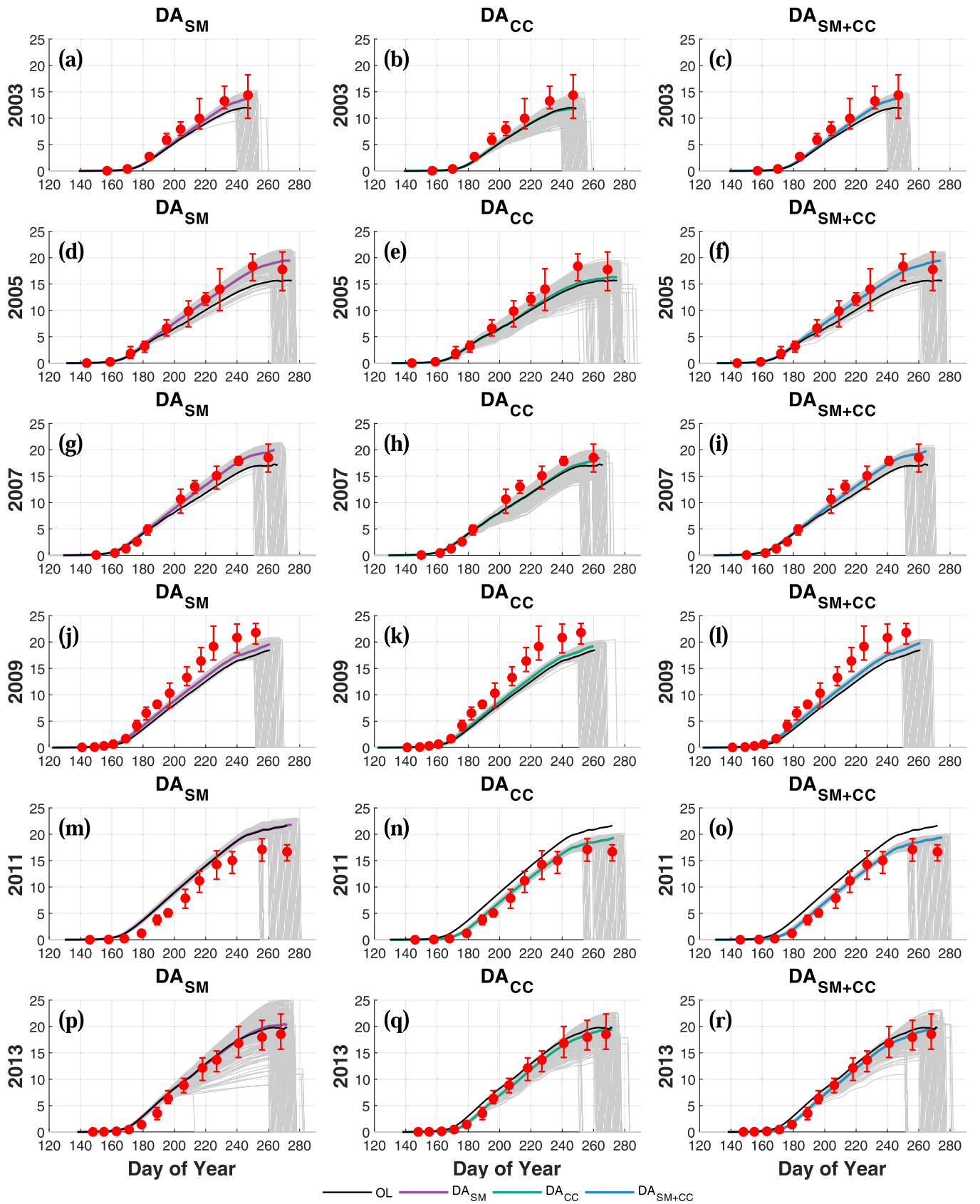


Fig. 6. Biomass estimates from the three data assimilation strategies. The gray lines represent the 200 assimilation ensemble realizations, and the thick line represent the mean estimates from different strategies. The OL mean is also plotted in black lines as a reference. The red circles are the field average measurements, and the error bars indicate the variation range of the 6 IMZ samplings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5

Goodness of fit metrics for biomass estimates in the 6 growing seasons under different simulation strategies.

Year	Strategy	R ²	RMSE (t/ha)	nRMSE (%)	Bias (t/ha)	d
2003	OL	0.997	1.69	24.72	-1.40	0.969
	DA _{SM}	0.990	0.97	14.16	-0.79	0.991
	DA _{CC}	0.997	1.60	23.46	-1.32	0.972
	DA _{SM+CC}	0.992	0.90	13.23	-0.74	0.992
2005	OL	0.995	2.06	24.55	-1.51	0.971
	DA _{SM}	0.992	0.60	7.18	-0.09	0.998
	DA _{CC}	0.994	1.55	18.38	-1.14	0.985
	DA _{SM+CC}	0.992	0.61	7.27	-0.15	0.998
2007	OL	0.993	1.48	17.50	-0.89	0.987
	DA _{SM}	0.991	0.66	7.84	0.02	0.998
	DA _{CC}	0.988	1.35	15.96	-0.78	0.990
	DA _{SM+CC}	0.992	0.67	7.96	-0.07	0.998
2009	OL	0.993	2.89	32.73	-2.23	0.959
	DA _{SM}	0.995	2.39	27.12	-1.77	0.972
	DA _{CC}	0.994	2.53	28.66	-1.95	0.969
	DA _{SM+CC}	0.995	2.27	25.70	-1.69	0.975
2011	OL	0.985	2.57	33.38	2.18	0.968
	DA _{SM}	0.985	2.59	33.54	2.20	0.967
	DA _{CC}	0.988	1.00	13.03	0.58	0.994
	DA _{SM+CC}	0.988	1.01	13.13	0.60	0.994
2013	OL	0.995	0.86	11.14	0.67	0.996
	DA _{SM}	0.993	0.95	12.35	0.73	0.996
	DA _{CC}	0.996	0.45	5.84	0.09	0.999
	DA _{SM+CC}	0.997	0.42	5.48	0.02	0.999

derived (RMSE = 1.24 t/ha, nRMSE = 15.42%), which significantly improved over OL estimates (RMSE = 2.08 t/ha, nRMSE = 25.87%) in all the goodness of fit metrics. Compared with previous AquaCrop maize simulation studies (Heng et al., 2009; Mebane et al., 2013; Sandhu and Irmak, 2019a,b; Ran et al., 2020), the joint assimilation reached similar accuracy in biomass estimation. This suggests that the proposed method effectively improved AquaCrop simulation performance, even though no routine calibration was performed.

3.5. Yield estimates

The maize yield estimates from OL and the three assimilation strategies are shown in Fig. 8. The OL significantly under-estimated yield in 2003, 2005, 2007 and 2009, and over-estimated yield in 2011, which is in line with OL biomass estimates (Fig. 3). Soil moisture assimilation significantly improved yield estimation in 2003 and 2005, while canopy cover assimilation performed better in 2011. The relative performance of DA_{SM} and DA_{CC} is consistent with the biomass estimates (Fig. 6), as yield is derived directly from biomass using a crop-specific harvest index. Overall, DA_{SM+CC} performed best, in which the RMSE decreased from 2.01 t/ha for OL to 1.24 t/ha, and the R² increased from 0.338 for OL to 0.695.

