# Assimilation of SMOS Soil Moisture for Quantifying Drought Impacts on Crop Yield in Agricultural Regions

Subit Chakrabarti, *Student Member, IEEE,* Tara Bongiovanni, Jasmeet Judge, *Senior Member, IEEE,* Lincoln Zotarelli, and Cimelio Bayer

***Abstract*—This study investigates the effects of agricultural drought on crop yields, through integration of crop growth models and remote sensing observations. The soil moisture (SM) product from SM and Ocean Salinity (SMOS) mission obtained at 25 km was downscaled to a spatial resolution of 1 km, compatible with the crop models. The downscaling algorithm is based upon information theoretic learning and uses data-driven probabilistic relationships between high-resolution remotely sensed products that are sensitive to SM and *in situ* SM. The downscaled SM values are assimilated in the crop model using an Ensemble Kalman filter-based augmented state-vector technique that estimates states and parameters simultaneously. The downscaling and assimilation framework are implemented for predominantly agricultural region of the lower La-Plata Basin (LPB) in Brazil during two growing seasons. This rain-fed region was affected by agricultural drought in the second season, indicated by markedly lower precipitation compared to the first growing season. The downscaled SM was compared with the *in situ* SM at a validation site and the root mean square difference (RMSD) was $0.045\ \mathrm{m}^3/\mathrm{m}^3$. The crop yields estimated by the downscaling–assimilation framework were compared with those provided by the Companhia Nacional de Asastecimento (CONAB) and Instituto Brasileiro de Geografia e Estatistica (IBGE). The assimilated yields are improved during both seasons with increased improvement during the second season that was affected by agricultural drought. The differences between the assimilated and observed crop yields were 16.8% during the first growing season and 4.37% during the second season.**

***Index Terms*—Agricultural drought, crop growth models, data assimilation, data fusion, downscaling, microwave remote sensing, soil moisture (SM).**

## I. Introduction

DROUGHT is one of the major natural hazards causing losses worth billions of dollars each year. Typically, drought indices (DIs) that are functions of parameters such as precipitation (PPT), evapotranspiration, soil moisture (SM), land cover (LC), and leaf area index (LAI) are used by various agencies such as International Production Assessment Division to help asses and respond to drought [1]–[3]. These indices are mostly applicable for assessing meteorological droughts that are characterized by lower than expected precipitation in a region. Many indices, such as in the standardized precipitation indices (SPIs) [4], use precipitation maps as the only input. Precipitation has also been used indirectly as an indication of SM using water balance models, e.g., the Palmer 2 layer model, that use the palmer drought severity index (PDSI); the crop condition data retrieval and evaluation (CADRE) data base management system (DBMS); and the multilayer SM schemes in land surface models (LSMs) in the North American Land Data Assimilation System (NLDAS) [5]. However, these DIs are not accurate in predicting and monitoring agricultural droughts that are characterized by prolonged periods of reduced SM in the root zone (RZSM). In addition, the models used to estimate the DI typically assume constant LS, soil properties, and topography [6]–[8]. For droughts in vegetated regions, Brown *et al.* [2] proposed a composite DI called VegDRI that uses supervised classification and regression to a database with four temporally static biophysical variables and micrometeorological conditions from the past 16 years to form a seasonal model. The model is applied to current geospatial data to classify drought-induced vegetation stress into seven classes [9], [10]. The methodology has been tested across seven states in the US and was found to perform much better at estimating crop stress than traditional DIs. However, this approach does not provide quantitative crop yield estimates.

Assessing agricultural droughts is important in accurately quantifying their impacts on crop yields, particularly in rain-fed systems. One of the reasons the global crop yields are not accurately estimated in the presence of agricultural drought is the imperfect regional characterization of RZSM, based on only sparsely available *in situ* measurements [7]. Satellite-based SM products, from microwave sensors, such as those aboard the European Space Agency-SM and Ocean Salinity (SMOS) mission [11], are available every 2–3 days at a spatial resolution of 25 km. The planned NASA-SM Active Passive (SMAP) mission will use a combined active and passive microwave system at L-Band to provide SM at spatial resolutions of 36 and 9 km [12]. However, satellite-derived SM estimates are indicative of SM in the top 5–10 cm of the soil (henceforth, near-surface SM) and may not be good indicators of RZSM for agricultural drought assessments. This near-surface SM is incorporated into a water balance model to provide RZSM, e.g., in [13]. However, for applications in heterogeneous areas under dynamic LC such as in agricultural regions, these coarse scale SM or RZSM have to be spatially downscaled.

Recent studies have implemented various approaches to downscale coarse-scale SM products to spatial resolutions that

Manuscript received November 01, 2013; accepted February 26, 2014. Date of publication May 12, 2014; date of current version November 04, 2014. This research was supported by the NASA-Applied Sciences Program NNX12AJ86G.

S. Chakrabarti, T. Bongiovanni, and J. Judge are with the Center for Remote Sensing, Agricultural and Biological Engineering Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32611 USA.

L. Zotarelli is with the Horticultural Sciences Department, University of Florida, Gainesville, FL 32608 USA.

C. Bayer is with the Department of Soil Science, UFRGS—Universidade Federal do Rio Grande do Sul, 90040-060 Porto Alegre, RS, Brazil (e-mail: subitc@ufl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2014.2315999

can be used for agricultural studies. Empirical models based upon statistical methods such as regression [14]–[16], multiscale tree [17], and geostatistical methods such as co-kriging and block kriging have been used to extract fine scale SM information from diverse coarse scale RS products. Another class of techniques employ the triangle method [18], [19] using statistical models to extrapolate the dependent data within a hypothetical triangle formed by the observed data, [20], [21]. Piles *et al.* [21] developed a linking model from the triangle formed by the observed data at coarse scale and applied the model to data at finer scale, assuming equivalence at both scales. Merlin *et al.* [20] calculated soil evaporative efficiency [22] as a function of NDVI and LAI and used the triangular relationship between the efficiency and SM at the two scales to obtain high-resolution SM estimates. Both the methods were validated using data from the Murrembridge River catchment area in Southern Australia during the National Airborne Field Experiment '06 [23]. Piles *et al.* [21] used data from 3 days during a 4-week period in January and February 2010. Lack of NDVI observations from MODIS limited their validation study to just 3 days. Although the annual precipitation and LCs in the area was heterogeneous, the validation limited the diversity. Drusch *et al.* [24] used model climatology to downscale SM but the estimates were found to have significant bias [25]. The robustness of these methods over different vegetation and precipitation conditions remains largely unknown.

