



# Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction



Amor V.M. Ines<sup>a,\*</sup>, Narendra N. Das<sup>b</sup>, James W. Hansen<sup>a</sup>, Eni G. Njoku<sup>b</sup>

<sup>a</sup> International Research Institute for Climate and Society, The Earth Institute at Columbia University, 61 Route 9W Palisades, NY 10964, United States

<sup>b</sup> NASA, Jet Propulsion Laboratory, M/S 300-323, 4800 Oak Grove Drive, Pasadena, CA 91109, United States

## ARTICLE INFO

### Article history:

Received 10 November 2012

Received in revised form 28 June 2013

Accepted 1 July 2013

Available online 15 August 2013

### Keywords:

Yield forecasting

Crop model

Data assimilation

Soil moisture

Leaf area index

## ABSTRACT

To improve the prediction of crop yields at an aggregate scale, we developed a data assimilation–crop modeling framework that incorporates remotely sensed soil moisture and leaf area index (LAI) into a crop model using sequential data assimilation. The core of the framework is an Ensemble Kalman Filter (EnKF) used to control crop model runs, assimilate remote sensing (RS) data and update model state variables. We modified the Decision Support System for Agro-technology Transfer – Cropping System Model (DSSAT-CSM)–Maize model (Jones et al., 2003) to be able to stop and start simulations at any given time in the growing season, such that the EnKF can update model state variables as RS data become available. The data assimilation–crop modeling framework was evaluated against 2003–2009 maize yields in Story County, Iowa, USA, assimilating AMSR-E soil moisture and MODIS-LAI data independently and simultaneously. Assimilating LAI or soil moisture independently slightly improved the correlation of observed and simulated yields ( $R = 0.51$  and  $0.50$ ) compared to no data assimilation (open-loop;  $R = 0.47$ ) but prediction errors improved with reductions in MBE and RMSE by  $0.5$  and  $0.5 \text{ Mg ha}^{-1}$  respectively for LAI assimilation while these were reduced by  $1.8$  and  $1.1 \text{ Mg ha}^{-1}$  for soil moisture assimilation. Yield correlation improved more when both soil moisture and LAI were assimilated ( $R = 0.65$ ) suggesting a cause–effect interaction between soil moisture and LAI, prediction errors (MBE and RMSE) were also reduced by  $1.7$  and  $1.8 \text{ Mg ha}^{-1}$  with respect to open-loop simulations. Results suggest that assimilation of LAI independently might be preferable when conditions are extremely wet while assimilation of soil moisture + LAI might be more suitable when conditions are more nominal. AMSR-E soil moisture tends to be more biased under the presence of high vegetation (i.e., when crops are fully developed) and that updating rootzone soil moisture by near-surface soil moisture assimilation under very wet conditions could increase the modeled percolation causing excessive nitrogen (N) leaching hence reducing crop yields even with water stress reduced at a minimum due to soil moisture assimilation. However, applying the data assimilation–crop modeling framework strategically by considering a-priori information on climate condition expected during the growing season may improve yield prediction performance substantially, in our case with higher correlation ( $R = 0.80$ ) and more reductions in MBE and RMSE ( $2.5$  and  $3.3 \text{ Mg ha}^{-1}$ ) compared to when there is no data assimilation. Scaling AMSR-E soil moisture to the climatology of the model did not improve our data assimilation results because the model is also biased. Better soil moisture products e.g., from Soil Moisture Active Passive (SMAP) mission, may solve the soil moisture data issue in the near future.

© 2013 The Authors. Published by Elsevier Inc. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

## 1. Introduction

When a crop model is used to predict crop yields early in the growing season, two sources of uncertainties prevail – those coming from climate and model uncertainties (Hansen, Challinor, Ines, Wheeler, & Moron, 2006). Climate-related uncertainty is greatest early in the growing season but tends to decrease as weather data

become available as growing season progresses. Model-related uncertainty due to errors in model structure, modeling assumptions and other ancillary data, generally remains constant through the growing season. Skillful climate forecasts can reduce climate-related uncertainty in crop yield prediction especially at the earlier stages of the growing season, while model-related uncertainty can potentially be reduced by assimilating remote sensing (RS) data during the growing season (de Wit & Van Diepen, 2007; Hansen et al., 2006; Vazifedoust, Van Dam, Bastiaanssen, & Feddes, 2009).

Remote sensing has been incorporated into crop simulation models either as a forcing function or simulation steering (Bouman, Van Diepen, Vossen, & Van Der Val, 1997). Forcing function is applied to replace simulated state variable with the RS observation while simulation

\* Corresponding author. Tel.: +1 8456804490.  
E-mail address: [ines@iri.columbia.edu](mailto:ines@iri.columbia.edu) (A.V.M. Ines).

steering is used to re-initialize (e.g., sowing date, planting density) or re-parameterize (e.g., canopy and growth parameters) the crop model in a way that minimizes the difference between simulated and measured data. Examples of the simulation steering approach include the works of Bouman (1992), Olivos et al. (2005), Fang, Liang, and Hoogenboom (2011) and Thorp et al. (2012) who linked radiative transfer models with crop models. Ines, Honda, Gupta, Droogers, and Clemente (2006) used remotely sensed evapotranspiration to re-parameterize soil properties, crop and water management parameters of a pseudo-regional Soil–Water–Atmosphere–Plant (SWAP) model.

When RS data are used to replace the value of a model-simulated state variable or to infer some soil–plant–atmosphere–continuum properties, one assumes that the RS data are free of error or assumes that the level of data error is acceptable to be propagated within the simulated system (Fang, Liang, Hoogenboom, Teasdale, & Cavigelli, 2008; Ines & Mohanty, 2008a,b,c; Ines & Mohanty, 2009). Thorp, Hunsaker, and French (2010) assimilated measured Leaf Area Index (LAI) in the DSSAT-CSM-Wheat model using forcing and updating mechanisms. The updating mechanism is a forcing scheme that accounts for back propagation of the change in LAI to the system. Their simple assimilation procedure is more successful in minimizing errors in ET and canopy weight, but had difficulty improving yield simulations because yield is controlled by other factors aside from LAI. Vazifedoust et al. (2009) conducted a simple sequential data assimilation using a constant gain Kalman filter to assimilate LAI and the ratio of actual ET to potential ET ( $ET/ET_p$ ) in SWAP-WOFOST and found significant improvements in simulated total dry matter but only one among three of their observation fields showed significant improvements in simulated yield. It should be noted that timing of the use and frequency of LAI data assimilation in crop model is critical as LAI (or NDVI) is more directly related to yield at silking and grain filling (Ozalkan, Sepetoglu, Daur, & Sen, 2010; Sehgal, Sastri, Karla, & Dadhwal, 2005; Teal et al., 2006).

Sequential data assimilation is a robust way of combining model and observations to minimize the uncertainty of a given modeled state as it enhances the use of information between imperfect model and observations. Of the several algorithms (e.g., particle filter, Kalman filter) capable of performing data assimilation to update sequentially model states and parameters, the Monte Carlo-based Ensemble Kalman Filter (EnKF) is the one that is widely used (Evensen, 2003). EnKF received a lot of attention in the geosciences because of its ease of implementation, computational efficiency and optimum performance. It uses the Monte Carlo approach to approximate the conditional second-order moments of variables of interest using a finite number of randomly generated model replicates, then corrects model forecast and error covariance (Evensen, 2003; Houtekamer & Mitchell, 1998). Many studies have implemented EnKF to assimilate RS data in meteorological and hydrological models with considerable success (e.g., Crow & Wood, 2003; Das & Mohanty, 2006; Das, Mohanty, Cosh, & Jackson, 2008; Dunne & Entekhabi, 2005; Evensen, 2003; Keppenne & Rienecker, 2002; Reichle, McLaughlin, & Entekhabi, 2002).

As with any other models, crop simulation models are also subject to structural and data (input and forcing) errors hence they are imperfect in simulating the truth. Sequential data assimilation can be used to improve crop model performance without altering its structure by periodically updating state variables within the growing season with RS observations. RS of vegetation (e.g., LAI) and soil moisture are potentially useful for sequential data assimilation because of their obvious influence on crop growth, hence on crop yields. Their spatial and temporal coverage also allows data assimilation for crop forecasting at regional scale.

EnKF had been used with crop models recently with some success and challenges especially when assimilating LAI (e.g., Curnel, de Wit, Duveiller, & Defourny, 2011). Most of these studies however were conducted under hypothetical conditions, so-called forward–backward simulations, and could be limited to explaining fully the strengths and limitations of the method under actual conditions, especially at

predicting yield at aggregate scale (Curnel et al., 2011; Nearing et al., 2012). de Wit and Van Diepen (2007) showed the utility of RS-derived rootzone soil wetness index to correct some of the errors in the soil water balance associated with imperfect model inputs e.g., gridded rainfall data, in crop yield prediction.

In this paper, we developed a data assimilation–crop modeling framework for assimilating remotely sensed data with a crop model that could be used to improve crop yield forecasting at a given lead-time within the growing season. We present our implementation of an EnKF data assimilation system, development of the stand-alone DSSAT-CSM-Maize model, and testing and evaluation of the method under actual growing conditions in Story County, Iowa. The testing and evaluation aims to quantify the use of remotely sensed soil moisture and LAI to improve simulated yields within the data assimilation–crop modeling framework, independently and simultaneously. A variant of EnKF called an Ensemble Square Root Filter (Whitaker & Hamill, 2002) (but we termed it EnKF in general) was implemented for this study to simplify the use of RS data in the data assimilation, especially crop growth observations e.g., LAI, as the square root filter allows data assimilation without perturbing the observed data. This kind of work is important to improving the applications of data assimilation in crop yield forecasting.

## 2. Methods

### 2.1. Development of EnKF-DSSAT-CSM-Maize

#### 2.1.1. EnKF data assimilation system

The core of data assimilation lies in the Kalman filter system, which assumes that observations are related to the true state  $x_t$  (e.g., soil moisture or LAI at time  $t$ ) as:

$$y = Hx_t + \varepsilon \quad (1)$$

where  $y$  is the observation vector,  $\varepsilon$  is a Gaussian random error vector with a mean of zero and observation error covariance  $R$ , and  $H$  is the operator that maps the model variable space to the observation space. Furthermore, the forecast of  $x_t$  at  $t = k$  is Gaussian with mean  $x_{t=k}^f$  and error covariance  $P_{t=k}^f$ . Under these assumptions, the estimated state and error covariance is updated as:

$$x_{t=k}^a = x_{t=k}^f + K(y - Hx_{t=k}^f) \quad (2)$$

$$P_{t=k}^a = (I - KH)P_{t=k}^f \quad (3)$$

where  $f$  and  $a$  are indices of the prior (called forecast) and posterior (called analysis) estimates, respectively,  $t$  is an index of time,  $I$  is the identity matrix, and  $K$  is the Kalman gain matrix defined as

$$K = P_{t=k}^f H^T (HP_{t=k}^f H^T + R)^{-1} \quad (4)$$

The EnKF forecast and analysis error covariance come directly from an ensemble of model simulations:

$$P^f H^T = (N_e - 1)^{-1} \sum_{n=1}^{N_e} (x_n^f - \bar{x}^f) (Hx_n^f - H\bar{x}^f)^T \quad (5)$$

where  $N_e$  is the number of ensemble members,  $n$  is a running index for ensemble member, and  $\bar{x}^f$  represents the ensemble mean calculated as:

$$\bar{x}^f = N_e^{-1} \sum_{n=1}^{N_e} x_n^f \quad (6)$$

Usually, the ensemble is generated by perturbing the observed data. The variance used in the perturbation is based on the uncertainty of the

data. On the other hand, model parameters are perturbed to generate the ensemble of model runs. In this system, ensemble members are integrated independently and updated in accordance with the Kalman filter method when new observations become available.