3.6. Impact of ensemble size

The data assimilation performance relies on a sufficient number of ensemble realizations to generate accurate covariance samplings (Yin et al., 2015). When a limited number of ensemble members are used, the assimilation results usually demonstrate large instability. Here an assimilation experiment was undertaken to determine the minimum number of ensemble realizations needed to generate stable results. DA_{SM+CC} was performed using 5, 10, 20, 30, 50, 100, 200, 300 and 400 ensembles, respectively. For each ensemble size, DA_{SM+CC} was performed multiple times, and the box plot of biomass estimation RMSE is shown in Fig. 9. While the figure clearly demonstrates that the result stability increases with larger ensemble size, comparatively little improvement is seen when more than 50 ensemble members are used. In some years (e.g., 2007), stable results can even be derived with as few as 10 ensemble members. This indicates that similar assimilation estimates

can be obtained using a much smaller ensemble size, which would significantly reduce the computational burden, particularly when the methodology is applied over a large area using remote sensing data.

3.7. Impact of soil moisture assimilation frequency

In this study, soil moisture observations were available at sub-daily scale and were assimilated on a daily basis. However, if the methodology is applied using remote sensing data, soil moisture is unlikely to be available every day, which may lead to reduced assimilation performance. Fig. 10 shows the biomass and yield estimation RMSEs from DA_{SM+CC} with soil moisture data assimilated every 1–14 days. The RMSEs from DA_{CC} are also plotted in dashed lines, as the DA_{SM+CC} simulations are expected to converge to DA_{CC} simulations when soil moisture data are assimilated at a very long interval. For biomass estimates, the RMSE slowly increases with larger assimilation interval in 2003, 2005, 2007 and 2009, but the added value of soil moisture assimilation is still evident even at the 14-day assimilation interval compared to DA_{CC}. In 2011, the impact of assimilation interval is marginal, while the RMSE fluctuates within a 0.2 t/ha range in 2013. This is consistent with previous research showing that a shorter assimilation interval has a direct impact on soil moisture estimates and a reduced impact on other model states (Lu et al., 2019). For yield estimates, the RMSE increases gradually while the R² decreases slightly with longer soil moisture assimilation interval. When soil moisture is assimilated at a 14-day interval, the yield RMSE increases by 0.22 t/ha and the R² decreases by 0.078 compared to daily soil moisture assimilation case. Overall, a shorter assimilation interval generally leads to improved biomass and yield estimates, but the added value of soil moisture assimilation is still evident at a 14-day assimilation interval in most cases. This suggests that the methodology can also effectively improve model performance in areas and applications with limited soil moisture data availability.

3.8. Potential for yield prediction

In this study, all available soil moisture and canopy cover observations during the growing season were used to improve maize simulation. The results demonstrate that maize yield can be estimated accurately by monitoring the soil wetness and canopy growth, which enables crop yield monitoring and retrospective yield estimation when the whole growing season is observed. Another interesting topic is whether the proposed method is capable of predicting the final maize yield using observations in the first few months of the growing season. Here the model was rerun using observations within 90 days after planting (i.e., no observations collected after this day were used for the phenological scaling or assimilation). Fig. 11 shows the yield prediction evaluation. As a smaller number of observations are used for phenological scaling and assimilation, the yield predictions deteriorate compared to using all observations in the growing season (Fig. 8). Despite the reduced performance, DA_{SM+CC} still outperforms all other experiments, and the yield prediction RMSE is reduced from 2.48 t/ha in OL to 1.7 t/ha (31.5% reduction). This suggests that the proposed method is able to predict the final yield around 3 months before harvest. If more observations are used or better knowledge on the maize phenological features is available, the yield prediction accuracy can be further improved.

4. Discussion

4.1. Comparison with other calibration/assimilation studies

Results from this study demonstrated that maize simulation can be improved by assimilating soil moisture and canopy cover data with much reduced calibration workload. This methodology is different from the calibration studies, which aim to accurately estimate key model parameters for a specific crop hybrid/cultivar. One advantage of full

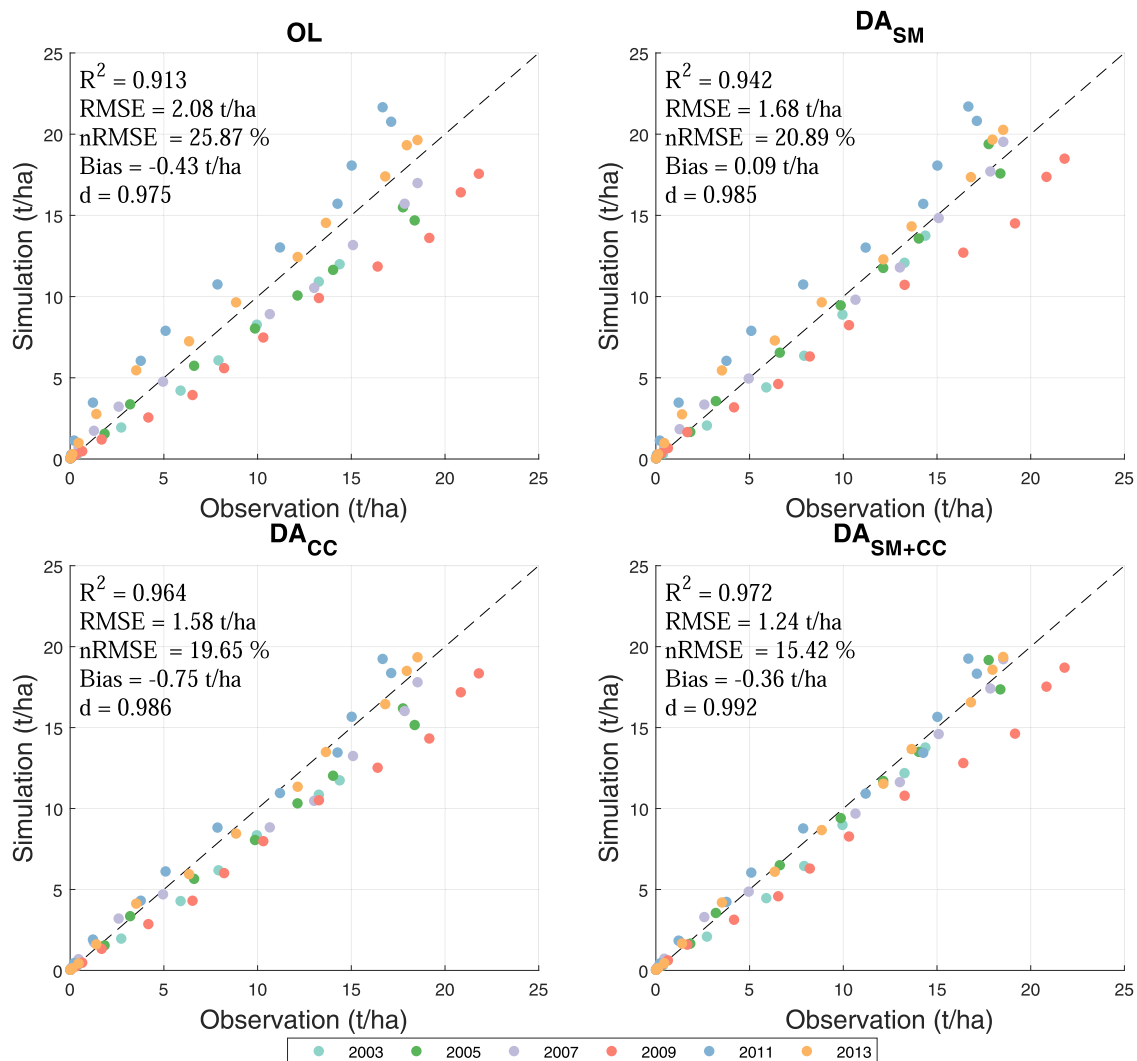


Fig. 7. Biomass estimates using different strategies from the 6 growing seasons.