The downscaling algorithm based on the principle of relevant information (PRI) offers a robust alternative to spatial downscaling algorithms that use hierarchical models or empirical relationships across different scales. The PRI is based upon the measures of entropy and mutual information and provides a hierarchical decomposition of spatial data that is optimal in terms of transfer of information across spatial scales [26]. Statistical data descriptors beyond the second-order moment are necessary to capture the complete stochastic structure of data and to preserve maximum information to achieve optimal downscaling, especially over heterogeneous and dynamic environments. The algorithm uses all the statistical moments of the data and other information theoretic data descriptors to downscale SM across scales. Because it is primarily data driven and does not employ complex stochastic models, unlike the SVMs and Neural Networks, it is computationally more tractable [27]. Recently, the PRI-based downscaling methodology was evaluated using synthetic observations in North Central Florida and compared with Universal Triangle method [21]. The PRI was found to be robust in heterogeneous and dynamic agricultural region and during changes in micrometeorological conditions [28].

As mentioned before, the remotely sensed near-surface SM can be assimilated in a land surface [29] or a crop growth model [30] to provide optimal estimates on RZSM. Such mechanistic crop growth models are typically used to simulate growth and development to estimate crop yield and biomass [31]–[33]. However, uncertainties in these estimates increase over time, due to errors in computation, initial conditions, forcings, and model parameters. Satellite-derived SM estimates can be used to reduce the model bias through data assimilation techniques in which state and parameters are simultaneously updated using

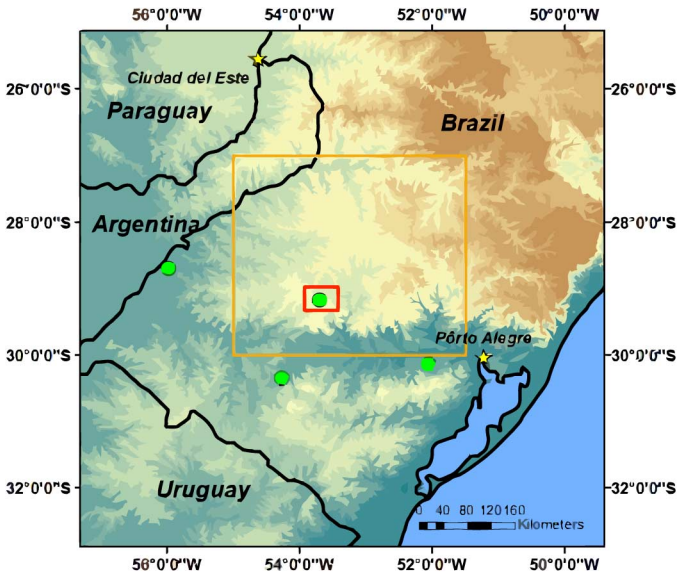



Fig. 1. Study region in LPB showing the *in situ* data sites. The green circles represent the *in situ* sites used for transformation function and red square represents the *in situ* site for validation of the methodology.

either Ensemble Kalman filter (EnKF) [29], [30] or particle filters [34]. Errors caused by parameter uncertainty can also be accounted for by using the augmented state vector technique in the EnKF-based assimilation algorithm. Li *et al.* [35] assimilated the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) SM retrievals into the Noah LSM and found that the SM fields are less biased if a mass conservation scheme is implemented. This did not result in the expected maximum minimization of the estimation error because the algorithm only updated states and not parameters in the EnKF, and because the retrievals were not downscaled prior to using them in the model [35]. Kaheil *et al.* [36] used support vector machines (SVM) combined with variogram analysis to downscale and assimilate SM observations from airborne imagery from a scale of 800 to 50 m. Extending SVM-based methodology to coarser scales of satellite observations may be difficult due to computational complexity and over-fitting the training data [37], further complicated during dynamic vegetation and heterogenous micrometeorological conditions.

The goal of this study is to understand the impacts of agricultural drought on crop yields through integration of remote sensing observations and crop growth models. The primary objectives of this study are to: 1) downscale near-surface SM from SMOS obtained at 25 to 1 km for two growing seasons in the rain-fed Brazilian region of the La-Plata Basin (LPB) (Fig. 1) using PRI. The two growing seasons primarily differ in the total PPT received and 2) assimilate the downscaled SM at 1 km into the decision support system for agrotechnology transfer (DSSAT) cropping system model [31] in agricultural regions under dynamic micrometeorological conditions. Fig. 2 shows the overall flow diagram to achieve the two objectives.

In Section II, we describe the region where the downscaling and assimilation algorithms were implemented, and the micrometeorological, remotely sensed and *in situ* data used in the study.

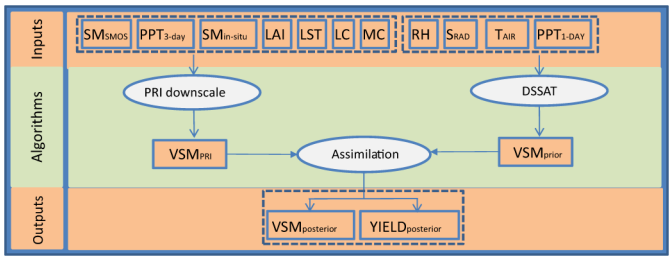



Fig. 2. Flow diagram of the integration of remote sensing observations and crop growth model in this study.

## II. Study Region and Data Sets

The study was conducted in region of a $161 \times 161$ km$^2$ in LPB in Brazil, stretching from 27 ° to 30 °S and 51.5 ° to 55 °W, as shown in Fig. 1. The LPB with an area of about $3 \times 10^6$ km$^2$ is one of the largest river basins in the world and includes parts of five countries, namely Argentina, Bolivia, Brazil, Paraguay, and Uruguay. The region is a major agricultural region of economic importance for all the countries [38]. The basin is vulnerable to high losses in crop yields due to agricultural drought. Climate change is predicted to continue to produce warmer and drier climate in the region.

To downscale SMOS SM at 25 km to 1 km, satellite-based observations of LAI, LC, LST, and PPT, *in situ* observations of SM and soil texture were used. The satellite-based observations used in this study along with their spatiotemporal resolutions are listed in Table I. The 3-h PPT data were averaged over a 3-day time period using a moving window to obtain 3-day averaged PPT. Observations at three *in situ* sites were used to obtain the transformation function for downscaling SM observations. These, along with the fourth site used for validation, are listed in Table II. The satellite data used for the DSSAT forcings are also listed in Table I. Daily aggregated values of PPT were used as DSSAT forcing.