In our study, however, a variant of the EnKF approach (i.e., Ensemble Square Root Filter) was sought to ensure that the analysis error covariance does not become unrealistically low because we intended not to perturb observations to minimize the risk of pairing LAI profiles with unusual planting date. Burgers, Van Leeuwen, and Evensen (1998) demonstrated that  $P^a$  is underestimated by a factor of  $(I - KH)$  when observations are not treated as random variables. This can cause the standard EnKF to reject observations in favor of the ensemble forecast, which can lead the analysis incrementally further away from reality, resulting in filter divergence (e.g., Burgers et al., 1998; Houtekamer & Mitchell, 1998; Mitchell & Houtekamer, 2000; Whitaker & Hamill, 2002). Whitaker and Hamill (2002) showed that adding random noise to observations further skews the distribution of  $P^a$  that resulted in a more erroneous analysis even though the covariance is increased. They suggested an alternative way of updating ensemble members, where the ensemble mean ( $\bar{x}_{t=k}^a$ ) is still updated by Eq. (2) but deviations ( $\tilde{x}_{t=k}^a$ ) from the mean are updated by:

$$\bar{x}_{t=k}^a = \bar{x}_{t=k}^f + K' (H\tilde{x}_{t=k}^f) \tag{7}$$

where ( $\bar{x}_{t=k}^f$ ) is a forecast value centered at the ensemble forecast mean,  $K' = \alpha K$  and  $\alpha = \left(1 + \sqrt{\frac{R}{HP'H^T + R}}\right)^{-1}$ . By this method the analysis error

covariance is guaranteed to be exactly equal to that of Eq. (2), and perturbed observations are no longer necessary.

Fig. 1 shows the overall framework of the EnKF data assimilation system. First, ensemble members are generated (in our case, 40, see Section 2.3.5), then independent crop model runs are invoked. For each model run, each time a new RS observation becomes available, the run is interrupted, EnKF updates the target model state variables, and the simulation is re-initialized with the updated states and re-run until the next update is available. The data assimilation step also includes small inflation parameters (1.05 for soil moisture and 1.5 for LAI) to perturb the forecast ensemble members in case their variability becomes too low. This step ensures that the observations are not systematically rejected during assimilation.

Earlier, we conceptualized the integration by embedding the EnKF within the DSSAT-CSM code (Jones et al., 2003). However, we found that this is not feasible if the data assimilation crop-modeling framework is not developed in a parallel computing infrastructure because the crop model should run multiple ensemble members simultaneously, then wait to update state variables and model parameters when RS data are available. That is why we developed a modified version of DSSAT-CSM-Maize (Section 2.1.2) that allows EnKF to control an ensemble of independent crop model runs.

With respect to the time scales of soil moisture process dynamics and the observed surface layer, which is typically shallow as compared to deeper soil layers, the propagation of near surface soil moisture information only through vertical model physics is relatively inefficient. In contrast, updating of the deeper layers soil moisture based on the modeled surface-rootzone soil moisture error correlations expressed

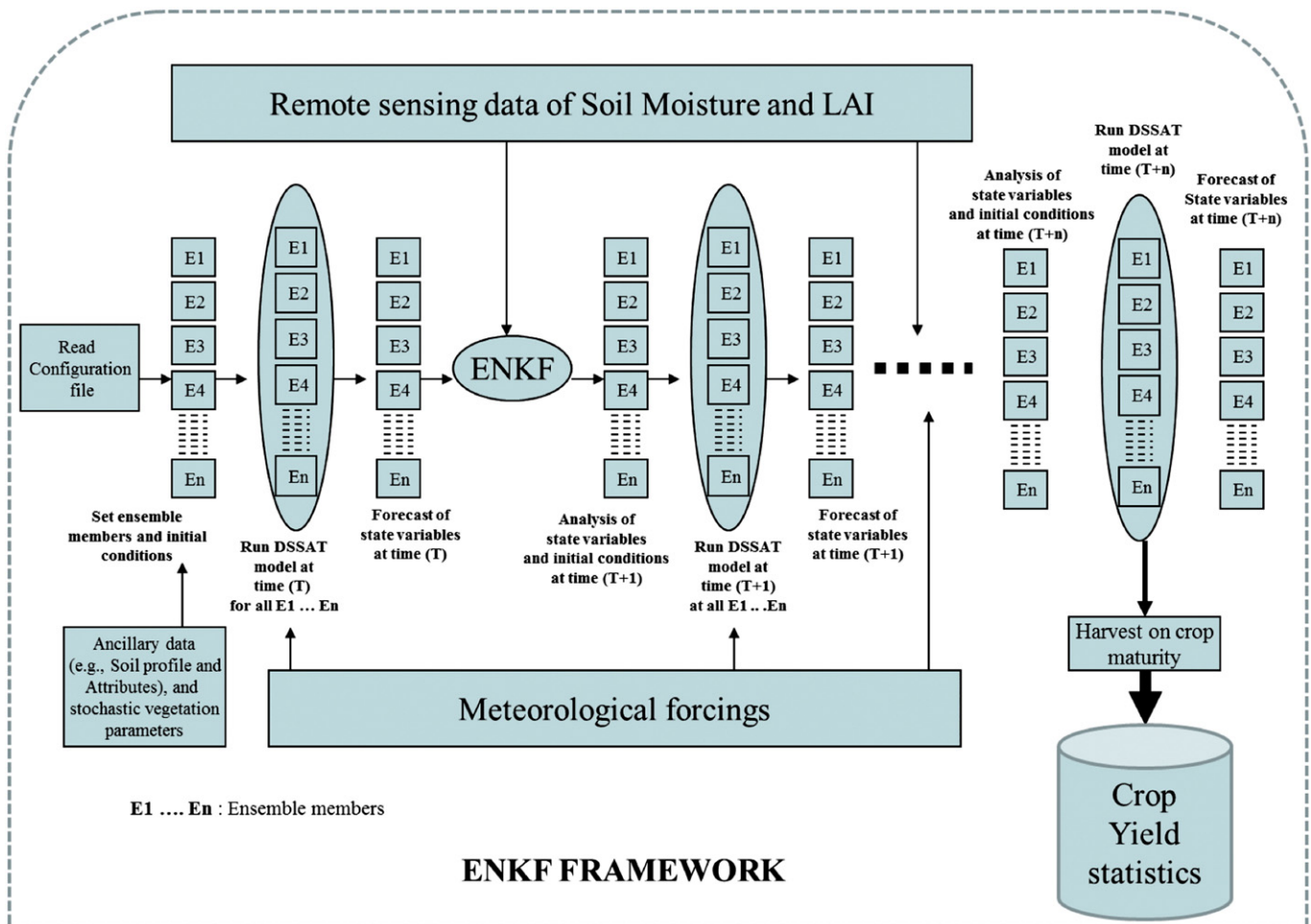


Fig. 1. Schematic of the data assimilation-crop modeling framework.



through the ensemble members can provide an efficient downward propagation of surface soil moisture (observation through remote sensing) information, as long as errors in the surface layer soil moisture are statistically correlated to errors in the soil moisture deeper layers via the model physics. The term  $P_t^f = kH^T$  in Eq. (4) is the cross-covariance between errors in the model state (for example, surface and rootzone soil moisture) and errors in the observed variables (i.e., the near surface soil moisture measurements). The DSSAT-CSM model propagation and update steps as illustrated in Fig. 1 mean that surface information is propagated into the rootzone in two ways. First, in the DSSAT-CSM model propagation step, soil moisture interactions take place between the surface and deeper layers according to the modeled soil moisture dynamics. Second, in the presence of a surface soil moisture observation, an increment to deeper layer soil moisture is computed and applied using the EnKF update step, that is based on the innovation (i.e., difference between predicted and observed) and the surface-rootzone error correlation (as expressed in the Kalman gain). All of these are done on a daily time step (Fig. 1).

The assimilation of near-surface RS soil moisture updates the rest of the rootzone soil moisture using the updating equations in Eqs. (7) and (2), for a 9-layered soil profile,  $H$  would be equal to  $(1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$  as the only layer with measurement is the first layer (see Section 2.2.2). Only near-surface and sub-surface soil moisture were updated here. The only link with the soil Nitrogen process is the updated profile soil moisture. The assimilation of RS LAI is also done using Eqs. (7) and (2), but now  $H = 1$ . However, LAI is related to other vegetation variables, using stage-state functions in DSSAT-CSM, we updated plant leaf area (PLA) and plant leaf weight (LFWT) using the updated LAI values. The updated plant variables are used to calculate carbohydrates (CARBO) for the next time step. In this study, the crop model propagates the link between the updated plant variables and plant Nitrogen process in the next time step.

### 2.1.2. Modified DSSAT-CSM-Maize

The DSSAT-CSM simulates growth, development and yield of a crop growing on a homogeneous land area, under prescribed or simulated management practices, as influenced by the dynamics of solar radiation, temperature, and soil water, carbon and nitrogen (Jones et al., 2003). Two dominant soil profiles in our study area were sampled to represent the landscape. Each realization was used as homogenous land unit to run the modified DSSAT-CSM-Maize. Crop models are generally designed to run continuously from sowing until maturity or harvest – unless the crop fails due to extreme stress. However, the Ensemble Kalman Filter must be able to interrupt the simulation of ensemble member as RS data become available, check and adjust target state variables, and re-start simulations with the updated state values as initial conditions. We therefore modified DSSAT-CSM-Maize to be able to stop and re-start at any point within the growing season, based on the timing and frequency of available RS data. Every time the modified DSSAT-CSM-Maize model stops, an ASCII file records the current values of model state variables and parameters. This file is used to access and modify assimilated state variables, and re-initialize the crop model when it is again invoked. We verified that, in the absence of data assimilation, model outputs using the EnKF-DSSAT-CSM-Maize implementation matched the original DSSAT-CSM version, within the bounds of truncation and rounding errors associated with the intermediate text file. Nitrogen (N) process was included in the simulations because we wanted to simulate actual yields. However, the only link with EnKF and the N-process is the updated soil moisture states.

## 2.2. Testing and evaluation

To test the data assimilation-crop modeling framework, we selected a large-scale, relatively homogenous farming system in Midwestern USA, Story County, Iowa, where the majority of the crops are maize

and soybean, but maize dominates the landscape (Fig. 2). In our study, we only considered maize (*Zea mays*) in the analysis.

### 2.2.1. Crop yields

Maize production was downloaded and processed for Story County from the USDA National Agricultural Statistics Service (NASS) website. Available data include area planted and harvested and corresponding yield. Yields measured in bushels were converted to mass units ( $\text{Mg ha}^{-1}$ ). The 2003–2009 mean yield was  $11.12 \text{ Mg ha}^{-1}$  with a standard deviation of  $0.7 \text{ Mg ha}^{-1}$ . The NASS yield data is the most readily available estimate of maize yield for aggregate modeling in the county. Plot and field scale yield data averaged across the county may be more accurate to represent county yield but were not available.

### 2.2.2. Soils

Soil data in shapefile format were downloaded and processed for Iowa from the Soil Survey Geographic (SSURGO) database operated by the National Resources Conservation Service (NRCS) of the United States Department of Agriculture (USDA). The tabular dataset contains estimated and measured physical and chemical soil properties and soil interpretations. Map units were dominated by a single soil or mixture of soils.