calibration is the ability to provide parameter estimates that are physically meaningful, comparable between hybrids/cultivars, and instrumental for the calibration of new hybrids/cultivars. Consequently, a full calibration study may perform better at field-level, as the specific features are more carefully accounted for. The strength of our methodology lies in the ability to address the spatial heterogeneity in crop phenology over large areas. At regional scale, multiple hybrids/cultivars from different seed companies are often used, and the hybrid/cultivar parameters are often unavailable (Jégo et al., 2010). In addition, the lack of high-quality weather, soil and management data is a major constraint for full calibration (Grassini et al., 2015; Morell et al., 2016), and the spatial distribution of the use of seeds is also unknown (Jégo et al., 2010). Therefore, the modeling capacity may be reduced if uniform parameters are applied over a large area with strong variability. By contrast, our work suggests that pixel-specific parameters can be quasi-calibrated in a cost-effective way without compromising the modeling capacity through the assimilation of key variables. This facilitates the pixel-level application for a large area, particularly with large spatial variability.

To the authors' knowledge, this study is the first to explore the possibility of reduced calibration among agricultural assimilation studies. Full calibration is normally performed before assimilation at field-level, which constrains the extension to a regional application. For remote sensing applications, calibration is often performed at aggregated scale (e.g., county-level) using averaged vegetation time series

(Kang and Özdoğan, 2019). This does not fully utilize the capacity of remote sensing to provide pixel-level measurements for precision agriculture, and ignores the phenological variability within the calibration region (e.g., each county). This may also be prone to uncertainties caused by the non-linearity in vegetation index (e.g., LAI) upscaling (Garrigues et al., 2006; Chen et al., 2019). Our work provides potential for improved crop modeling that could accommodate computational efficiency as well as pixel-level calibration for precision agricultural applications.

4.2. Potential for high-resolution yield estimation

As both soil moisture and canopy cover are retrievable using satellite sensors, the methodology has considerable potential to be applied using remote sensing data over a large area. Soil moisture has been retrieved from microwave observations for decades, but typically at 20–50 km scale, which is too coarse for agricultural use (Peng et al., 2020). With recent developments in synthetic aperture radar (SAR) retrieval of soil moisture, observations have been available at 1 km scale from C-band SAR (Paloscia et al., 2013) or active-passive microwave data merging (Das et al., 2019). Some studies have also derived soil moisture data at 100 m or finer scale (Escorihuela et al., 2018; Lei et al., 2020; Vergopolan et al., 2020; Ma et al., 2020). Though remote sensing soil moisture data may not currently be available on a daily basis at high resolution, improvement in model simulation is still expected as is suggested by

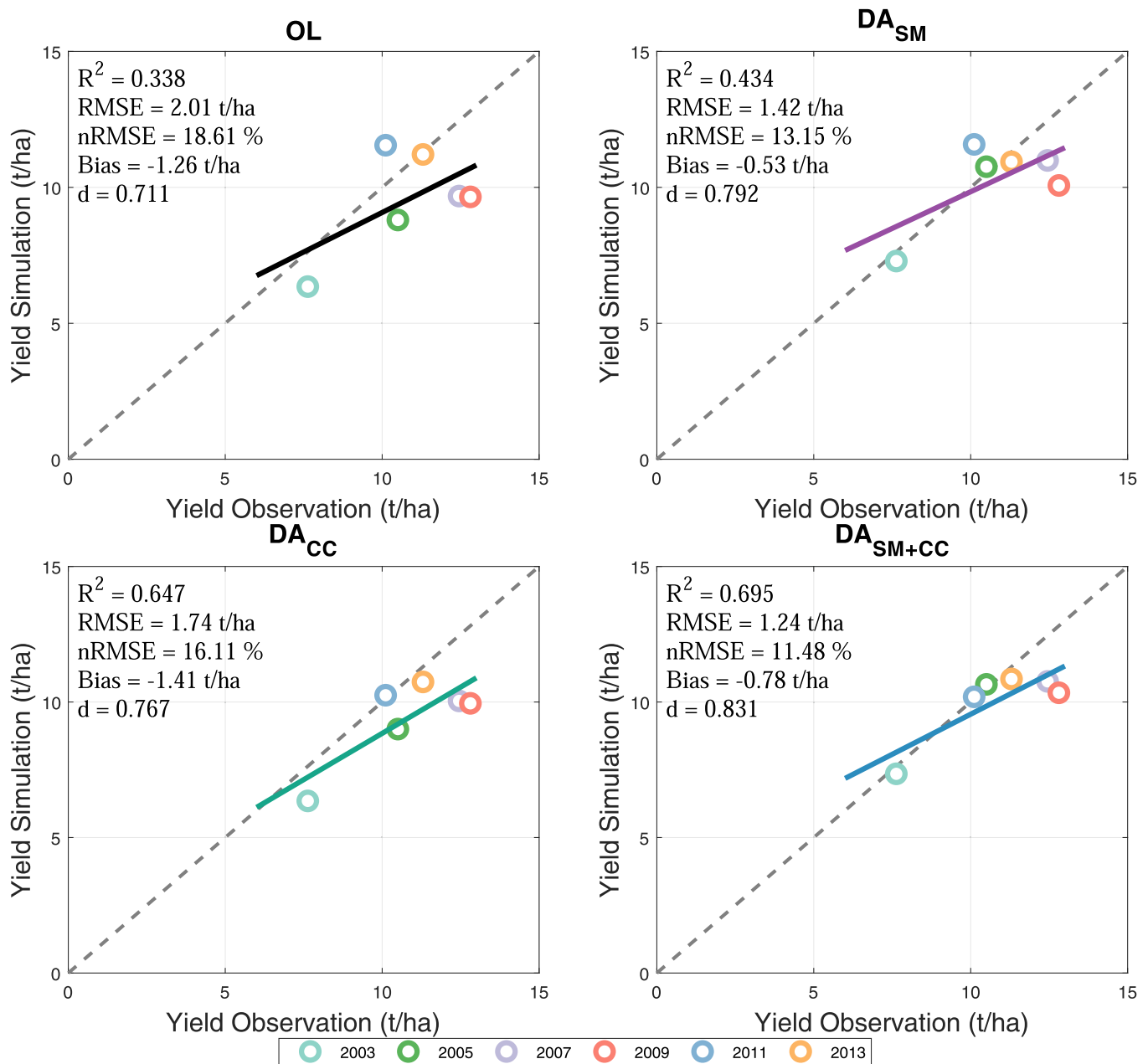


Fig. 8. Yield estimates using different strategies from the 6 growing seasons. The thick solid lines represent the simulation-observation linear regression.