The predominate soils in the Brazilian LPB region are red Latosols (Oxisols), which are characterized by their structural stability due to a high Fe and Al oxide content and a clayey texture, with 30%–60% clay [39]. As clay is the major soil component in the region, mass fraction of clay (MC) is used as an indicator of soil texture in the algorithm. The soil texture information was obtained from the ISRIC-WISE soil profile database [40]. The region consisted of primarily soybean crop, with the growing season from September to April. The wilting point, field capacity, porosity, and hydraulic conductivity were obtained from the DSSAT V4.5 soils database.

## III. Theory

### A. PRI Downscaling Framework

The downscaling methodology using PRI is implemented in two steps.

1) An initial estimate of SM is obtained at 1 km using transformation functions that relate LST, LAI, PPT, LC, and MC to *in situ* SM at the same spatial resolution of 1 km.
2) This initial estimate is merged with the SM observations at the coarser resolutions of 25 km using PRI, to obtain improved estimates of SM at 1 km.

TABLE I
Satellite Observations

| Sl. no. | Physical quantity | Original resolution | Interpolated resolution (km) | Satellite | Temporal resolution |
|---|---|---|---|---|---|
| 1[a] | SM | 25 km | – | NASA–SMOS | 2–3 days |
| 2[a,b] | Precipitation | 0.25° | 1 | TRMM | 3 h |
| 3[a] | LAI | 1 km | – | NASA Aqua/Terra (MODIS) | 4 days |
| 4[a] | Land surface temperature | 1 km | – | NASA Aqua/Terra (MODIS) | 1 day |
| 5[a] | LC | 0.5 km | – | NASA Aqua/Terra (MODIS) | 1 year |
| 6[b] | Total daily solar radiation | 0.125° | 1 | GOES-GSIP | 3 h |
| 7[b] | Maximum and minimum temperatures | 0.25° | 1 | GOES-5 | Daily |

[a]Data used for transformation function training.
[b]Data used for forcing DSSAT.

*1) Step 1: Transformation of LST, LAI, PPT, LC, and MC at 1 km for Initial Estimates of SM at 1 km:* The main objective of the transformation process is to obtain a probabilistic relationship between SM at 1 km using information from LST, LAI, 3-day PPT, LC, and MC obtained at the same scale. A discrete formulation of the Bayes rule is used to estimate *in situ* SM, as given in (1), whereas SM is discretized into $k$ classes, $i \in [1, k]$, LST into $k_1$ classes $LST_{i_1}, i_1 \in [1, k_1]$, LAI into $k_2$ classes $LAI_{i_2}, i_2 \in [1, k_2]$, 3-day PPT into $k_3$ classes $PPT_{i_3}, i_3 \in [1, k_3]$, LC into $k_4$ classes $LC_{i_4}, i_4 \in [1, k_4]$, and MC into $k_5$ classes $MC_{i_5}, i_5 \in [1, k_5]$

$$
\begin{aligned}
&p(SM_i|LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}) \\
&\quad = \frac{p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}|SM_i)p(SM_i)}{p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5})} \\
&SM_i = \underset{SM_i}{\operatorname{argmax}} \frac{p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5})p(SM_i)}{p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5})} \\
&p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}) \\
&\quad = \sum_{i=1}^{k} p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}|SM_i)p(SM_i).
\end{aligned} \tag{1}
$$

A nonparametric probability density function (pdf) estimation technique known as Parzen windows [41] was used to estimate the pdf, $p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}|SM_i)$, in (1). In this technique, symmetric functions, known as kernels, are used to estimate the pdf of a random variable based on a finite set of data samples. The kernels assign weights to each of the samples based upon their distance from the point at which the pdf is estimated, with higher weights being assigned to the samples that are closer to the point. In this study, a zero-mean, unity-variance Gaussian function was used as the kernel. With Gaussian kernels, the weight decreases exponentially with the square of the distance, rendering the distant points irrelevant.

TABLE II
SITES FOR *IN SITU* SM DATA

| Site no. | Latitude (°S) | Longitude (°W) | Location | Source | Duration (temporal resolution) | Soil texture |
|---|---|---|---|---|---|---|
| 1[a] | 29.17 | 53.7 | Julio Castilhos, Rio Grande do Sul | SINDA | 10/2004 – 07/2012 (every 3 h) | Neosols |
| 2 | 28.69 | 55.96 | Saõ Borja, Rio Grande do Sul | SINDA | 10/2004 – 07/2012 (every 3 h) | Nitosols |
| 3 | 30.34 | 54.26 | Saõ Gabriel, Rio Grande do Sul | SINDA | 10/2004 – 07/2012 (every 3 h) | Latosols |
| 4 | 23.20 | 51.18 | Londrina, Paraná | CREX | 05/2003 – 09/2003 (every 15 min | Latosols |
| 5 | 30.09 | 51.67 | Eldorado do Sul, Rio Grande do Sul | UFRGS | 12/2010 – 11/2012 (every 2 to 45 days) | Acrosols |

[a]Site used for validation.
SINDA, Sistema Nacional de Dados Ambientais National System of Environmental Data.
UFRGS, Universidade Federal do Rio Grande do SUL.
CRE X, crop rotation experiment.

Given $n$ training samples $LST_j, LAI_j, LC_j, PPT_j, MC_j (j \in [1, n])$, the probability of LST, LAI, LC, PPT, and MC belonging to classes $LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}$, respectively, for a known $SM_i$ is represented as $p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}|SM_i)$ and computed using (2). The pdf is estimated at all combinations of LST, LAI, PPT, LC, and MC classes resulting in a pdf of dimension $k_1 \times k_2 \times k_3 \times k_4 \times k_5$

$$\begin{aligned} p(LST_{i_1}, LAI_{i_2}, PPT_{i_3}, LC_{i_4}, MC_{i_5}|SM_i) &= \frac{1}{n}\sum_{j=1}^{n} \frac{1}{h_1 h_2 h_3 h_4 h_5} \psi\left(\frac{LST_{i_1} - LST_j}{h_1}\right) \\ &\times \psi\left(\frac{LAI_{i_2} - LAI_j}{h_2}\right)\psi\left(\frac{LC_{i_3} - LC_j}{h_3}\right) \\ &\times \psi\left(\frac{PPT_{i_4} - PPT_j}{h_4}\right)\psi\left(\frac{MC_{i_5} - MC_j}{h_5}\right) \end{aligned} \quad (2)$$