We selected the two dominant soil profiles present within Story County. The soil profiles comprised mostly of loamy and clay loam soil of glacial till origin. Loamy soil is well-drained with high water holding capacity, making it suitable to agricultural use. Clay loam soil has lower drainage capacity and high water holding capacity due to clay texture. Soil profiles were simulated using nine soil layers (0–5, 5–15, 15–30, 30–45, 45–60, 60–90, 90–120, 120–150, 150–180 cm) to a depth of 180 cm using these two dominant soil types.

### 2.2.3. Weather data

DSSAT-CSM-Maize requires daily rainfall (mm), minimum and maximum temperature ( $^{\circ}\text{C}$ ) and solar radiation ( $\text{MJ m}^{-2} \text{ day}^{-1}$ ). We used Iowa State University AgClimate (<http://mesonet.agron.iastate.edu/agclimate/>) weather data from 2003 to 2009 (Fig. 3).

### 2.2.4. Remote sensing data

**2.2.4.1. AMSR-E soil moisture.** For this study, we used the AMSR-E soil moisture product (Njoku, Jackson, Lakshmi, Chan, & Nghiem, 2003). AMSR-E data was downloaded from the NSIDC web portal (<http://nsidc.org/data/amsre/>) and are further processed at the Jet Propulsion Lab (JPL). The crop model runs on a daily time step, and the best soil moisture estimate we can use to assimilate is the AMSR-E descending (A.M. passing) product. A fundamental assumption used in the soil moisture retrieval algorithm is the uniformity of temperature profile between the soil and vegetation stand, and the morning retrieval period is the best time to satisfy this assumption, hence more accurate retrievals. An example of the AMSR-E soil moisture product (–0–2 cm depth) over North America gridded at 25 km is shown in Fig. 4. The AMSR-E soil moisture level-3 product retrieval algorithm uses polarization ratios (PR, i.e., difference between vertical and horizontal brightness temperatures at a given frequency, divided by their sum) of the AMSR-E channel brightness temperatures at X-band (10.27 GHz) (Njoku & Chan, 2006).

A time series of AMSR-E soil moisture data over Story County (Fig. 3) shows a distinct signature typical to AMSR-E X-band measurements. Fluctuations in soil moisture values due to wetting and drying are apparent earlier in the year, but soil moisture fluctuations are dampened later in the year when vegetation cover is high. The AMSR-E X-band measurements are unable to penetrate enough through lush vegetation to detect surface soil moisture. The measurements are attenuated by high vegetation water content (VWC). Uncertainty/errors linearly interpolated from 2% for bare soil and 7.5% for VWC of  $6 \text{ kg m}^{-2}$  (Bindlish,

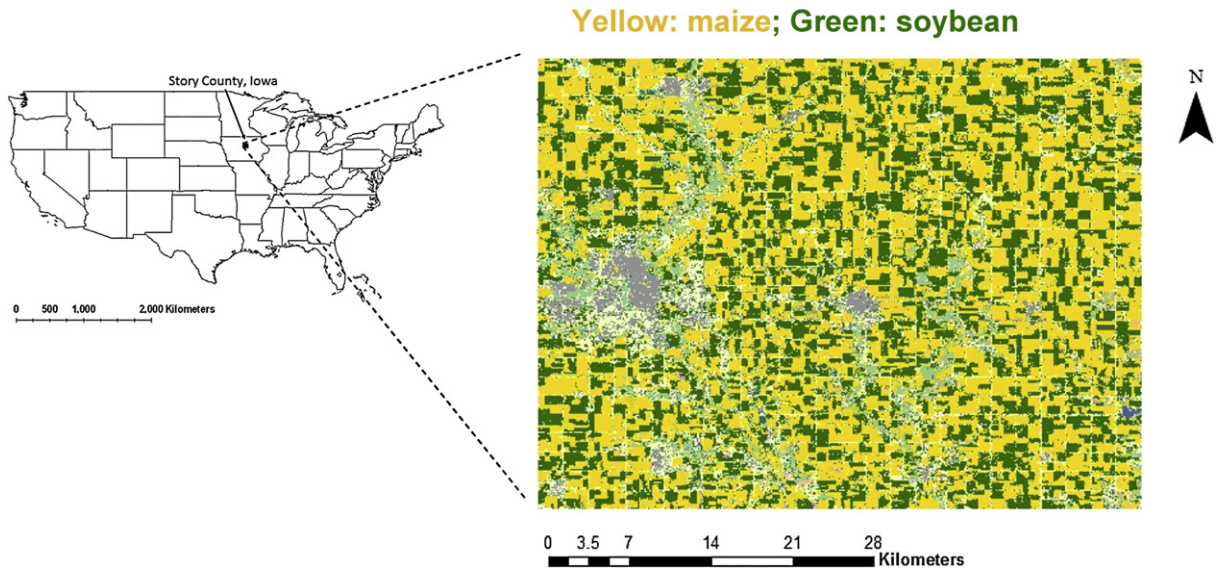


Fig. 2. Location of the study area (Story County, Iowa) and sample land use land cover map for 2005 (source: <http://nassgeodata.gmu.edu/CropScape/>).

Jackson, Gasiewski, Klein, & Njoku, 2006) were assigned to AMSR-E soil moisture estimates during this period during data assimilation.

**2.2.4.2. MODIS leaf area index (LAI).** We downloaded MOD15A2 from NASA ECHO portal (<http://reverb.echo.nasa.gov/reverb/>) and processed MODIS-LAI for Iowa. We used MODIS Reprojection Tool (MRT) to process 4 tiles of MODIS data for the state. The MODIS-LAI product is available as an 8-day composite at 1 km resolution. For testing and evaluation of the data assimilation-crop modeling framework we processed LAI from 2003 to 2009. In sampling county-wide LAI, we randomly selected maize pixels from the landuse-landcover maps of USDA-NASS crop data layer (<http://nassgeodata.gmu.edu/CropScape/>) and extracted LAI profiles from the co-located, stacked MODIS-LAI images. The extracted LAI profiles were further processed (smoothed, then averaged) to arrive at an aggregate Story County LAI time series.

#### 2.2.5. Data assimilation strategies

We used a hybrid maize variety in the crop simulations, roughly calibrated using 2005 data. The crop model was run in open-loop and the genotype coefficients and planting date were adjusted (manually) to match the growth and phenological characteristics of the crops based on USDA-NASS, remote sensing data and crop information from Iowa State University (ISU) Extension and Outreach (<http://www.extension.iastate.edu>). Sequential assimilation of remote sensing data could also correct some uncertainty associated with calibrated model parameters (Das et al., 2008).

We used DOY 130 as sowing dates (all years) for the data assimilation runs based on the examination of the MODIS-LAI profiles. This is an approximate as the planting window in Central Iowa is between April 15–May 18 (<http://www.extension.iastate.edu>). Curnel et al. (2011) noted that a phenological shift between model and observations can have a large impact thus sowing date should be chosen with care. Simulations were started at DOY 100 to allow the soil water balance model to initialize for 30 days. Fertilizer applications were fixed at 200 kg-N ha<sup>-1</sup> (based on typical practices for maize in Iowa) applied at sowing. Plant stand density was set at 7.2 plants m<sup>-2</sup>. Evapotranspiration was modeled using Priestley–Taylor method. No irrigation was applied.

The study compared an open-loop simulation and simulations that assimilated (i) LAI only, (ii) soil moisture only and (iii) soil moisture and LAI. All simulations were conducted in 2003–2009, reinitialized for each growing season. Multiple year analyses ensure that the data assimilation-crop modeling framework was subjected to actual

variability of climate. A provision was made in the data assimilation-crop modeling framework that if simulated LAI was above a threshold level (see Section 2.3.4), EnKF increases retrieval error of AMSR-E soil moisture (see Eqs. (2) and (7)).

**2.2.5.1. Monte Carlo simulations.** The success of data assimilation using EnKF depends greatly on the Monte Carlo setup. Crop model parameters that have a major influence on DSSAT-CSM-Maize model include physical attributes of soil profile (residual water content, field capacity and saturated water content), thermal time for seedling emergence, thermal time from silking to physiological maturity, maximum number of kernel per plant and phylochron interval. They were all considered fixed parameters during data assimilation. An uncertainty level of 10% was introduced to each model parameter above and perturbed using a Gaussian distribution. Then ensemble members were generated by randomly sampling model parameter combinations from the perturbed arrays. Forty ensemble members were selected to optimize the EnKF framework performance in terms of accuracy and computational time. At the start of simulation, the modified DSSAT-CSM-Maize model also randomly sampled values of leaf weight at emergence and plant leaf area at emergence for each selected ensemble member to increase the variability of the ensemble.

**2.2.5.2. Nitrogen and scaling effects.** Coupling the nitrogen (N) process is necessary in actual yield estimation. In addition to our coupled N experiments, we made an experiment when N is decoupled in the process i.e., simulations are not impacted by N stress, only by water stress, to test how the data assimilation-crop modeling framework will perform. Also, we tested scaling the AMSR-E soil moisture data to the climatology of the crop model with the aim of reducing the RS data bias (Reichle & Koster, 2004). We used a cumulative distribution function (CDF) mapping to scale the RS soil moisture data to the distribution of the modeled soil moisture (Eq. (7)), where  $x$  and  $x'$  are the non-scaled and scaled AMSR-E soil moisture and  $F(\cdot)$  is the CDF of the model or AMSR-E soil moisture; here we used normal distributions for the CDFs.

$$x' = F_{model}^{-1}[F_{AMSR-E}(x)] \quad (8)$$

Note however that if the model is also biased there is a danger of scaling RS soil moisture to the climatology of the model. If available, correcting the RS data with ground observations upscaled at the proper resolution maybe a better option (e.g., Ines & Hansen, 2006; Ines, Hansen, & Robertson, 2011). Better soil moisture measurements from