Fig. 10. Canopy cover data on the other hand, may be retrieved from global operational platforms such as LandSat (30 m, 14-daily) and Sentinel-2 (10 m, 5-daily), which have been used in agricultural monitoring (Jin et al., 2019) and simulation (Kang and Özdoğan, 2019). Additionally, even higher-resolution canopy cover data may be derived from air-borne unmanned aerial vehicles (UAVs) or space-borne CubeSats. UAVs are very effective in providing centimeter-level measurements similar to traditional field observations, but covering a much greater spatial domain with a unique top-down view (McCabe et al., 2017b; Ziliani et al., 2018). The CubeSat concept is to make satellites smaller, lighter and cheaper (Puig-Suari et al., 2001), thus easier to build a constellation with hundreds of satellites. For example, the Planet Labs company has launched over 280 CubeSats since 2013 (McCabe et al., 2017a), which provides a capacity of sub-daily global coverage. Based on CubeSat data, daily NDVI and LAI data at 3 m resolution have been retrieved (Houborg and McCabe, 2018a,b), which significantly exceeds the LAI data availability in this study. With this level of observation

capacity, the spatial variability of canopy cover can be accurately characterized, and the simulation uncertainties can be effectively constrained.

The reference crop phenological parameters may be obtained from previous studies performed under similar environments, as is the case in this study. Alternatively, the parameters may be calibrated using in situ data in one field and applied to the whole study area. The uniform scaling parameter range was determined manually in this study, but the method can be easily improved to automatically determine the phenological scaling factor by seeking the scaling factor that generates the highest correlation between simulated and observed canopy cover, which facilitates remote sensing applications. In addition, many studies have suggested that time series of remote sensing vegetation indices can be used to detect crop phenology (Sakamoto et al., 2005; Pan et al., 2015; Zeng et al., 2020; Diao, 2020), and that planting date can also be detected with an RMSE of less than 2 days using CubeSat images (Sadeh et al., 2019).

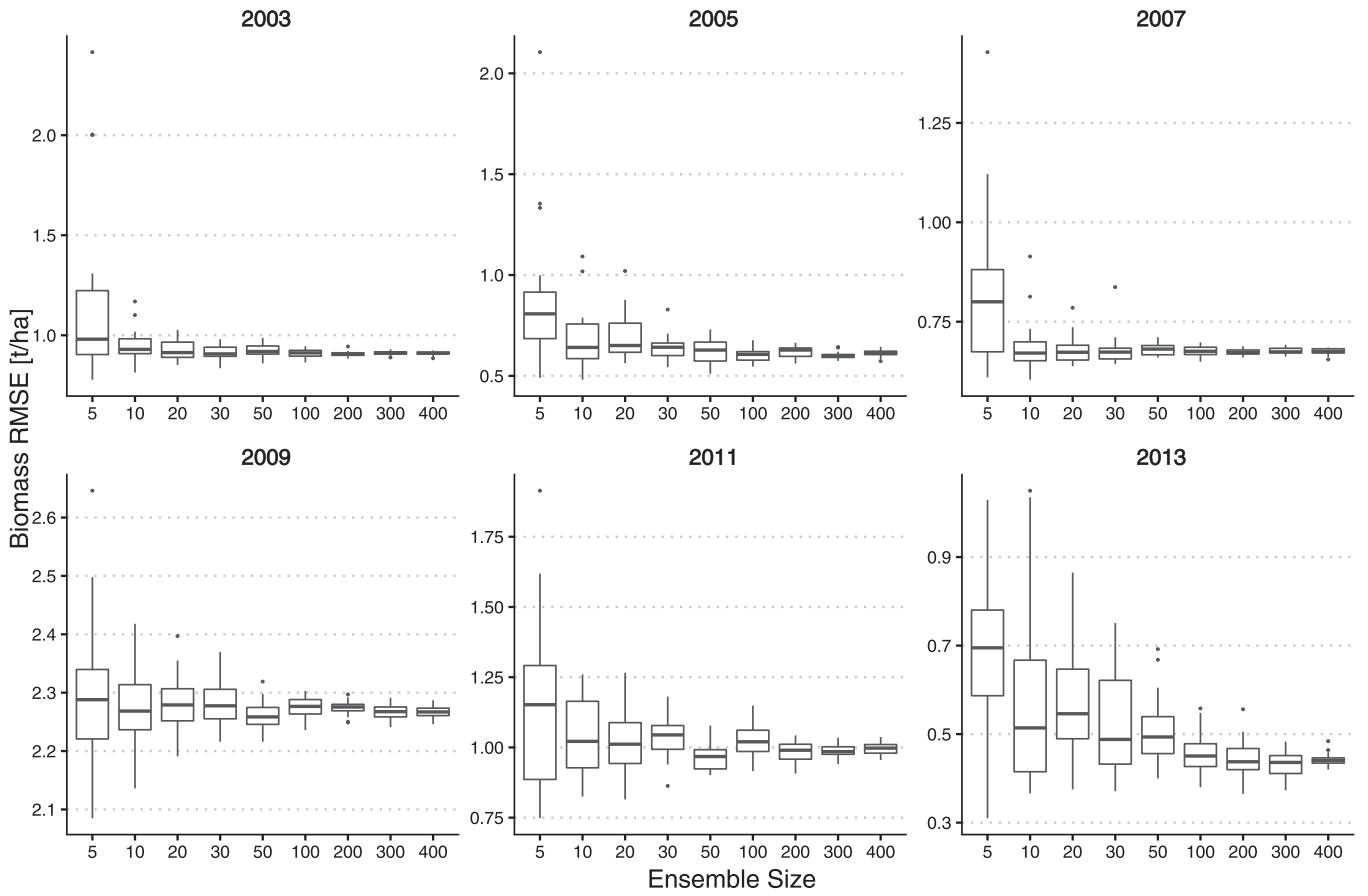


Fig. 9. Biomass estimation RMSE from DA_{SM+CC} with increasing ensemble size in the 6 growing seasons.