where $\psi$ is the Gaussian kernel and $h_i, i \in [1, 2, 3, 4, 5]$, is a smoothing parameter, which relates to the variance in $\psi$ for LST, LAI, PPT, LC, and MC, respectively. The parameter $h_i$ was determined based on Silverman's rule, which is a standard rule-of-thumb that is used in kernel-based estimation algorithms [42], as given in (3). The pdf $p(SM_i)$ in (1) is a one-dimensional pdf that was computed at $k$ points

$$h_i = 1.06 \min\left[\sigma_i, \frac{I_{QR_i}}{1.34}\right] n^{-1/5} \quad (3)$$

where $\sigma_i$ and $I_{QR_i}, i \in [1, 2, 3, 4, 5]$ are the standard deviation and interquartile range, respectively, computed from the $n$ samples for LST, LAI, PPT, LC, and MC.

Computing the above pdfs from the $n$ training samples can be viewed as a learning step. Given LST, LAI, PPT, LC, and MC information at 1 km, the initial estimates of SM at 1 km are obtained using (1).

*2) Step 2: Downscaling SM Using PRI:* The PRI provides a hierarchical decomposition of spatial data that is optimal in terms of transfer of information across scales, by minimizing or maximizing a balance between entropy and the Kullback–Leibler (KL) divergence of the original data with respect to its upscaled or downscaled versions [41]. While entropy is a redundancy reduction term, the KL divergence is an information preserving term. If entropy is maximum, while KL-divergence is 0, the observations remain the same, while if KL-divergence is high with zero entropy, the data set is represented by a single point, the mean. Depending on the proportion of the two metrics, one can find a range of structures, which can be cast as an information extraction process. The optimization equation for PRI is

$$\max_X J(X) = H(X) + \beta KL(p_X \| p_S) \quad (4)$$

where $J(X)$ is the cost function, $p_S$ is the pdf of the original data, $p_X$ is the pdf at each iteration, $H(X)$ is the entropy, and $KL$ is the KL divergence. The $\beta$ is a user-defined weighting parameter that balances the redundancy and information preservation in $J(X)$. As the value of $\beta$ increases, the cost function gives more and more emphasis to KL, thus preserving more information about the data at the cost of extremely high-redundancy reduction. Equation (4) may be rewritten in a different form by using Renyi's quadratic entropy ($H_2(X)$) and cross-entropy ($H_2(X, S)$), respectively, [43], to obtain a more practical cost function $J(X)$, namely

$$\max_X J(X) = (1 - \beta)H_2(X) + 2\beta H_2(X, S). \quad (5)$$

An iterative solution $x_k, k = [1, 2, \ldots N]$ can be obtained by taking the derivation of $J(X)$ with respect to $x$ and equating it to 0. The resulting fixed-point rule is given as

$$x_k(t+1) = c\frac{1-\beta}{\beta}\frac{\sum_{j=1}^{N}\psi\big(x_k(t) - x_j(t)\big)x_j(t)}{\sum_{j=1}^{M}\psi\big(x_k(t) - s_j\big)} + \frac{\sum_{j=1}^{M}\psi\big(x_k(t) - s_j\big)s_j}{\sum_{j=1}^{M}\psi\big(x_k(t) - s_j\big)} \quad (6)$$

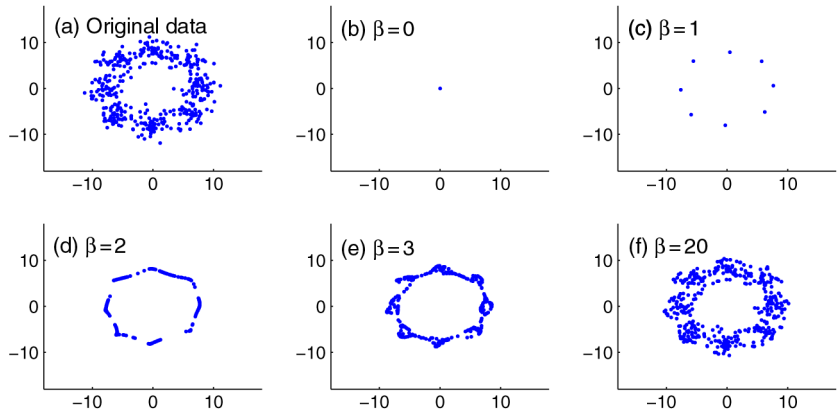



Fig. 3. (a) Conventional implementation of PRI where $X$ is initialized to a sample data set $S$, producing different nonlinear tranformations in $X$-based upon $\beta$. (b) Implementation of PRI for downscaling by morphing $X$ and $S$, where $X$ and $S$ are initialized to the observations obtained from the transformation step at 1 km and the coarse-resolution satellite observation at 10 km.

where

$$c = \frac{V(X,S)M}{V(X)N}$$

$V(X)$ is the information potential [44] of data points in $X, V(X,S)$ is the cross-information potential between data points in $X$ and $S, N$, and $M$ are the number of data points in $X$ and $S$, respectively, $t$ represents the iteration, and $\psi$ is the Gaussian kernel.

To illustrate the functionality of the PRI, a simple example created by a set of Gaussian clusters placed in a circle is shown in Fig. 3(a). The variable $X$ in (5) is initialized to the simulated Gaussian cluster data set ($S$). Applying the PRI for different values of $\beta$ reveals, various structures relevant to the data set. At one extreme, we have a single point, and at the other extreme, we have the data itself as the solution. As we move from one end to another, the PRI displays the modes and principal curves revealing different hierarchical levels of structure from the data.

The implementation of PRI was slightly modified for downscaling SM. Instead of initializing $X$ to $S$, the initial estimates of SM at 1 km obtained from Section III-A1 were used as $X$ in (5). SM at 10 km was used as $S$. Under the assumption that $X$ and $S$ represent the same imagery using different measurements, the crossentropy term $H_2(X,S)$ should be largest when they coincide. This condition is obtained when $\beta = 1$ in (5). However, this case provides the best downscaled $X$ supported by $S$, in spite of the fact that $S$ represents only an approximate and coarse-level information on SM. In this study, an intermediate value of $\beta = 2$ was chosen so that the PRI-image would approximate the mean level of SM at coarse scales but will also embed the level of detail provided by the initial estimates of SM at 1 km, to obtain morphed estimates of SM at 1 km.