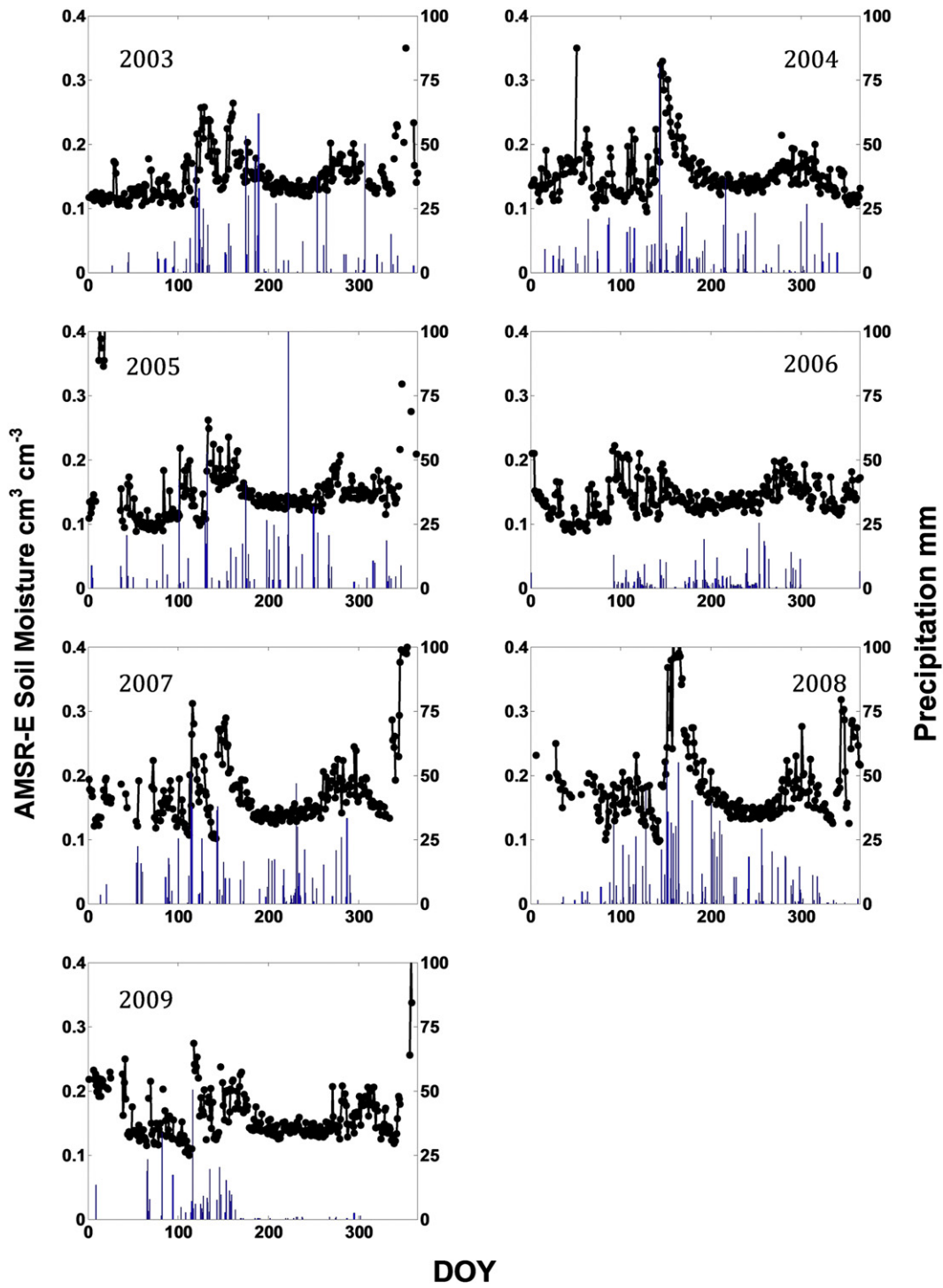


Fig. 3. AMSR-E soil moisture (line) and station rainfall (bar) at Story Co., Iowa, 2003–2009.

new generation satellites e.g., SMAP, may reduce this problem in the future.

2.2.5.3. Analysis. The performance of the data assimilation-crop modeling framework was evaluated using county level maize yields recorded by USDA National Agricultural Statistics Service (NASS) for Story County, Iowa for the year 2003–2009. Results of the three data assimilation modeling strategies and open-loop runs were compared. Standard goodness-of-fit statistics such as Pearson's correlation (R, unitless)

(Eq. (9)), Mean Bias Error (MBE, Mg ha<sup>-1</sup>) (Eq. (10)) and Root Mean Squared Error (RMSE, Mg ha<sup>-1</sup>) (Eq. (11)) were used to measure performance,

$$R = \frac{\left( M \sum_{i=1}^M \bar{y}_i y_{obs,i} - \sum_{i=1}^M \bar{y}_i \sum_{i=1}^M y_{obs,i} \right)}{\sqrt{M \left( \sum_{i=1}^M \bar{y}^2 - \left( \sum_{i=1}^M \bar{y}_i \right)^2 \right) \left( M \sum_{i=1}^M y_{obs,i}^2 - \left( \sum_{i=1}^M y_{obs,i} \right)^2 \right)}} \quad (9)$$



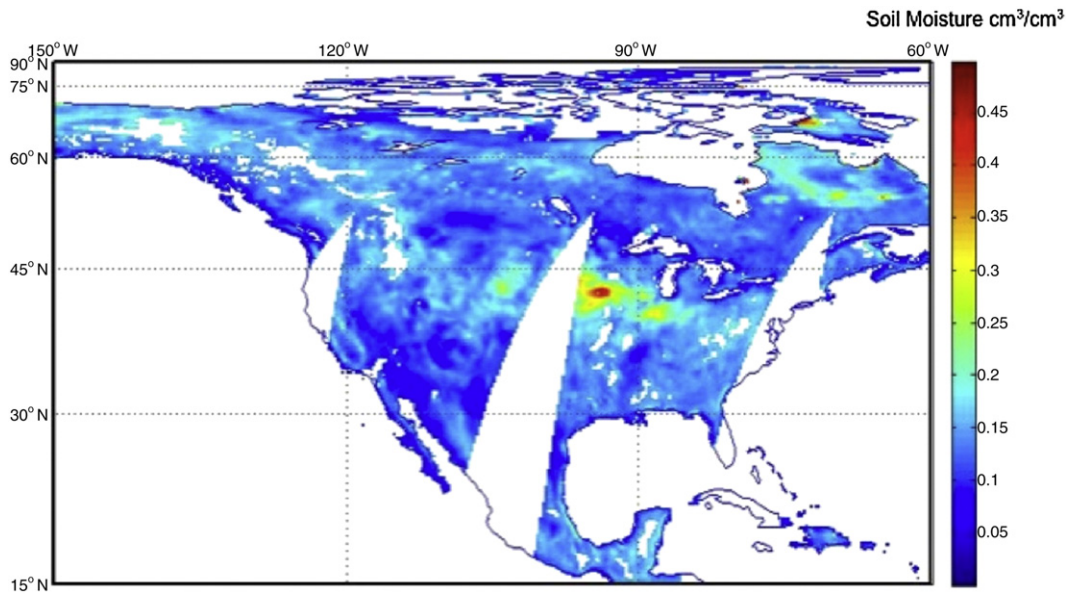


Fig. 4. Snapshot of AMSR-E soil moisture, June 6, 2008.

$$MBE = M^{-1} \sum_{i=1}^M (\bar{y} - y_{obs,i}) \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^M (\bar{y}_i - y_{obs,i})^2}{M-1}} \quad (11)$$

where  $\bar{y}_t = N_e^{-1} \sum_{n=1}^{N_e} y_{n,i}$  is the average yield ( $Mg\ ha^{-1}$ ) of the ensemble run,  $i$  is an index of year,  $M$  is the number of years,  $n$  is an index of ensemble member,  $N_e$  is the size of the ensemble,  $y_{obs,i}$  is the county level observed yield for year  $i$  and  $y_{n,i}$  is an ensemble member yield for year  $i$ .

### 3. Results and discussion

#### 3.1. Open-loop simulations

The crop model without data assimilation was able to capture a substantial part of the year-to-year variability of county-level yields ( $R = 0.47$ , Fig. 5a, Table 1). However, simulations generally under-predicted observed yields, leading to high MBE ( $-3.7\ Mg\ ha^{-1}$ ) and RMSE ( $4.7\ Mg\ ha^{-1}$ ) (Table 1; actual errors could be lower as the model predicted yields in dry weight). The yield mismatch was particularly pronounced in 2006 and 2009. Based on the weather station data, these years received the lowest rainfall (Fig. 3) but records show that yields are high in the county. We hypothesize that this performance of the open-loop simulation could be attributed to several factors, which may include uncertainties in forcing data e.g., rainfall at the station (due to scale effects) and modeling assumptions, including model parameters. Assuming that the USDA-NASS yield data is accurate, which may be valid in the US, the open-loop simulation results suggest that the rainfall measured at the weather station in 2006 and 2009 may not have represented well what occurred in the county (scale effects) as the recorded yields of maize (fully rainfed) during those years are above average. A quick verification with CMORPH rainfall data supported this hypothesis, but one should note that there are also uncertainties associated with this data (<http://iridl.ldeo.columbia.edu/expert/SOURCES/NOAA/NCEP/CPC/CMORPH>). However, it is also possible that the crops' deeper root system may have allowed the crops to access deeper soil moisture during those years, hence the higher yields. During

2008, which is a very wet year, the slightly lower county yields could be attributed to the lower temperature and solar radiation, associated with very wet conditions, and water logging or leaching of nutrients due to runoff or deep percolation. The correlation of simulated yields and observations with open-loop simulation suggests that the weather station

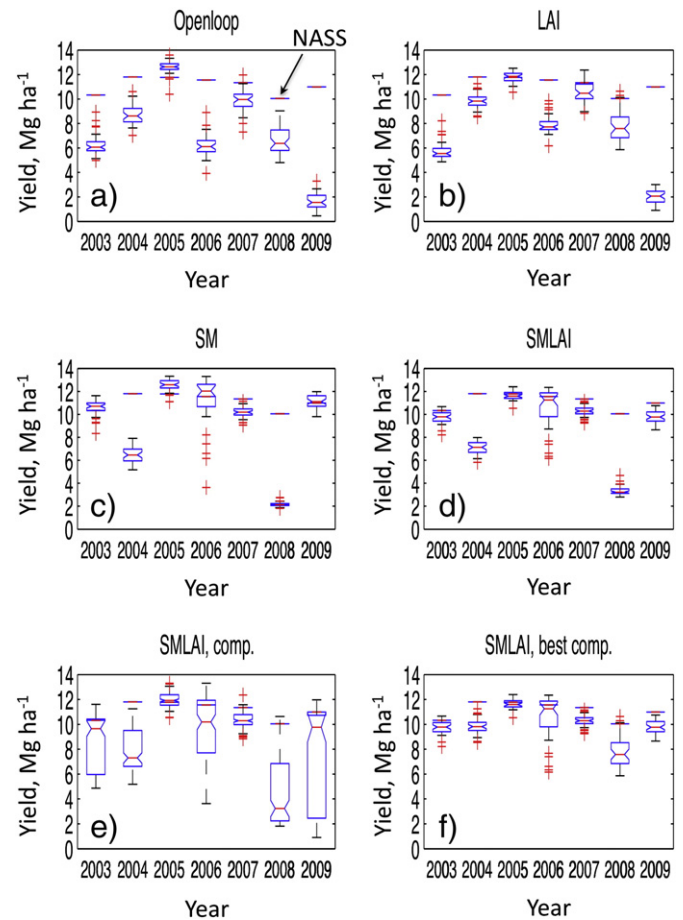


Fig. 5. Performance of open-loop simulation (a), assimilation of MODIS-LAI (b), assimilation of AMSR-E soil moisture (c), assimilation of AMSR-E soil moisture + MODIS-LAI (d), composite all (e), composite best (f).

**Table 1**  
Performance (average) of the EnKF data assimilation system for simulating maize yields, Story County, Iowa.

Simulation strategies	R	MBE, Mg ha <sup>-1</sup>	RMSE, Mg ha <sup>-1</sup>
<i>No scaling</i>			
Open-loop	0.47	-3.7	4.7
DA with LAI	0.51	-3.2	4.2
DA with Soil moisture	0.50	-1.9	3.6
DA with Soil moisture + LAI	0.65	-2.0	2.9
Composite all	0.73	-2.4	3.0
Composite best	0.80	-1.2	1.4
<i>Nitrogen decoupling, and scaling</i>			
N-process decoupled <sup>a</sup>	-0.43	2.8	3.7
RS soil moisture scaled to model climatology <sup>b</sup>	0.61	-3.0	3.4

R – Pearson's correlation; MBE – Mean Bias Error; RMSE – Root Mean Squared Error.

<sup>a</sup> Data assimilation with Soil moisture + LAI with non-limited N (i.e., N-process is OFF).

<sup>b</sup> Data assimilation with Soil moisture + LAI with AMSR-E scaling, N-process is ON.

used in the simulations had captured some of the inter-annual variability of rainfall in the county.