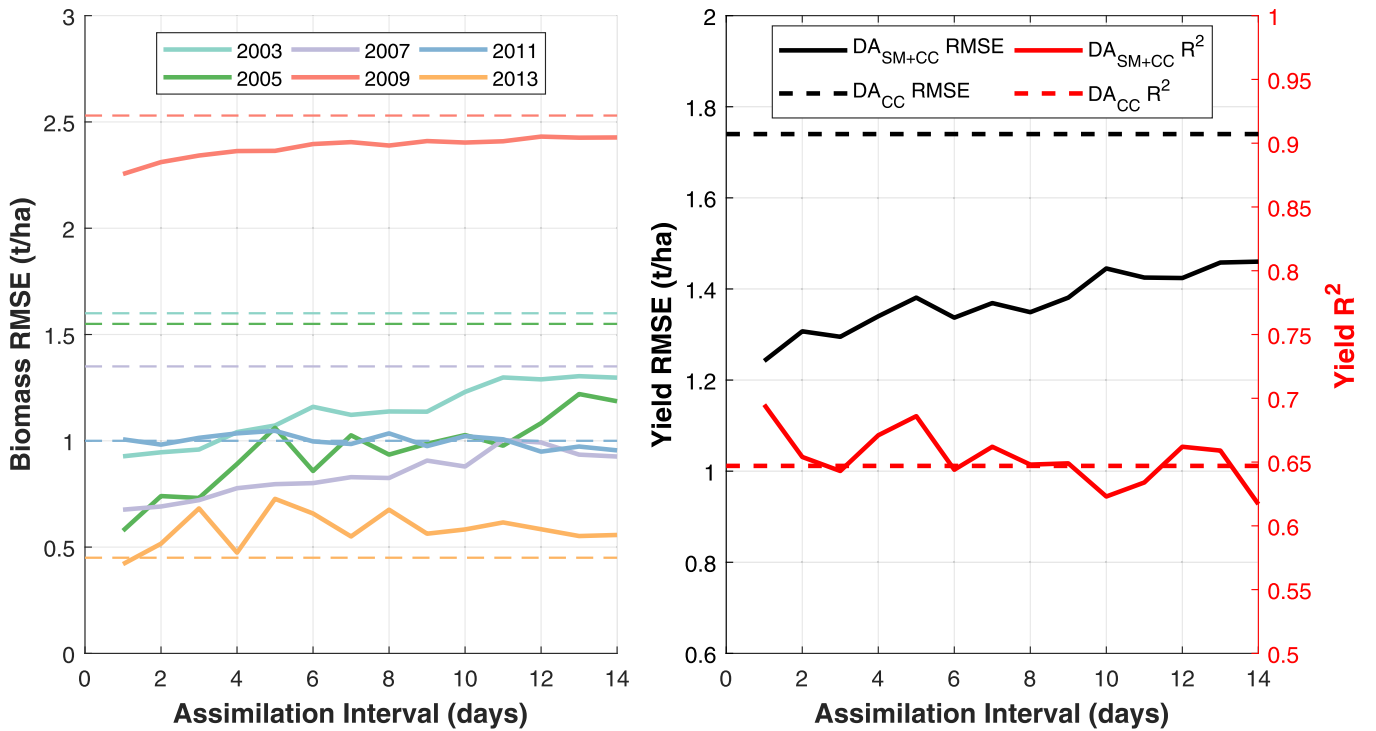


Fig. 10. Biomass and yield estimation RMSE using DA_{SM+CC} with soil moisture assimilated every 1–14 days. The dashed lines in the biomass plot represent the RMSEs using DA_{CC} for each growing season, respectively.

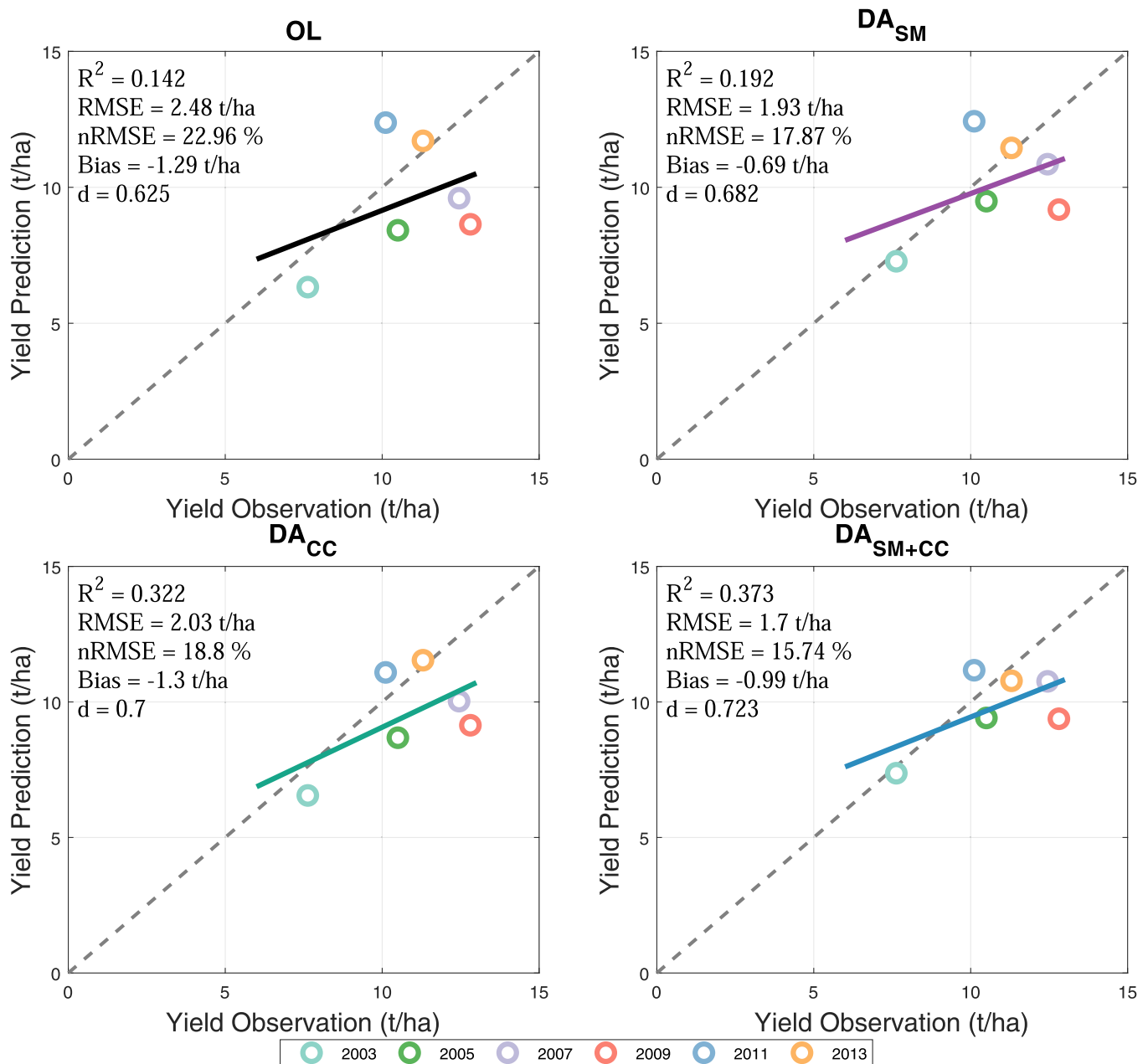


Fig. 11. Yield prediction from OL and the three assimilation strategies using only soil moisture and canopy cover observations within 90 days after planting.

4.3. Potential for yield prediction using remote sensing

Aside from estimating maize yield using all available observations in the growing season in a retrospective manner, this study also shows potential for yield prediction using observations from early stages in the growing season. Maize normally reaches maximum canopy cover before the 90th day after planting, with senescence yet to start. As a result, the phenological scaling factor determined using data within 90 days after planting captures the canopy cover dynamics in the vegetative growth stages, but may not well characterize the canopy decline in some years. This can be seen in the yield difference for 2009 and 2011 in Figs. 8 and 11. Despite this, DA_{SM+CC} still significantly improves yield estimates in almost all years (nRMSE = 15.74%) with good accuracy (Jamieson et al., 1991). The lead time and prediction accuracy are comparable to other yield prediction studies (Peng et al., 2018; Kang et al., 2020). Here only the assimilation of data within 90 days after planting (around 3 months before harvest) was tested. If more observations are used, the yield

prediction accuracy is expected to improve and converge to the performance in a retrospective application. This implies that as remote sensing data become available during the growing season, they can be assimilated sequentially into the model for improved yield prediction. This provides potential for an operational yield forecast system. In addition, prior knowledge of the crop phenology may also improve prediction accuracy. If data within 90 days after planting are assimilated with phenological scaling factors determined using all data, the yield estimation (included in Fig. S1 in the Supplementary material) accuracy is comparable to the retrospective estimation in Fig. 8. This suggests that with a priori knowledge of crop phenology, the yield prediction can be further improved.