### *B. Decision Support System for Agrotechnology Transfer*

The DSSAT is a suite of crop growth models for 28 different crops. The DSSAT model used in this study is the nonlinear CROPGRO soybean model [45], which has been extensively applied in different regions of the world [46]–[48]. The model simulates crop growth and development at a daily time step and estimates yield and biomass. The model consists of modules for weather, field management, soil, and development and growth of the plant. The minimum weather inputs to the weather module are daily solar radiation, maximum and minimum temperature, and precipitation. The field management module contains the planting information such as date, row spacing, and plant population, irrigation (if applicable), fertilizer, and harvest date. The soil module includes soils that are divided into user-defined layers, and each layer has specific properties. In this study, the soil was composed of two layers, a surface layer from 0 to 5 cm and a deeper layer from 5 to 120 cm, with the same constitutive properties, characteristic of soils with high clay content. The parameters of the first layer are updated during assimilation, whereas the parameters of the second layer are assumed to stay the same during each season. Root-zone SM (RZSM) was calculated using the following equation, where $k$ indicates the total number of layers within the root zone and $\Delta z$ is the thickness of the layer

$$RZSM = \sum_{i=1}^{k} SM_i \Delta z_i \tag{7}$$

where $SM_i$ is the SM at depths 0–5 cm, for $i = 1$ and 5–120 cm, for $i = 2$.

### *C. Assimilation Algorithm—EnKF*

The SM in the top 5 cm of the soil ($\mathrm{SM}_{0-5}$) is the SM obtained using the PRI downscaling algorithm. These are assimilated using an augmented state vector technique with the EnKF-based algorithm. The EnKF algorithm propagates an ensemble of state vectors in parallel such that each state vector represents one realization of model replicates. For each ensemble member, the state equation in the EnKF is [49], [50]

$$x_t^{i-} = f\left(x_{t-1}^{i+}, u_{t-1}^{i}, \theta_{t-1}^{+}, t-1\right) + \omega_{t-1}^{i} \tag{8}$$

where $f(\cdot)$ is the DSSAT model and $x_t^{i-}$ is the state of the $i$th realization prior to the update at time $t$; $x_{t-1}^{i+}$ is the posterior state of the $i$th realization at time $t-1$, $u_{t-1}^{i}$ represents the meteorological forcings (e.g., temperature, precipitation, and solar

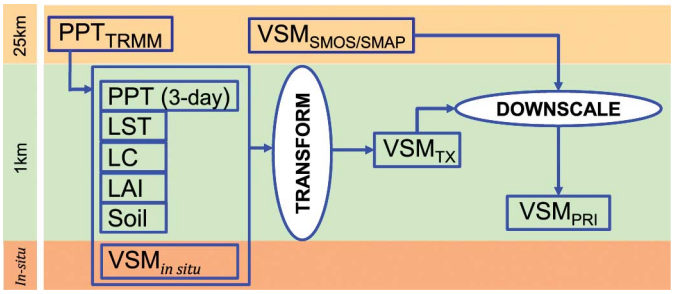



Fig. 4. Flow diagram of the PRI downscaling algorithm.

radiation), $\theta^{+}_{t-1}$ is the posterior estimate of parameters of the nonlinear model (e.g., soil and cultivar parameters), and $\omega^{i}_{t-1}$ is the model error. The model physics were assumed to be perfect, $\omega^{i}_{t-1} = 0$ in this study.

The A matrix, which holds the ensemble of state vectors, can be written as, $A = \{x^1, x^2 \ldots, x^N\}$, where $x^i$ represents each member of the ensemble and $N$ represents number of ensemble members. The observation equation that relates the prior state to the observations ($d_i$) can be expressed as

$$d^i = hx^{i-}_t + \epsilon^i_t \tag{9}$$

where $h(\cdot)$ is the measurement operator relating the state variables to the observations, and $\epsilon$ is the error associated with the observations. The ensemble of perturbed observations can be expressed as $D = \{d^1, d^2 \ldots, d^N\}$ and the ensemble of perturbations as $\gamma = \{\epsilon^1, \epsilon^2 \ldots, \epsilon^N\}$. Then, mathematically, the EnKF can be represented as

$$A^+ = A^- + K(D - HA^-) \tag{10}$$

where $A^-$ denotes the ensemble of prior states and $A^+$ denotes the ensemble of posterior states, $K$ is the Kalman gain, $D$ is the ensemble of perturbed observations, and $H$ is the operator relating the ensemble of perturbed observations to the ensemble of states. The Kalman gain can be calculated as [51]

$$K = P_e H^T \left(HP_e H^T + R_e\right)^{-1} \tag{11}$$

where $P_e$ represents the prior ensemble covariance matrix and $R_e$ is the observation error covariance matrix. Then, the final update equation is

$$A^+ = A^- + P_e H^T \left(HP_e H^T + R_e\right)^{-1}(D - HA^-). \tag{12}$$

## IV. Methodology

### A. PRI-Based Downscaling Algorithm

The PRI downscaling methodology has two steps, as described in Section III-A (Fig. 4). The transformation function is obtained using *in situ* SM at the three data sites listed in Table II and satellite observations of LAI, LST, 3-day PPT, LC, and MC at those sites. The soil texture, indicated by MC, is discretized into three ranges of mass fractions, 0–0.36; 0.37–0.45; and 0.45–0.6, as shown in Fig. 5. In this study, the transformation function is assumed spatially invariant over areas with similar soil texture, given similar soil characteristics, and the predominant LC of soybean in the region. The initial estimate of SM is obtained during the two seasons, from September 1 (DoY 244) to April 30 (DoY 120), in 2010 and 2011, using (1). The SM estimated using the transformation function were compared with the *in situ* data at the validation site (Site 1 in Table II) to test the validity of the assumption of spatial invariance.

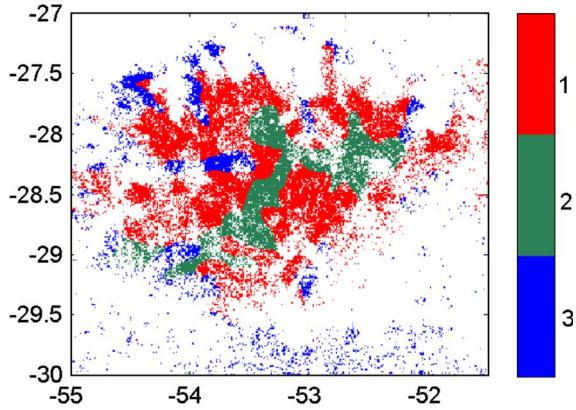



Fig. 5. Soil texture classes, as indicated by volume fraction of clay in the region. Red indicates a clay volume fraction of 0–0.36, green indicates a clay volume fraction of 0.37–0.45 and blue indicates a clay mass fraction of 0.45–0.6.