### 3.2. Data assimilation of LAI, soil moisture and combination

#### 3.2.1. Assimilating LAI

Assimilating MODIS-LAI showed a slight improvement in simulated yields compared with the open-loop simulation ( $R = 0.51$ ), including a small reduction of systematic error (Table 1, Fig. 5b). Assimilation improved simulated yields in some years (2004, 2005, 2006, 2007 and 2008), but did not show improvement in 2009. For brevity, we present detailed analyses only for the contrasting years, 2006, 2009 and 2008. Based on weather station records, 2006, 2008, and 2009 were moderately dry, extremely wet, and very dry, respectively (Fig. 3).

Fig. 6 shows the vegetation, soil moisture dynamics (near-surface), water and nitrogen stress in 2006, with open-loop simulation (a–d) and data assimilation of LAI (e–f). Enhanced canopy growth resulting from the assimilation of remotely sensed LAI is associated with the improved yield simulation that year. In both cases, the crops were subjected to high intermittent water stress due to low rainfall (Fig. 6c, g), but with no apparent nitrogen (N) stress (Fig. 6d, h). The small shifts in simulated near-surface soil moisture (Fig. 6b, f, from DOY 200 hence) with LAI assimilation show that the enhanced canopy also altered the simulated water balance, especially for evapotranspiration and rootzone soil moisture (not shown, respectively).

In 2009, rootzone soil moisture was very low (data not shown) and the crop was subjected to extreme levels of prolonged water stress due to lack of rain after anthesis as recorded by the weather station. This led to drastic reductions in LAI under the open-loop simulation (Fig. 7a, c). Instead of improving simulated yield, assimilation of remotely sensed LAI increased water stress and reduced simulated yield because the limited rootzone soil moisture could not satisfy the increase in water demand that resulted from increased LAI (Fig. 7e, g). This increase in canopy growth from assimilation of LAI data also increased N stress (Fig. 7d, h) because of non-availability of labile N due to limited soil moisture and increased N demand by the crops.

The improvements in simulated yields with the assimilation of LAI in 2008 were due to enhanced canopy growth (Fig. 8a, e) combined with abundant soil moisture availability due to being a wet year (Fig. 8b, f). Because of the extremely wet condition, the simulated LAI by open-loop simulation were robust but more inferior than when there is data assimilation. With LAI assimilation, water stress was negligible until the later part of the growing season (Fig. 8c, g). Nitrogen stress is apparent due to higher N leaching and N loss associated with increased percolation and runoff (Fig. 8d, h).

#### 3.2.2. Assimilating soil moisture

When assimilating RS soil moisture alone, we did not constrain vegetation dynamics. Assimilating AMSR-E soil moisture improved

simulated yields compared with open-loop simulations, except for 2004 and 2008 (Table 1, Fig. 5c).

The median of the yield ensemble improved substantially when soil moisture was assimilated in 2006 (Fig. 5c). The assimilation of soil moisture led in an overall reduction in water stress, improvement in canopy growth (Fig. 9a, c) and increased near-surface soil moisture (Fig. 9b). The durations and intensities of water stress experienced by the crops in open-loop simulation were minimized after soil moisture assimilation. de Wit and Van Diepen (2007) observed that RS soil moisture when assimilated with a crop model could correct some scale issues associated to rainfall. But note that the AMSR-E signal is known to be attenuated by high vegetation moisture content (canopy fully developed) causing bias in measurements (Fig. 9b). Because of the bias in AMSR-E soil moisture, the assimilation of near-surface soil moisture to update soil moisture in the rootzone increased cumulative percolation by roughly 4 orders of magnitude, which is linked to N stress at the later part of the growing season (Fig. 9d). The bias in AMSR-E could be corrected by upscaled observations from the ground. When in service in 2014, the SMAP mission may reduce the severity of soil moisture signal saturation due to high vegetation volume during the growing season (Das, Entekhabi, & Njoku, 2011; Entekhabi et al., 2010).

The case of 2009 suggests the potential use of remotely sensed soil moisture to correct shortcomings in station weather data as inputs to simulate aggregate yield due to scale issues or instrument failure (Fig. 5c) as also mentioned by Bolten and Crow (2012) in their study. When soil moisture was assimilated, most of the water stress experienced by the crops during open-loop simulation and LAI data assimilation were minimized, resulting in enhanced canopy growth and yield (Fig. 10a–c, Fig. 5c). But as observed in 2006, the assimilation of AMSR-E soil moisture also increased total percolation compared with open-loop simulation, which is linked to N stress at the later part of the growing season (Fig. 10d).

The case of 2008 (Fig. 11) shows another aspect of soil moisture assimilation especially with a biased AMSR-E soil moisture data. This bias appears to be a disadvantage in very wet conditions. While assimilation of AMSR-E works well in nominal conditions, when condition is extremely wet, the performance can severely retrogressed (Fig. 5c). Due to excess water entering into the soil profile as a result of rootzone soil moisture updating and excessive rainfall, N losses appear to be high due to excessive leaching as well as by surface runoff, resulting in extreme (duration and intensity) N stress in the growing season (Fig. 11c, d). Even though water stress is completely eliminated after soil moisture assimilation (Fig. 11b, c), the crops in the simulations failed due to extreme N stress (Fig. 11a, d). In DSSAT-CSM, the law of the minimum is used to account for the effects of N and water stress to daily plant growth (i.e.,  $\text{Min}[(1-\text{WaterStress}), (1-\text{NStress})]$ ) thus, if N stress dominates and depending on timing (vegetation and anthesis growth stages are critical), crops are adversely impacted resulting to lower yields or worse, crop failure. This is also what happened in 2004



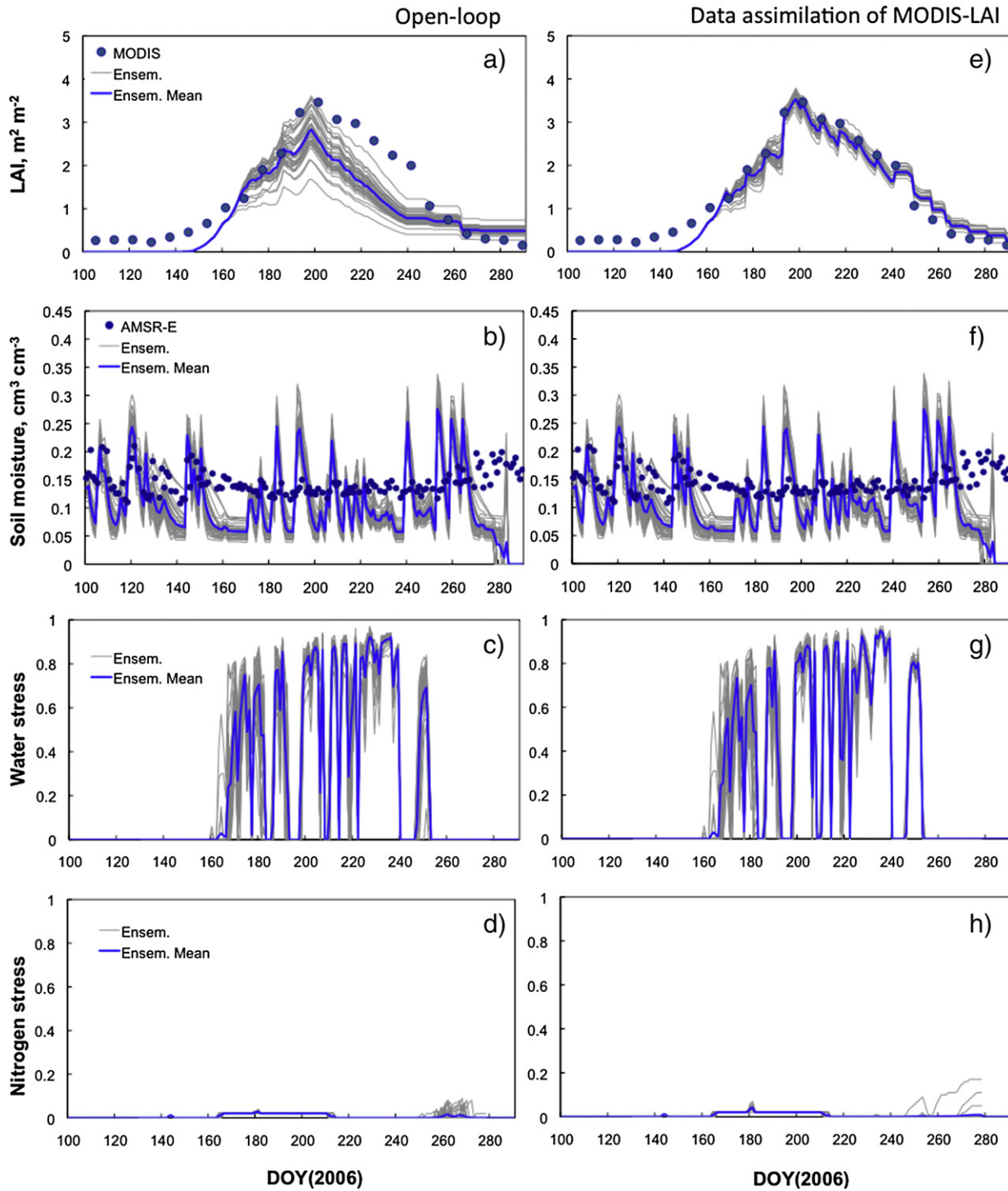


Fig. 6. LAI, soil moisture, water and nitrogen stress (unitless) simulations with (a–d) open-loop and (e–h) data assimilation of MODIS-LAI in 2006.

simulations but not as severe as 2008. It appears that when a system is not soil moisture-controlled, assimilating only LAI may give better results (Fig. 5b).

### 3.2.3. Assimilating soil moisture and LAI

The performance of the data assimilation-crop modeling framework showed further improvement when both soil moisture and LAI were assimilated ( $R = 0.65$ ,  $MBE = -2.0 \text{ Mg ha}^{-1}$ ,  $RMSE = 3.9 \text{ Mg ha}^{-1}$ , Table 1; Fig. 5d), suggesting an interaction between the two.

Assimilating both LAI and soil moisture (Fig. 9e–h) tightened the ensemble yield distribution in 2006 (Fig. 5c,d), compared to soil moisture assimilation alone. The resulting median was closer to the recorded county yield. Assimilating soil moisture reduced simulated crop water stress, while the LAI assimilation corrected the LAI profiles, simultaneously (Fig. 9e,f).