4.4. Potential improvement in future studies

In this study, only canopy cover and surface soil moisture observations were assimilated to mimic a remote sensing application. Though

root-zone soil moisture is not observable from remote sensing, deeper-layer soil moisture data can be simulated by land surface models, which may provide additional information on crop simulation. To evaluate the added-value of deeper-layer soil moisture on crop modeling, the experiment is rerun by assimilating canopy cover and the soil moisture profile observations (included in Fig. S2 in the Supplementary material). The results suggest that assimilating soil moisture profile data does not lead to improved yield estimates. This may relate to the improvement in water stress determination in AquaCrop v6.0, which essentially enhances the influence of top-layer (10 cm) soil moisture on crop growth (FAO, 2017). Consequently, the significance of adding deeper-layer soil moisture is reduced.

Yield estimates may be further improved if other variables containing complementary information are assimilated. For example, assimilation of biomass and *ET* observations may lead to improved yield estimates as they are closely related to yield formation. Recently, dry matter productivity (DMP) and *ET* products have been developed by Copernicus Global Land Service (Swinnen et al., 2015) and the NASA ECOSTRESS mission (Fisher et al., 2020; Anderson et al., 2021) at fine resolution, which may be incorporated into the assimilation framework. Alternatively, the AquaCrop model may be coupled to a land surface model to facilitate the assimilation of more variables. Using advanced data assimilation algorithms may also improve simulation performance, such as using smoothing techniques instead of filtering, or adopting joint state-parameter updates (also called state-augmentation) in the assimilation.

5. Conclusion

Parameter calibration is normally required for crop modeling but can be time-consuming or infeasible in regions with large heterogeneity. In this study, a method was proposed to improve maize simulation performance using data assimilation with reduced parameter calibration prior to simulation. The phenological parameters were determined by uniformly scaling the parameters calibrated in a different location for a different maize hybrid, and other key model parameter values were taken from default settings in the model documentation. The phenological scaling factors were randomly sampled within an arbitrary ± 0.05 range, and surface soil moisture and canopy cover observations were assimilated both jointly and separately to constrain simulation uncertainties using the ensemble Kalman filter. The methodology was applied for maize simulation in six growing seasons in a rain-fed field using AquaCrop, and the model estimates were evaluated against canopy cover, evapotranspiration, biomass and yield data. A further experiment was also conducted to evaluate the method for yield prediction using observations from early crop growth stages.

The results indicated that simulations with uniformly scaled phenological parameters successfully captured the trends of canopy cover and biomass development in the growing season, but may lead to significant estimation bias. Assimilation of soil moisture and canopy cover observations effectively constrained simulation uncertainties and improved model performance. In general, soil moisture assimilation performed better when the water stress was high, while canopy cover observations were more informative under wet conditions. When soil moisture and canopy cover data were jointly assimilated, the model performed best, and the overall yield estimates (RMSE = 1.24 t/ha, nRMSE = 11.48%, $R^2 = 0.695$) were significantly improved over the no-assimilation case (RMSE = 2.01 t/ha, nRMSE = 18.61%, $R^2 = 0.338$). Sensitivity analyses suggested that the improvement was still evident with temporally sparse soil moisture observations and a small ensemble size. A yield prediction experiment using data within 90 days after planting suggested that the method was capable of predicting maize yield (RMSE = 1.7 t/ha, nRMSE = 15.74%) around 3 months before harvest. Overall, the proposed method can effectively improve maize yield estimation and prediction without performing a thorough parameter calibration prior to model simulation. This provides potential

for regional applications with remote sensing soil moisture and canopy cover data, particularly when the heterogeneity in the crop phenology is evident.

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.

Acknowledgments

This work was partly funded through the 'A new paradigm in precision agriculture: assimilation of ultra-fine resolution data into a crop-yield forecasting model' project, supported by the King Abdullah University of Science and Technology, Grant number OSR-2017-CRG6, and through the 'Building REsearch Capacity for sustainable water and food security In drylands of sub-saharan Africa (BRECCIA)' project, which is supported by UK Research and Innovation as part of the Global Challenges Research Fund, Grant number NE/P021093/1. The authors thank David Scoby from University of Nebraska-Lincoln for providing data.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.agwat.2021.106884.

References

- Ajaere, N.M., 2012. Multi-Temporal Analysis of Crop Biomass Using Selected Environmental Variables and Remote Sensing Derived Indices.
- Allen, R., Pereira, L., Raes, D., Smith, M., 1998. Crop evapotranspiration: guidelines for computing crop water requirements. *Irrig. Drain. Pap.* 56, 300.
- Anderson, M.C., Yang, Y., Xue, J., Knipper, K.R., Yang, Y., Gao, F., Hain, C.R., Kustas, W. P., Cawse-Nicholson, K., Hulley, G., 2021. Interoperability of ECOSTRESS and landsat for mapping evapotranspiration time series at sub-field scales. *Remote Sens. Environ.* 252, 112189.
- Araya, A., Hoogenboom, G., Luedeling, E., Hadgu, K.M., Kisekka, I., Martorano, L.G., 2015. Assessment of maize growth and yield using crop models under present and future climate in southwestern Ethiopia. *Agric. For. Meteorol.* 214, 252–265.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* 82, 2415–2434.
- Boote, K., 2019. *Advances in Crop Modelling for a Sustainable Agriculture*. Burleigh Dodds Science Publishing.
- Boote, K.J., Jones, J.W., Pickering, N.B., 1996. Potential uses and limitations of crop models. *Agron. J.* 88, 704–716.
- Boote, K.J., Jones, J.W., White, J.W., Asseng, S., Lizaso, J.L., 2013. Putting mechanisms into crop production models. *Plant Cell Environ.* 36, 1658–1672.
- Bouman, B., Van Keulen, H., Van Laar, H., Rabbinge, R., 1996. The 'School of de Wit' crop growth simulation models: a pedigree and historical overview. *Agric. Syst.* 52, 171–198.
- Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoulaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Dürr, C., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie* 18, 311–346.
- Chakrabarti, S., Bongiovanni, T., Judge, J., Zotarelli, L., Bayer, C., 2014. Assimilation of SMOS soil moisture for quantifying drought impacts on crop yield in agricultural regions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7, 3867–3879.
- Chen, H., Wu, H., Li, Z.L., Tang, B.H., Tang, R., Yan, G., 2019. Spatial upscaling of remotely sensed leaf area index based on discrete wavelet transform. *Int. J. Remote Sens.* 40, 2343–2358.
- Chen, Y., Zhang, Z., Tao, F., 2018. Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. *Eur. J. Agron.* 101, 163–173.
- Das, N.N., Entekhabi, D., Dunbar, R.S., Chaubell, M.J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P.E., 2019. The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sens. Environ.* 233, 111380.
- De Wit, A.d., Van Diepen, C., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agric. For. Meteorol.* 146, 38–56.
- Diao, C., 2020. Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. *Remote Sens. Environ.* 248, 111960 <https://doi.org/10.1016/j.rse.2020.111960>. (<http://www.sciencedirect.com/science/article/pii/S0034425720303308>).