In the second step of the PRI algorithm, the initial SM at 1 km was merged with the SMOS SM data at 25 km, using (6) to obtained the downscaled SM at 1 km. The SMOS data were bias-corrected by the regional mean of the SMOS SM data at 25 km and the regional mean of the first estimate, during the season.

### B. DSSAT Simulation

The DSSAT simulations were conducted for two seasons in 2010–2011 and 2011–2012, both starting 75 days prior to planting on September 1 (DoY 244) and ending about 30 days after harvest on April 30 (DoY 120). The first season represents a "normal" year with a total PPT of 859 mm, whereas during the second year, there was an agricultural drought, with a total PPT of 389 mm over the region. Soybean planting density was 18 plants per m with a row spacing of 0.45 m [52]. Three distinct soil types, clay (C), clay loam (CL), and sandy clay loam (SCL) were created using the DSSAT V4.5 soil database. The soil properties for each are listed in Table III. These properties match the three clay categories used in the transformation function, as shown in Fig. 5.

The structure of DSSAT V4.5 does not allow the model to restart within season with posterior states, after assimilation. In this study, the DSSAT simulations were interrupted on the days when SMOS SM were available observations by setting the harvest date to the observation dates. Setting the harvest to an earlier date does not impact the growth or development of the crop in the model. The posterior states and parameters from assimilation were stored in a file. The simulation was restarted from the planting date with the original initial values of the states, but with updated parameters. At the time of observation, the

TABLE III
SOIL PROPERTIES FOR THREE SOIL LAYERS

| Soil type | LL ($m^3/m^3$) | DUL ($m^3/m^3$) | $K_{SAT}$ ($cm \cdot h^{-1}$) | $\rho$ ($g \cdot cm^{-3}$) | SAT ($m^3/m^3$) | Drainage rate (fraction/day) | Initial SM ($m^3/m^3$) |
|---|---|---|---|---|---|---|---|
| Clay | 0.283 | 0.399 | 0.39 | 1.32 | 0.413–0.488 | 0.05–0.5 | 0.34 |
| CL | 0.187 | 0.328 | 1.2 | 1.37 | 0.417–0.512 | 0.1–0.6 | 0.26 |
| SCL | 0.161 | 0.259 | 2.97 | 1.56 | 0.360–0.418 | 0.2–0.7 | 0.21 |

LL is the lower limit, DUL is the drain upper limit, $K_{SAT}$ is the saturated hydraulic conductivity, $\rho$ is the bulk density, SAT is the saturated soil water content, and SM is the soil moisture.

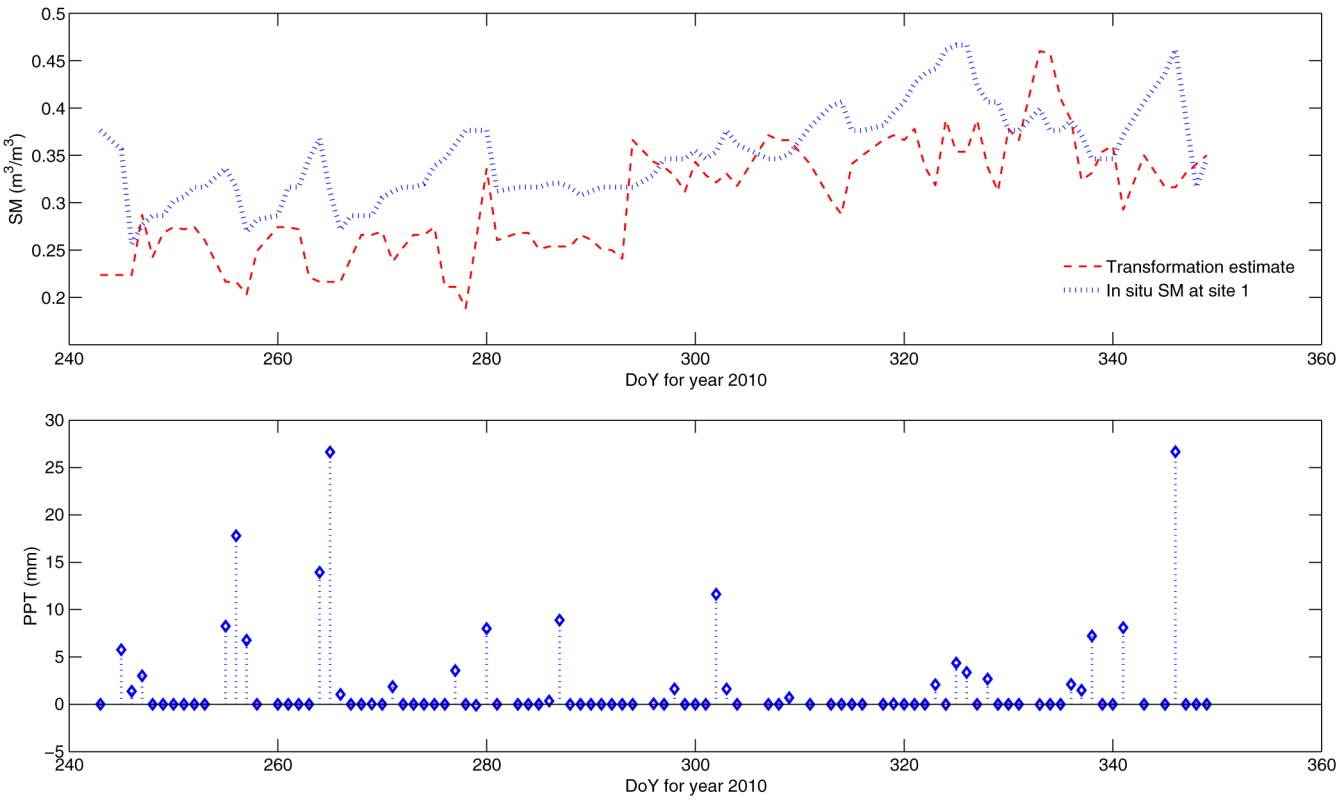



Fig. 6. First estimate of SM using transformation function and *in situ* SM during September to December 2010, at the SINDA station in Julio Castilhos, Rio Grande do Sul, which is the SM validation site. The precipitation, during that period, is also shown.

updated states from the file were inserted in the model and the simulation propagated until the next observation date, as shown in Fig. 2.