Assimilating the combination of AMSR-E soil moisture and MODIS-LAI increased the cumulative percolation (results not shown), which led to N stress in the later part of the growing season (Fig. 9h). It had been observed that assimilating AMSR-E soil

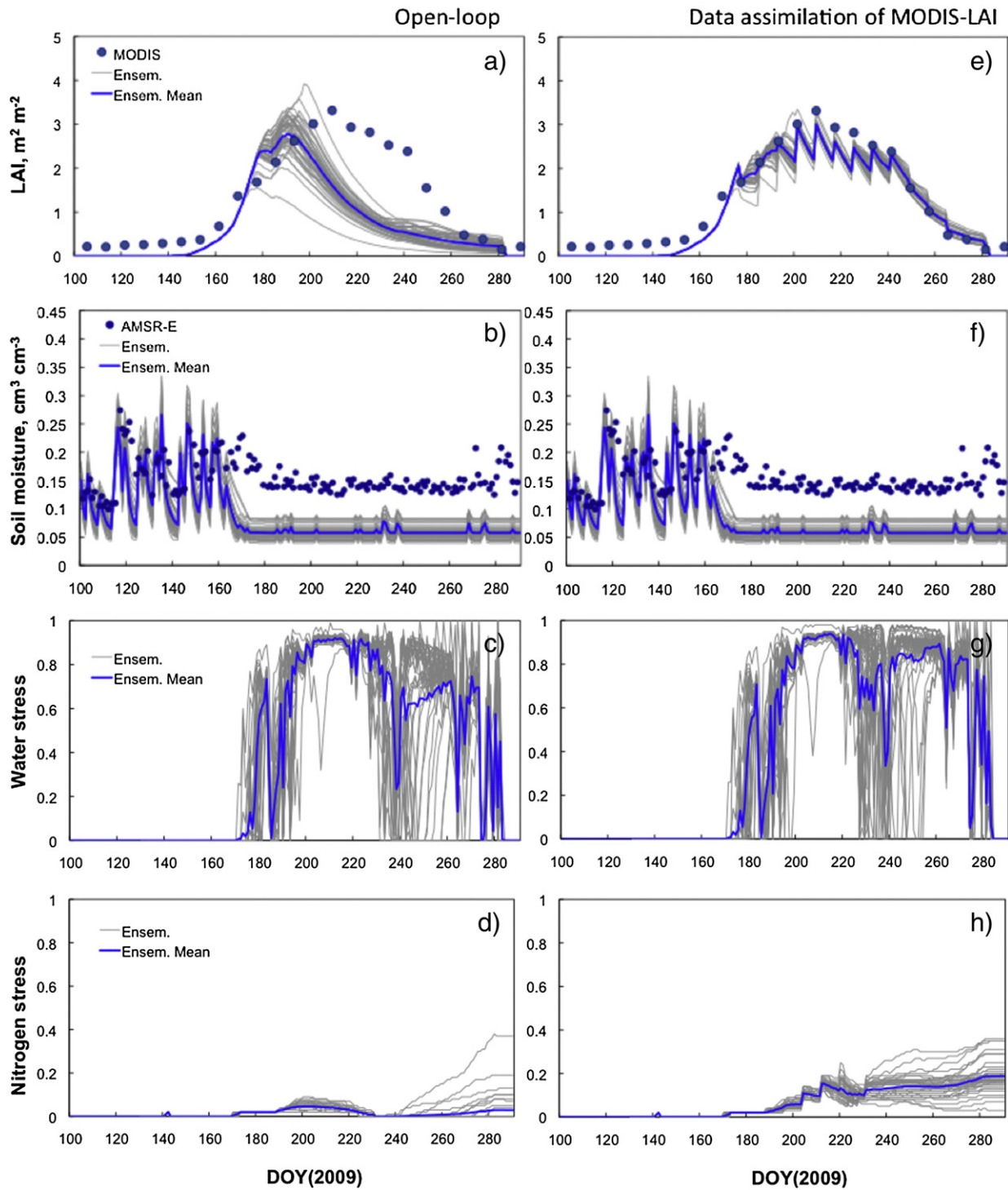


Fig. 7. LAI, soil moisture, water and nitrogen stress (unitless) simulations with (a–d) open-loop and (e–h) data assimilation of MODIS-LAI in 2009.

moisture alone increased percolation and N stress (Section 3.2.2.). However, the effect of data assimilation on N stress was not as extreme/severe when both LAI and soil moisture were assimilated (Fig. 9d,h). This was due to the more members in the ensemble having above-average LAI values when assimilating soil moisture alone (Fig. 9a). We observed the same in the case of 2009 (Fig. 10a–h) but the simulated yield distribution shifted down when LAI and soil moisture were assimilated (Fig. 5d).

Simultaneous assimilation of soil moisture and LAI improved simulated yields slightly in 2004 and 2008 (Fig. 5c,d). However, assimilating

LAI did not compensate for the over-prediction of leaching, N stress and yield reduction that apparently resulted from assimilating soil moisture data in these relatively wet years (Fig. 11e–h; see Section 3.2.2). We believe this could be an artifact of the bias in AMSR-E soil moisture data, with effects more pronounced under extremely wet conditions.

### 3.2.4. Composite analysis

With the current limitations of data and our data assimilation system, a strategic application of the data assimilation-crop modeling framework appears to be desired to further improve performance

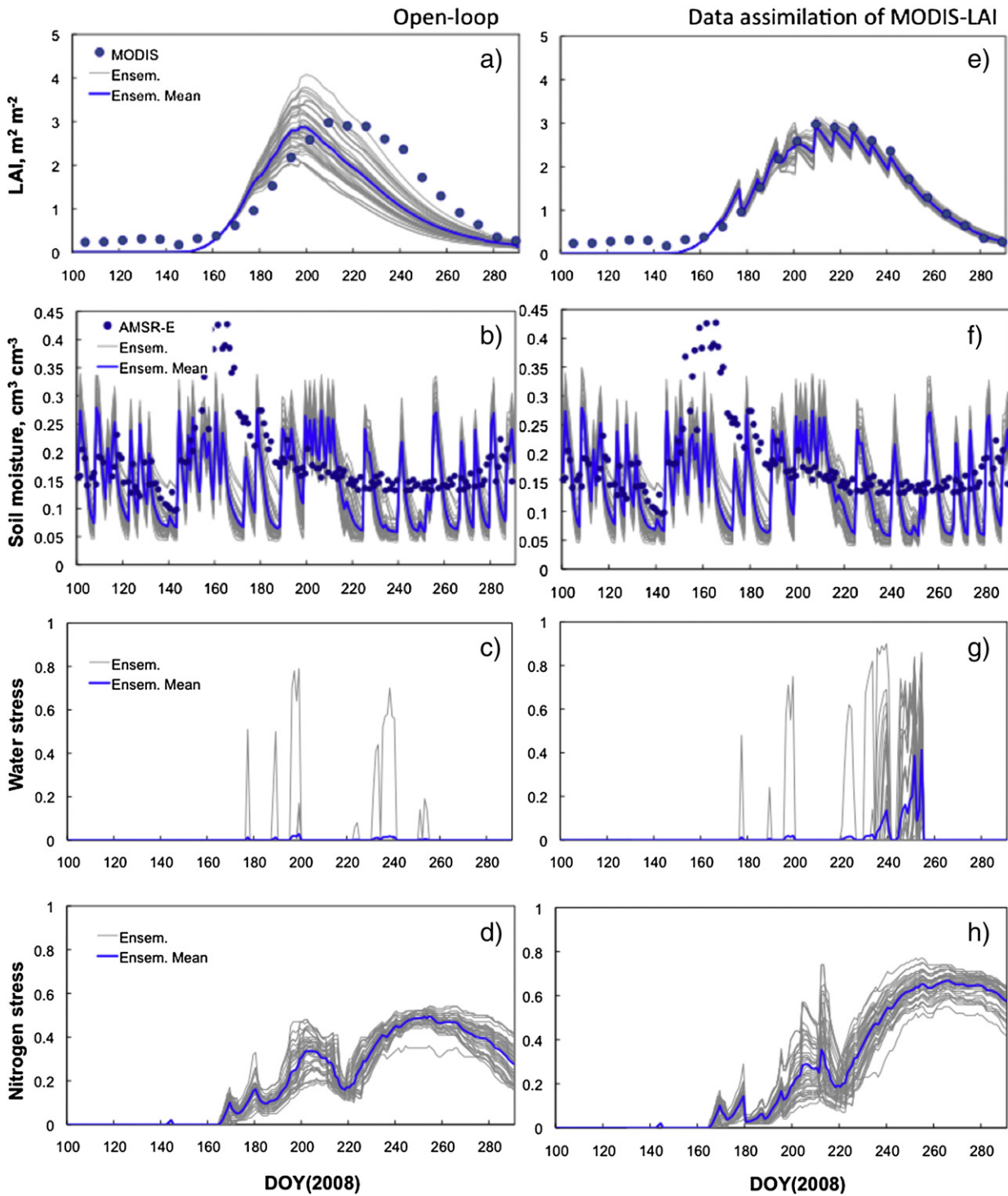
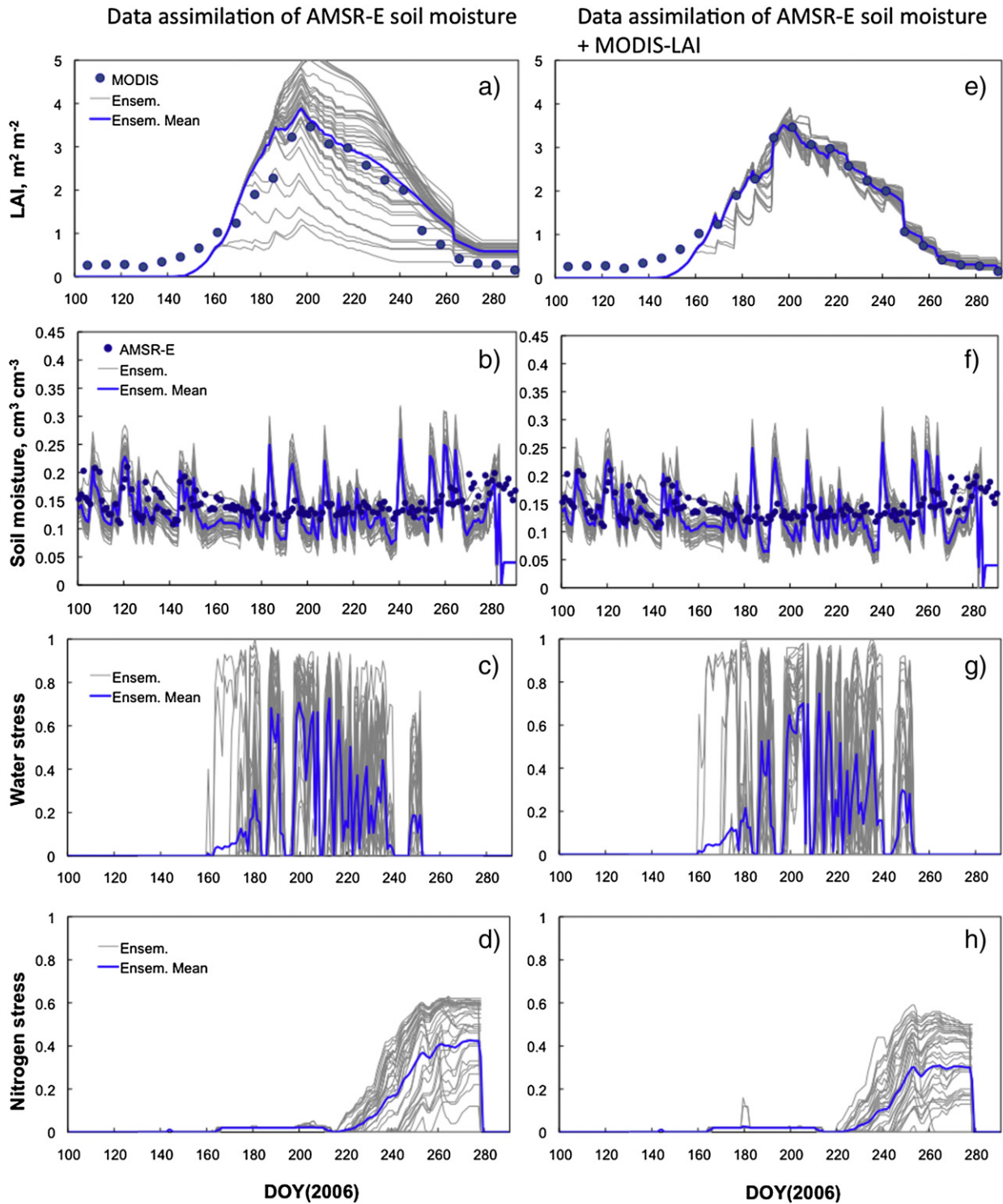


Fig. 8. LAI, soil moisture, water and nitrogen stress (unitless) simulations with (a–d) open-loop and (e–h) data assimilation of MODIS-LAI in 2008.