- Dorigo, W.A., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinf.* 9, 165–193.
- Escorihuela, M.J., Merlin, O., Stefan, V., Moyano, G., Eweys, O.A., Zribi, M., Kamara, S., Benahi, A.S., Ebbe, M.A.B., Chihirane, J., 2018. SMOS based high resolution soil moisture estimates for desert locust preventive management. *Remote Sens. Appl. Soc. Environ.* 11, 140–150.
- Estes, L., Bradley, B., Beukes, H., Hole, D., Lau, M., Oppenheimer, M., Schulze, R., Tadross, M., Turner, W., 2013a. Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. *Glob. Ecol. Biogeogr.* 22, 1007–1018.
- Estes, L.D., Beukes, H., Bradley, B.A., Debats, S.R., Oppenheimer, M., Ruane, A.C., Schulze, R., Tadross, M., 2013b. Projected climate impacts to South African maize and wheat production in 2055: a comparison of empirical and mechanistic modeling approaches. *Glob. Chang. Biol.* 19, 3762–3774.
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53, 343–367.
- Fang, H., Liang, S., Hoogenboom, G., Teasdale, J., Cavigelli, M., 2008. Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. *Int. J. Remote Sens.* 29, 3011–3032.
- FAO, 2017. **Aquacrop Update and New Features: Version 6.0. Food and Agriculture Organization of the United Nations, Rome, Italy.**
- Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., 2020. ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station. *Water Resour. Res.* 56, e2019WR026058.
- Foolad, F., Franz, T.E., Wang, T., Gibson, J., Kilic, A., Allen, R.G., 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. *Hydrol. Earth Syst. Sci.* 21, 1263.
- Foster, T., Brozović, N., Butler, A., Neale, C., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: an open source version of FAO's crop water productivity model. *Agric. Water Manag.* 181, 18–22.
- Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sens. Environ.* 105, 286–298.
- Grassini, P., van Bussel, L.G., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., van Ittersum, M.K., Cassman, K.G., 2015. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crop. Res.* 177, 49–63.
- Gupta, H.V., Beven, K.J., Wagener, T., 2006. Model calibration and uncertainty estimation. *Encycl. Hydrol. Sci.*
- Heng, L.K., Hsiao, T., Evett, S., Howell, T., Steduto, P., 2009. Validating the FAO AquaCrop model for irrigated and water deficient field maize. *Agron. J.* 101, 488–498.
- Houborg, R., McCabe, M.F., 2018a. A cubesat enabled spatio-temporal enhancement method (CESTEM) utilizing planet, Landsat and MODIS data. *Remote Sens. Environ.* 209, 211–226.
- Houborg, R., McCabe, M.F., 2018b. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. *Remote Sens.* 10, 890.
- Hsiao, T.C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., Fereres, E., 2009. AquaCrop—the FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. *Agron. J.* 101, 448–459.
- Hu, S., Shi, L., Huang, K., Zha, Y., Hu, X., Ye, H., Yang, Q., 2019. Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation. *Field Crop. Res.* 232, 49–61.
- Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.H., Wu, Y., 2019. Assimilation of remote sensing into crop growth models: current status and perspectives. *Agric. For. Meteorol.* 276, 107609.
- Hunt, E.D., Svoboda, M., Wardlaw, B., Hubbard, K., Hayes, M., Arkebauer, T., 2014. Monitoring the effects of rapid onset of drought on non-irrigated maize with agronomic data and climate-based drought indices. *Agric. For. Meteorol.* 191, 1–11.
- Hunt, L., Boote, K., 1998. Data for model operation, calibration, and evaluation. *Understanding Options for Agricultural Production*. Springer, pp. 9–39.
- Ines, A.V., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* 138, 149–164.
- Jamieson, P., Porter, J., Wilson, D., 1991. A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. *Field Crop. Res.* 27, 337–350.
- Jégo, G., Pattey, E., Bourgeois, G., Morrison, M., Drury, C., Tremblay, N., Tremblay, G., 2010. Calibration and performance evaluation of soybean and spring wheat cultivars using the STICS crop model in Eastern Canada. *Field Crop. Res.* 117, 183–196.
- Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020. Estimation of maize yield by assimilating biomass and canopy cover derived from hyperspectral data into the AquaCrop model. *Agric. Water Manag.* 227, 105846.
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* 228, 115–128.
- Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., 2017. Brief history of agricultural systems modeling. *Agric. Syst.* 155, 240–254.
- Kalman, R., 1960. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* 82, 35–44.
- Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with landsat using a hierarchical data assimilation approach. *Remote Sens. Environ.* 228, 144–163.
- Kang, Y., Özdoğan, M., Zhu, X., Ye, Z., Hain, C., Anderson, M., 2020. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environ. Res. Lett.* 15, 064005.
- Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N., Meinke, H., Hochman, Z., 2003. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* 18, 267–288.
- Lei, F., Crow, W.T., Kustas, W.P., Dong, J., Yang, Y., Knipper, K.R., Anderson, M.C., Gao, F., Notarnicola, C., Greifeneder, F., 2020. Data assimilation of high-resolution thermal and radar remote sensing retrievals for soil moisture monitoring in a drip-irrigated vineyard. *Remote Sens. Environ.* 239, 111622.
- 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. *Ecol. Model.* 291, 15–27.
- Linker, R., Ioslovich, I., 2017. Assimilation of canopy cover and biomass measurements in the crop model AquaCrop. *Biosyst. Eng.* 162, 57–66.
- Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resour. Res.* 43.
- Lu, Y., Dong, J., Steele-Dunne, S.C., 2019. Impact of soil moisture data resolution on soil moisture and surface heat flux estimates through data assimilation: a case study in the Southern Great Plains. *J. Hydrometeorol.* 20, 715–730.
- Ma, C., Li, X., McCabe, M.F., 2020. Retrieval of high-resolution soil moisture through combination of Sentinel-1 and Sentinel-2 data. *Remote Sens.* 12, 2303.
- Ma, H., Huang, J., Zhu, D., Liu, J., Su, W., Zhang, C., Fan, J., 2013. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST-ACRM model with Ensemble Kalman Filter. *Math. Comput. Model.* 58, 759–770.
- Marin, F., Jones, J.W., Boote, K.J., 2017. A stochastic method for crop models: including uncertainty in a sugarcane model. *Agron. J.* 109, 483–495.
- McCabe, M., Aragon, B., Houborg, R., Mascaro, J., 2017a. CubeSats in hydrology: ultrahigh-resolution insights into vegetation dynamics and terrestrial evaporation. *Water Resour. Res.* 53, 10017–10024.
- McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucier, A., Houborg, R., Verhoest, N.E., Franz, T.E., 2017b. The future of Earth observation in hydrology. *Hydrol. Earth Syst. Sci.* 21, 3879.
- Mebane, V.J., Day, R.L., Hamlett, J.M., Watson, J.E., Roth, G.W., 2013. Validating the FAO AquaCrop model for rainfed maize in Pennsylvania. *Agron. J.* 105, 419–427.
- Morell, F.J., Yang, H.S., Cassman, K.G., Van Wart, J., Elmore, R.W., Licht, M., Coulter, J. A., Ciampitti, I.A., Pittelkow, C.M., Brouder, S.M., 2016. Can crop simulation models be used to predict local to regional maize yields and total production in the US Corn Belt? *Field Crop. Res.* 192, 1–12.
- Nearing, G.S., Crow, W.T., Thorp, K.R., Moran, M.S., Reichle, R.H., Gupta, H.V., 2012. Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: an observing system simulation experiment. *Water Resour. Res.* 48.
- Nguy-Robertson, A., Suyker, A., Xiao, X., 2015. Modeling gross primary production of maize and soybean croplands using light quality, temperature, water stress, and phenology. *Agric. For. Meteorol.* 213, 160–172.
- Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., Reppucci, A., 2013. Soil moisture mapping using Sentinel-1 images: algorithm and preliminary validation. *Remote Sens. Environ.* 134, 234–248.
- Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. *Eur. J. Agron.* 35, 103–114.
- Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G.A., Yan, J., Liu, J., 2015. Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. *Int. J. Appl. Earth Obs. Geoinf.* 34, 188–197.
- Pauwels, V.R., Verhoest, N.E., De Lannoy, G.J., Guissard, V., Lucau, C., Defourny, P., 2007. Optimization of a coupled hydrology-crop growth model through the assimilation of observed soil moisture and leaf area index values using an ensemble Kalman filter. *Water Resour. Res.* 43.
- Peng, B., Guan, K., Pan, M., Li, Y., 2018. Benefits of seasonal climate prediction and satellite data for forecasting US maize yield. *Geophys. Res. Lett.* 45, 9662–9671.
- Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M.H., Crow, W.T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M.W., 2020. A roadmap for high-resolution satellite soil moisture applications—confronting product characteristics with user requirements. *Remote Sens. Environ.*, 112162.
- Puig-Suari, J., Turner, C., Ahlgren, W., 2001. Development of the standard CubeSat deployer and a CubeSat class PicoSatellite. In: *Proceedings of the IEEE Aerospace Conference* (cat. No. 01TH8542), 1, p. 347.
- Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. AquaCrop—the FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agron. J.* 101, 438–447.
- Raes, D., Steduto, P., Hsiao, T., Fereres, E., 2012. Chapter 3. Calculation procedures. *Reference Manual AquaCrop*.
- Raes, D., Steduto, P., Hsiao, T., Fereres, E., 2018. AquaCrop-The FAO Crop Model to Simulate Yield Response to Water: Reference Manual Annexes. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Ran, H., Kang, S., Hu, X., Li, S., Wang, W., Liu, F., 2020. Capability of a solar energy-driven crop model for simulating water consumption and yield of maize and its comparison with a water-driven crop model. *Agric. For. Meteorol.* 287, 107955.
- Reichle, R.H., 2008. Data assimilation methods in the Earth sciences. *Adv. Water Resour.* 31, 1411–1418.
- Sadeh, Y., Zhu, X., Chenu, K., Dunkerley, D., 2019. Sowing date detection at the field scale using CubeSats remote sensing. *Comput. Electron. Agric.* 157, 568–580.

- Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., Ohno, H., 2005. A crop phenology detection method using time-series MODIS data. *Remote Sens. Environ.* 96, 366–374.
- Sandhu, R., Irmak, S., 2019a. Assessment of AquaCrop model in simulating maize canopy cover, soil-water, evapotranspiration, yield, and water productivity for different planting dates and densities under irrigated and rainfed conditions. *Agric. Water Manag.* 224, 105753.
- Sandhu, R., Irmak, S., 2019b. Performance of AquaCrop model in simulating maize growth, yield, and evapotranspiration under rainfed, limited and full irrigation. *Agric. Water Manag.* 223, 105687.
- 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.
- Seidel, S.J., Palosuo, T., Thorburn, P., Wallach, D., 2018. Towards improved calibration of crop models—where are we now and where should we go? *Eur. J. Agron.* 94, 25–35.
- Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agron. J.* 101, 426–437.
- Steduto, P., 2003. Biomass water-productivity. Comparing the growth-engines of crop models. *FAO Expert Consultation on Crop Water Productivity under Deficient Water Supply*, pp. 26–28.
- Suyker, A.E., Verma, S.B., 2008. Interannual water vapor and energy exchange in an irrigated maize-based agroecosystem. *Agric. For. Meteorol.* 148, 417–427.
- Swinnen, E., Van Hoolst, R., Toté, C., 2015. Gio Global Land Component-Lot I “Operation of the Global Land Component”. Framework Service Contract N 388533 (JRC). Quality Assessment Report. Dry Matter Productivity (DMP)-PROBA-V, Algorithm Theoretical Basis Document, Issue II. 01. VITO, EC Copernicus Global Land, Brussels.
- Twine, T.E., Kustas, W., Norman, J., Cook, D., Houser, P., Meyers, T., Prueger, J., Starks, P., Wesely, M., 2000. Correcting eddy-covariance flux underestimates over a grassland. *Agric. For. Meteorol.* 103, 279–300.
- Vergopolan, N., Chaney, N.W., Beck, H.E., Pan, M., Sheffield, J., Chan, S., Wood, E.F., 2020. Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. *Remote Sens. Environ.* 242, 111740.
- Verma, S.B., Dobermann, A., Cassman, K.G., Walters, D.T., Knops, J.M., Arkebauer, T.J., Suyker, A.E., Burba, G.G., Amos, B., Yang, H., 2005. Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems. *Agric. For. Meteorol.* 131, 77–96.
- Wallach, D., 2011. Crop model calibration: a statistical perspective. *Agron. J.* 103, 1144–1151.
- Wang, J., Li, X., Lu, L., Fang, F., 2013. Estimating near future regional corn yields by integrating multi-source observations into a crop growth model. *Eur. J. Agron.* 49, 126–140.
- Willmott, C.J., 1981. On the validation of models. *Phys. Geogr.* 2, 184–194.
- Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., Wang, L., 2017. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-wheat model. *Agric. For. Meteorol.* 246, 194–206.
- Yin, J., Zhan, X., Zheng, Y., Hain, C.R., Liu, J., Fang, L., 2015. Optimal ensemble size of ensemble Kalman filter in sequential soil moisture data assimilation. *Geophys. Res. Lett.* 42, 6710–6715.
- Zeng, L., Wardlaw, B.D., Xiang, D., Hu, S., Li, D., 2020. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sens. Environ.* 237, 111511.
- Zhang, Z., Li, Z., Chen, Y., Zhang, L., Tao, F., 2020. Improving regional wheat yields estimations by multi-step-assimilating of a crop model with multi-source data. *Agric. For. Meteorol.* 290, 107993.
- Ziliani, M.G., Parkes, S.D., Hoteit, I., McCabe, M.F., 2018. Intra-season crop height variability at commercial farm scales using a fixed-wing UAV. *Remote Sens.* 10, 2007.