### C. EnKF Implementation

Among the inputs to the DSSAT model, precipitation observations typically have the highest error compared with other micrometeorological parameters. The errors can range between 2.9% and 12.3% of the total rainfall, depending upon the intensity and duration of the rainfall [53], [54]. A Gaussian observation error with a standard deviation of 12% of the observed value of precipitation was introduced at the time of precipitation. No error was introduced when there was no precipitation. The uncertainty in two soil parameters were included in the DSSAT simulations. Drainage rate and porosity were randomly generated from a uniform distribution. The cultivar parameters were taken from the DSSAT model. Out of the 49 total soybean varieties in the DSSAT database, 12 varieties were chosen within maturity groups 4–6 [55]. Fifty ensemble members are created from this data for assimilation. The four other soil parameters were obtained from the DSSAT V4.5 Soils as mentioned in Section II. This ensured that the relationships among the coefficients were maintained.

The downscaled SM obtained from Section IV-A was perturbed with a Gaussian error of zero mean and $0.04\ m^3/m^3$ standard deviation, to obtain the 50 ensemble members for the EnKF implementation [11]. The augmented state vector involved $SM_{0-5}$ and $SM_{5-120}$ states and drainage rate and porosity parameters that were updated during assimilation, with RZSM estimated from (7). The parameters estimated are examined over the two seasons to understand the impact of assimilation.

The SM values with and without assimilation obtained from the DSSAT were compared with the SINDA validation

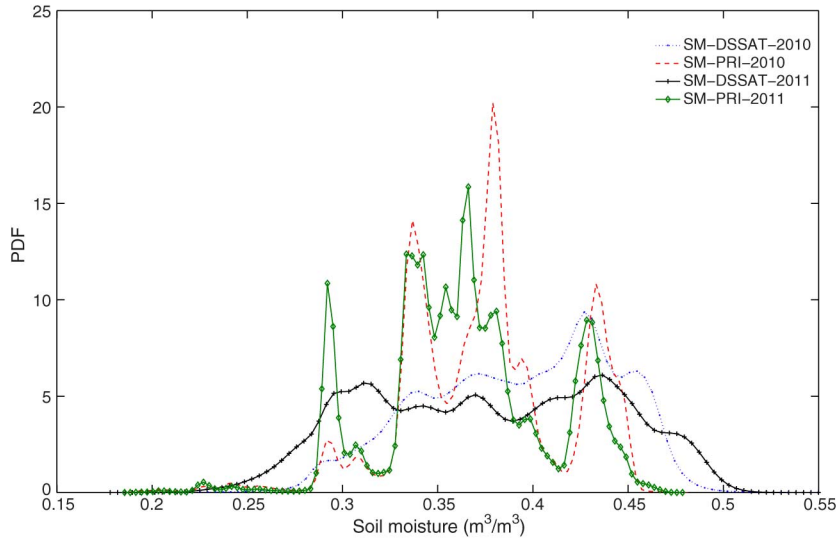



Fig. 7. KDE of the PRI estimated SM and the assimilated SM using Ensemble Kalman filters.

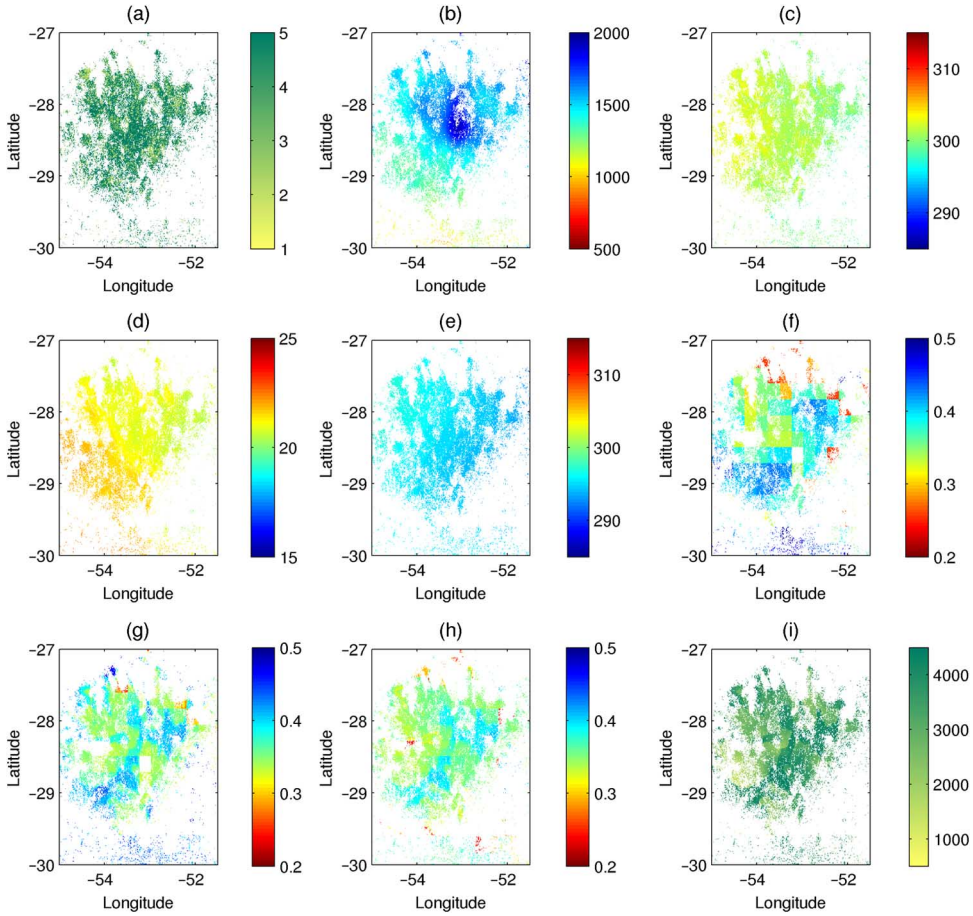



Fig. 8. Inputs and Results for 2010–2011 season over the study region. (a) Maximum land area index. (b) Total precipitation (mm). (c) Mean land surface temperature (K). (d) Mean daily solar radiation ($\mathrm{J \cdot m^{-2}}$). (e) Mean air temperature (K). (f) Mean SMOS SM at 25 km ($\mathrm{m^3/m^3}$). (g) Mean SM-PRI is the downscaled SM at 1 km ($\mathrm{m^3/m^3}$). (h) Assimilated SM ($\mathrm{m^3/m^3}$). (i) Final crop yields (kg).

site (Site 1 in Table II). The assimilated and open loop crop yields obtained from the DSSAT model are also compared with the crop yield statistics provided by the Companhia Nacional de Asastecimento (CONAB) and Instituto Brasileiro de Geografia e Estatistica (IBGE) to understand the feasibility of the downscaling and assimilation framework to estimate crop yields during normal and drought-impacted growing seasons.