(i.e., one should know when to best use LAI alone or LAI plus soil moisture in data assimilation). Our analysis showed that climate condition expected during the growing season could provide information as to when a variable is best to be assimilated. As a base scenario, we exploited all sources of RS information for predicting crop yields. In this case, if we run data assimilation with soil moisture, LAI, and LAI + soil moisture then aggregate all their results, we get a yield performance of  $R = 0.73$ ,  $MBE = -2.4 \text{ Mg ha}^{-1}$  and  $RMSE = 3.0 \text{ Mg ha}^{-1}$  (composite all, Table 1), but the spread of prediction is wider (Fig. 5e). In our earlier analysis, we found

that when the system is not soil moisture-controlled (extremely wet conditions), assimilating LAI alone might be more ideal because the plants will not be experiencing extreme water stress anyway, and when conditions are nominal, assimilating soil moisture and LAI might be better. This is to avoid the scenario that the data assimilation-crop modeling framework will deplete Nitrogen from the soil due to very high percolation. Using this information, we applied the data assimilation-crop modeling framework to predict crop yields and the performance substantially improved with  $R = 0.80$ ,  $MBE = -1.2 \text{ Mg ha}^{-1}$  and  $RMSE = 1.4 \text{ Mg ha}^{-1}$  (Fig. 5f, Table 1;





**Fig. 9.** LAI, soil moisture, water and nitrogen stress (unitless) simulations with (a–d) data assimilation of AMSR-E soil moisture and (e–h) data assimilation of AMSR-E soil moisture and MODIS-LAI in 2006.

composite best). The challenge in applying this rule of thumb would be the availability of skillful seasonal climate forecasts. Better soil moisture products e.g., SMAP, would be still a better alternative to fully utilize this kind of methodology. Some components of the data assimilation-crop modeling could be improved as well e.g., a variable inflation parameter.

### 3.2.5. Nitrogen and scaling effects

Accounting for the N process is important in simulating actual yields even when fertilizer application is at a very high level because of the

interaction between water and N in crop growth. Fig. 12a shows the tendency of over-predicting crop yields when N is non-limiting in the data assimilation; the overall performance is poor (Table 1).

In our case, scaling RS soil moisture with model climatology did not work well because our model was also biased in simulating county scale soil–water–plant–atmosphere processes (Fig. 12b) as shown by open-loop simulation results (Fig. 5a). Yield correlation was preserved as when there is no scaling performed, but the prediction errors were very high compared to no scaling (Table 1). However, the potential

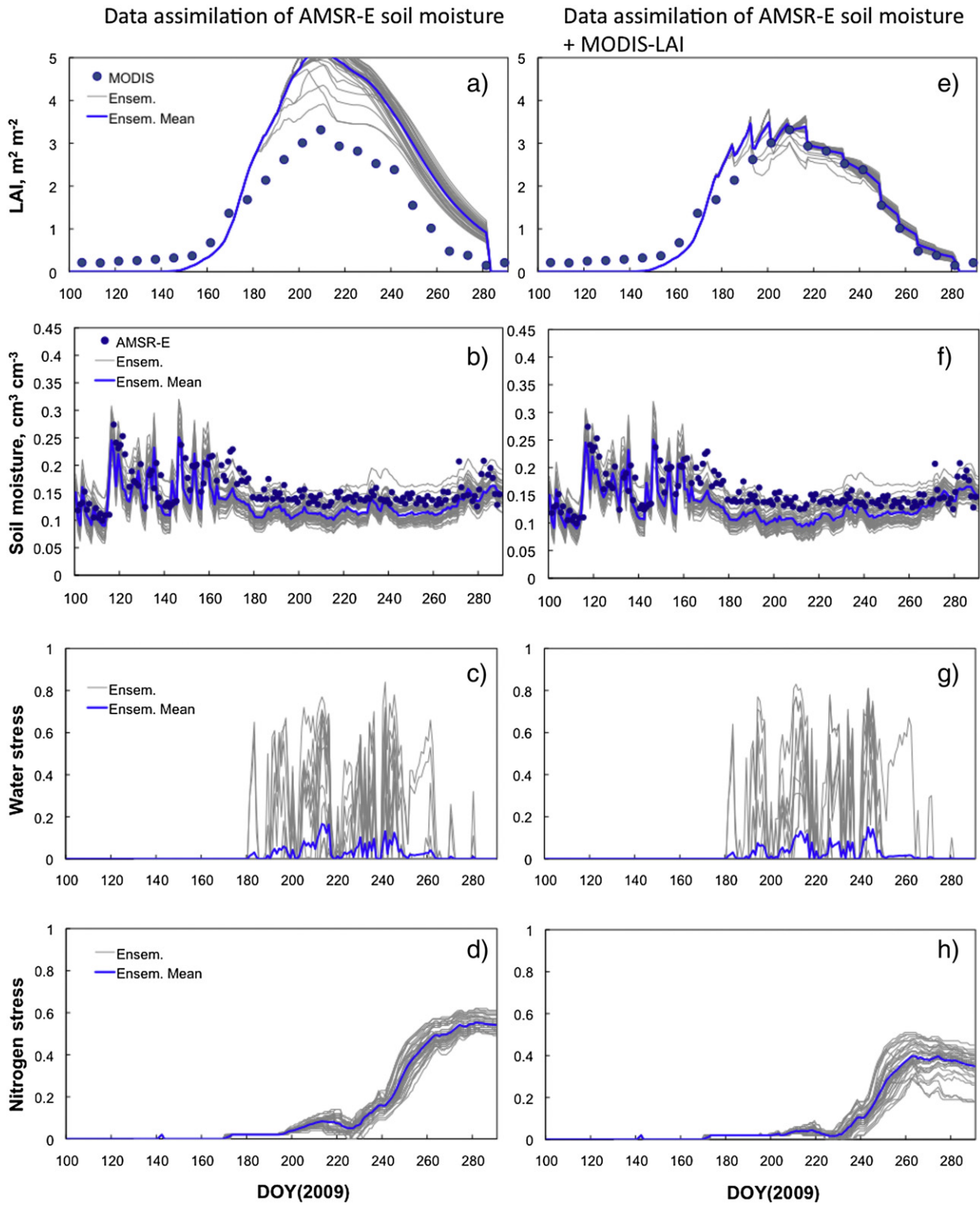


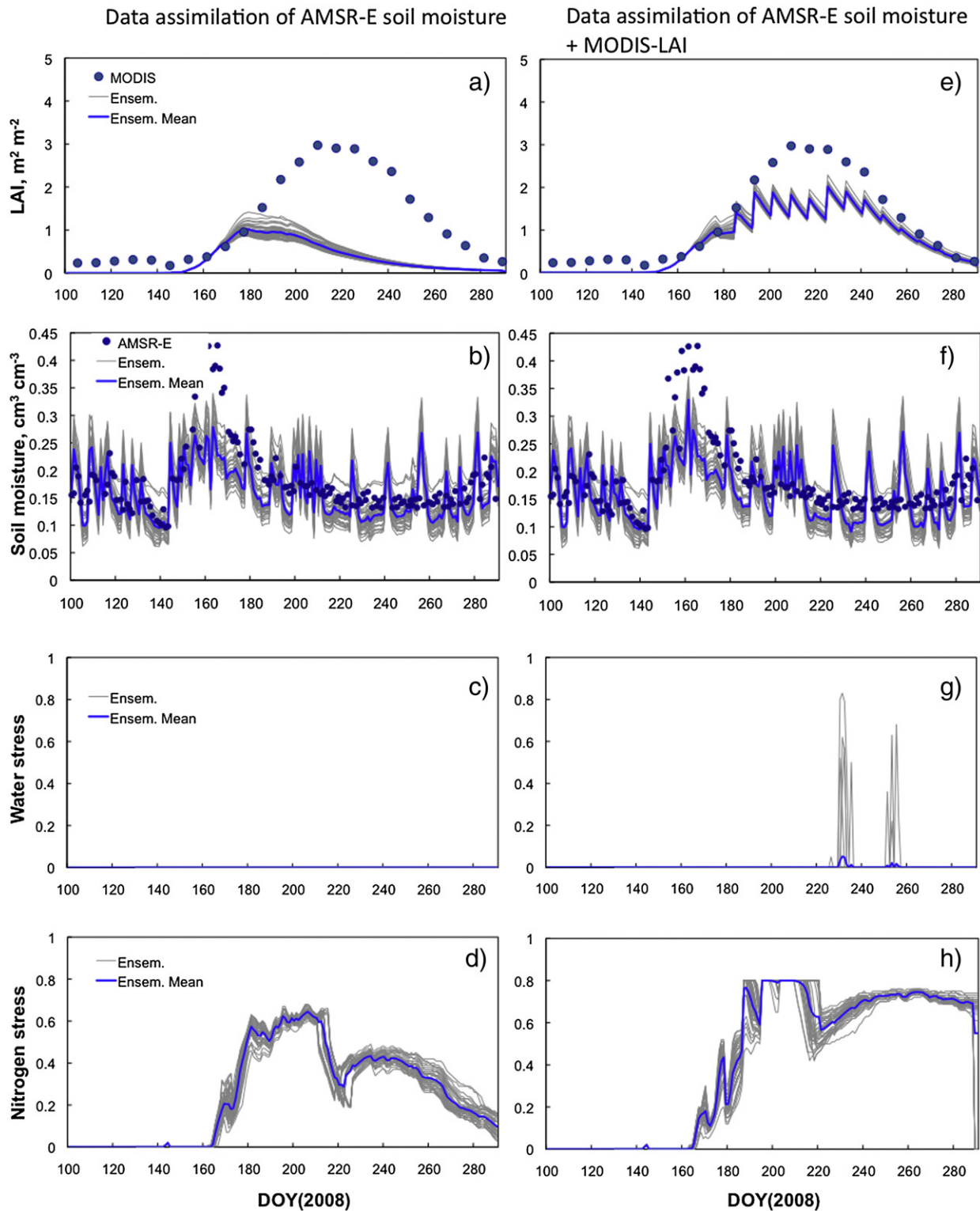
Fig. 10. LAI, soil moisture, water and nitrogen stress (unitless) simulations with (a–d) data assimilation of AMSR-E soil moisture and (e–f) data assimilation of AMSR-E soil moisture and MODIS-LAI in 2009.

use of better soil moisture products is evident in 2008 results (Fig. 12b, Fig. 5d).

#### 4. Summary and conclusions

We developed and tested a data assimilation-crop modeling framework for assimilating RS soil moisture and LAI to predict crop yields at aggregate scale. We found that DSSAT-CSM-Maize run

as a Monte Carlo ensemble without data assimilation was able to capture part of the inter-annual variability of maize yields in Story County, but with a strong negative mean bias (better model calibration could have improved this further). However, there were two interesting years (2006, 2009) that shed light about the risk of using only point-scale weather data (station) as inputs for pseudo-regional crop modeling application to simulate aggregate yield.



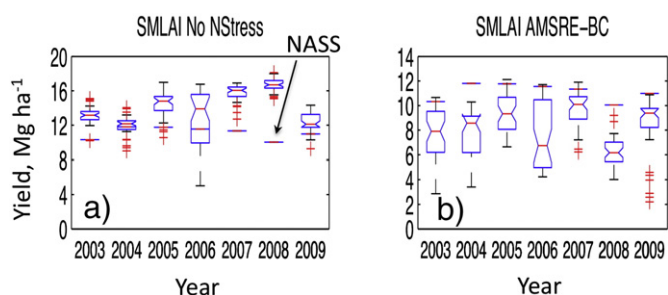
**Fig. 11.** LAI, soil moisture, water and nitrogen stress (unitless) simulations with (a–d) data assimilation of AMSR-E soil moisture and (e–h) data assimilation of AMSR-E soil moisture and MODIS-LAI in 2008.

When MODIS-LAI was assimilated, we observed some improvements in simulated yields, including in 2006. However, under very dry conditions, as in 2009, LAI assimilation could not improve simulated yields because the rootzone soil moisture could not meet the increased water demand that result from improved canopy growth.

We also observed an incremental improvement in simulated yields when AMSR-E soil moisture was assimilated, especially for the cases

of 2006 and 2009, suggesting the potential use of remotely sensed soil moisture data to correct the shortcomings of point-scale weather data as input to simulating aggregate yields, in addition to correcting model errors. However, it was observed that assimilating near-surface soil moisture to update the rootzone wetness had caused excessive drainage (vertical) that when conditions are extremely wet can result in the underestimation of yields, this is due to excessive N leaching/





**Fig. 12.** Performance of assimilation of AMSR-E soil moisture + MODIS-LAI (a) when Nitrogen is not limiting (Nitrogen module is OFF), and (b) AMSR-E soil moisture is scaled to the climatology of the model, Nitrogen module is ON. The flat boxplot shows Story county (IA) yield data from USDA-NASS from 2003 to 2009.

loss. We believe that this is an artifact of the biased AMSR-E soil moisture data, which impact is more pronounced under extremely wet climate condition. Unfortunately, when the AMSR-E signal becomes insensitive due to dense vegetation it is the time when accurate soil moisture measurements are needed for data assimilation.

There were more improvements in simulated yields when soil moisture + LAI were assimilated, but because of the current limitation of the data assimilation system, a composite analysis was tested to decide when LAI or SM + LAI assimilation is best to be done. We found that applying the data assimilation-crop modeling framework strategically using a-priori information of climate expected during the growing season could improve overall the simulated yield performance. Under very wet condition, SM assimilation may not be necessary and that LAI assimilation may be more useful.

Applying the commonly used scaling of RS soil moisture data to model climatology did not improve our data assimilation results because our model is also biased (as shown in open-loop) in simulating county-scale soil–water–plant–atmosphere processes.

The AMSR-E instrument aboard the Aqua satellite malfunctioned in October 2011 and is no longer in service. NASA's upcoming Soil Moisture Active Passive (SMAP) mission shall be a new source of soil moisture estimates. The L-band radar and L-band radiometer measurements from the SMAP mission will help alleviate the problem of soil moisture detection under dense vegetation conditions, and the penetration depth would be ~0–5 cm. The SMAP radar and radiometer are designed to remotely sense soil moisture with high accuracy ( $RMSE \leq 0.04 \text{ cm cm}^{-3}$ ) under vegetation with water content up to  $5.0 \text{ kg m}^{-2}$  (Entekhabi et al., 2010). This will ensure that the SMAP measurements are effective throughout most of the growing season for many crops, hence the availability of better soil moisture data during the critical period of crop growth. For future work, we plan to use the SMAP radiometer-radar combined 9 km soil moisture product (Das et al., 2011) or the SMAP radar-only 3 km soil moisture product to test the data assimilation-crop modeling framework presented here. Also, the Global Precipitation Measurement (GPM) rainfall product will be of great value for aggregate scale modeling of crop yields.

## Acknowledgement

This work was done with the support of the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) with the financial assistance of the European Union, the United States Agency for International Development (USAID), Canadian International Development Agency (CIDA), New Zealand Ministry of Foreign Affairs and Trade, the Danish International Development Agency (DANIDA), the UK Department for International Development (DfID), Irish Aid, and Instituto de Investigação Científica Tropical, Portugal (IICT) with technical support from IFAD. The views expressed in this document cannot be taken to reflect the official opinions of these agencies, nor the official position of the CGIAR.

We also acknowledge NOAA cooperative agreement #NA05OAR4311004 and Steve Zebiak for the conduct of this study. We thank Gerrit Hoogenboom for guidance in DSSAT-CSM code. Michael Norton and Yufang Zhu are also acknowledged for their help preparing remote sensing and other ancillary data. We acknowledge Susan Moran of SMAP Applications – some components of this study are contributions by IRI/CCAFS to NASA's SMAP Early Adopter. We thank the editors and reviewers for helping us improve the paper.

## References

- Bindlish, R., Jackson, T. J., Gasiewski, A. J., Klein, M., & Njoku, E. G. (2006). Soil moisture mapping and AMSR-E validation using the PSR in SMEX02. *Remote Sensing of Environment*, 103, 127–139.
- Bolten, J. D., & Crow, W. T. (2012). Improved prediction of quasi-global vegetation conditions using remotely-sensed surface soil moisture. *Geophysical Research Letters*, 39, L19406, <http://dx.doi.org/10.1029/2012GL053470>.
- Bouman, B. A. M., (1992). Linking physical remote sensing models with crop growth simulation models, applied for sugar beet. *International Journal of Remote Sensing*, 13, 2565–2581.
- Bouman, B. A. M., Van Diepen, C. A., Vossen, P., & Van Der Val, T. (1997). Simulation and systems analysis tools for crop yield forecasting. In P.S. Teng (Ed.), *Applications of Systems Approaches at the Farm and Regional Levels* (pp. 325–340). Kluwer Academic Publishers.
- Burgers, G., Van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Bulletin*, 126, 1719–1724.
- Crow, W. T., & Wood, E. F. (2003). The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97. *Advances in Water Resources*, 26, 137–149.
- Curnel, Y., de Wit, A. J. W., Duveiller, G., & Defourmy, P. (2011). Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agricultural and Forest Meteorology*, 151, 1843–1855.
- Das, N. N., Entekhabi, D., & Njoku, E. G. (2011). An algorithm for merging SMAP radiometer and radar data for high-resolution soil moisture retrieval. *IEEE Transaction of Geosciences and Remote Sensing*, 49, 1504–1512.
- Das, N. N., & Mohanty, B. P. (2006). Root zone soil moisture assessment using passive microwave remote sensing and vadose zone modeling. *Vadose Zone Journal*, 5, 296–307.
- Das, N. N., Mohanty, B. P., Cosh, M. H., & Jackson, T. J. (2008). Modeling and assimilation of root zone soil moisture using remote sensing observations in Walnut Gulch watershed during SMEX04. *Remote Sensing of Environment*, 112, 415–429.
- de Wit, A. J. W., & Van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146, 38–56.
- Dunne, S., & Entekhabi, D. (2005). An ensemble-based reanalysis approach to land data assimilation. *Water Resources Research*, 41, <http://dx.doi.org/10.1029/2004>.
- Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al. (2010). The Soil Moisture Active Passive (SMAP) mission. *Proceedings of the IEEE*, 98, 107–116.
- Evensen, G. (2003). The Ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 343–367.
- Fang, H., Liang, S., & Hoogenboom, G. (2011). Integration of MODIS products and a crop simulation model for crop yield estimation. *International Journal of Remote Sensing*, 32, 1039–1065.
- Fang, H., Liang, S., Hoogenboom, G., Teasdale, J., & Cavigelli, M. (2008). Crop yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. *International Journal of Remote Sensing*, 29, 3011–3032.
- Hansen, J. W., Challinor, A., Ines, A., Wheeler, T., & Moron, V. (2006). Translating climate forecasts into agricultural terms: advances and challenges. *Climate Research*, 33, 27–41.
- Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126, 796–811.
- Ines, A. V. M., & Hansen, J. W. (2006). Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology*, 138, 44–53.
- Ines, A. V. M., Hansen, J. W., & Robertson, A. W. (2011). Enhancing the utility of daily GCM rainfall for crop yield prediction. *International Journal of Climatology*, 31, 2168–2182.
- Ines, A. V. M., Honda, K., Gupta, A.D., Droogers, P., & Clemente, R. S. (2006). Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture. *Agricultural Water Management*, 83, 221–232.
- Ines, A. V. M., & Mohanty, B. P. (2008a). Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties under different hydroclimatic conditions. *Vadose Zone Journal*, 7, 39–52.
- Ines, A. V. M., & Mohanty, B. P. (2008b). Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithm: 1. Conceptual modeling. *Water Resources Research*, 44, W06422, <http://dx.doi.org/10.1029/2007WR005990>.
- Ines, A. V. M., & Mohanty, B. P. (2008c). Parameter conditioning with a noisy Monte Carlo genetic algorithm to estimate effective soil hydraulic properties from space. *Water Resources Research*, 44, W08441, <http://dx.doi.org/10.1029/2007WR006125>.
- Ines, A. V. M., & Mohanty, B. P. (2009). Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithms: 2. Using airborne remote sensing during SGP97 and SMEX02. *Water Resources Research*, 45, W01408, <http://dx.doi.org/10.1029/2008WR007022>.

- Jones, J. W., Hoogenboom, G., Porter, C., Boote, K. J., Batchelor, W. D., Hunt, L. A., et al. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235–265.
- Keppenne, C. L., & Rienecker, M. M. (2002). Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Monthly Weather Review*, 130, 2951–2965.
- Mitchell, H. L., & Houtekamer, P. L. (2000). An adaptive ensemble Kalman filter. *Monthly Weather Review*, 128, 416–433.
- 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 Resources Research*, 48, W05525, <http://dx.doi.org/10.1029/2011WR011420>.
- Njoku, E. G., & Chan, T. K. (2006). Vegetation and surface roughness effects on AMSR-E land observations. *Remote Sensing of Environment*, 100, 190–199.
- Njoku, E. G., Jackson, T. L., Lakshmi, V., Chan, T., & Nghiem, S. V. (2003). Soil moisture retrieval from AMSR-E. *IEEE Transactions of Geosciences and Remote Sensing*, 41, 215–229.
- Olioso, A., Inoue, Y., Ortega-Farias, S., Demarty, J., Wigneron, J. -P., Braud, I., et al. (2005). Future directions for advanced evapotranspiration modeling: assimilation of remote sensing data into crop simulation models and SVAT models. *Irrigation and Drainage Systems*, 19, 377–412.
- Ozalkan, C., Sepetoglu, H. T., Daur, I., & Sen, O. F. (2010). Relationship between some plant growth parameters and grain yield of chickpea (*Cicer arietinum* L.) during different growth stages. *Turkish Journal of Field Crops*, 15, 79–83.
- Reichle, R., & Koster, R. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, 3, 149–154.
- Reichle, R., McLaughlin, D., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130, 103–114.
- Sehgal, V. K., Sastri, C. V. S., Karla, N., & Dadhwal, V. K. (2005). Farm-level yield mapping for precision crop management by linking remote sensing inputs and a crop simulation model. *Journal of the Indian Society of Remote Sensing*, 33, 131–136.
- Teal, R. K., Tubuna, B., Girma, K., Freeman, W., Arnall, D. B., Walsh, O., et al. (2006). In-season prediction of corn grain yield potential using Normalized Difference Vegetation Index. *Agronomy Journal*, 98, 1488–1494.
- Thorp, K. R., Hunsaker, D. J., & French, A. N. (2010). Assimilating Leaf Area Index estimates from remote sensing into the simulations of a cropping systems model. *Transactions of the ASABE*, 53, 251–262.
- Thorp, K. R., Wang, G., West, A. L., Moran, M. S., Bronson, K. F., White, J. W., et al. (2012). Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models. *Remote Sensing of Environment*, 124, 224–233.
- Vazifedoust, M., Van Dam, J. C., Bastiaanssen, W. G. M., & Feddes, R. A. (2009). Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. *International Journal of Remote Sensing*, 30, 2523–2545.
- Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130, 1913–1924.