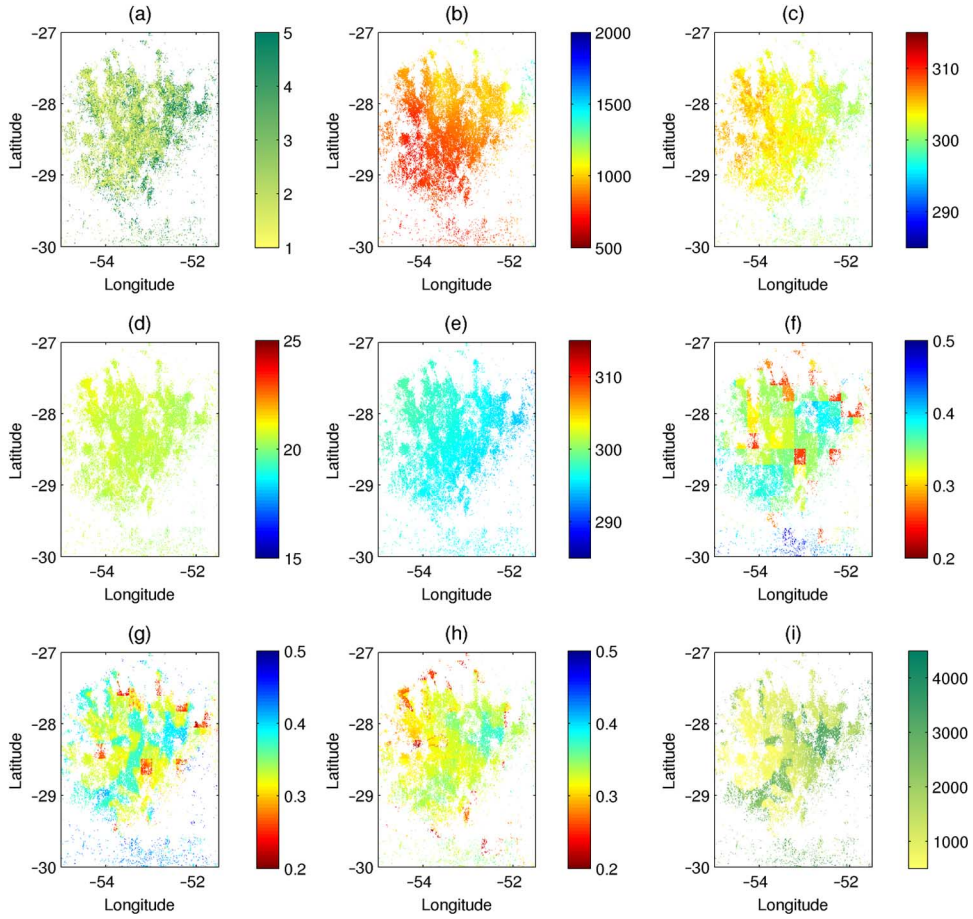



Fig. 9. Inputs and Results for 2011–2012 season over the study region. (a) Maximum land area index. (b) Total precipitation (mm). (c) Mean land surface temperature (K). (d) Mean daily solar radiation ($J \cdot m^{-}2$). (e) Mean air temperature (K). (f) Mean SMOS SM at 25 km ($m^3/m^3$). (g) Mean SM-PRI is the downscaled SM at 1 km ($m^3/m^3$). (h) Assimilated SM ($m^3/m^3$). (i) Final crop yields (kg).

## V. Results and Discussions

### A. PRI-Based Downscaling Algorithm

The initial estimates of SM were obtained at 1 km using the transformation function that extracts the structure and heterogeneity in the spatial distribution of SM. Fig. 6 shows the comparison of the *in situ* SM and the initial estimate, with the 3-day PPT, at the SINDA station in Julio Castilhos, Rio Grande do Sul, during the growing season in 2010. The *in situ* SM is well approximated by the initial SM estimated from the transformation function, with a root mean square difference (RMSD) of $0.044\ m^3/m^3$. The maximum difference of $0.17\ m^3/m^3$ is during high *in situ* moisture condition, but low 3-day averaged PPT. The initial SM estimate follows the trend of the PPT data closely.

For the second step of the PRI algorithm, the bias-corrected SMOS SM at 25 km is merged with the initial SM estimate at 1 km. Fig. 7 shows the kernel density estimate (KDE) of the downscaled SM data, for both the growing season, aggregated over the region. The growing season of 2010 has higher SM probabilities than during 2011 season, due to higher PPT. The modes for the first season suggests that the SM in the region is between $0.30\ m^3/m^3$ and $0.35\ m^3/m^3$ with dry and flooded conditions being less probable. For the second season, the probability is more for lower SM values. Figs. 8 and 9 show the mean SMOS SM, and the mean downscaled SM, along with the mean LST, total PPT, and maximum LAI inputs for the 2010–2011 and 2011–2012 season, respectively. The downscaled SM at 1 km captures the variability in the inputs and the SMOS SM. The structure in the soil texture map is preserved by the PRI while also accounting for the SMOS SM observations. On wet days, due to cloud cover, the LST data were mostly unavailable. In spite of some data gaps, PRI can generate an estimate, using correlations with other physical variables, demonstrating its robustness.

### B. DSSAT Simulation and EnKF Implementation

DSSAT was run using the forcings as described in Section IV-B. Fig. 7 also shows the comparison of the KDE obtained for the PRI SM at 1 km and the DSSAT estimated SM after assimilation, for both seasons, averaged over the region. The KDE of the assimilated SM is smoother compared with PRI SM, because of the reduction in uncertainties in the observations by EnKF. The Kullback–Liebler divergence (KLD) between the downscaled SM KDE, assimilated SM KDE, and a Gaussian distribution with mean and standard deviation equal to that of the respective data set was calculated.

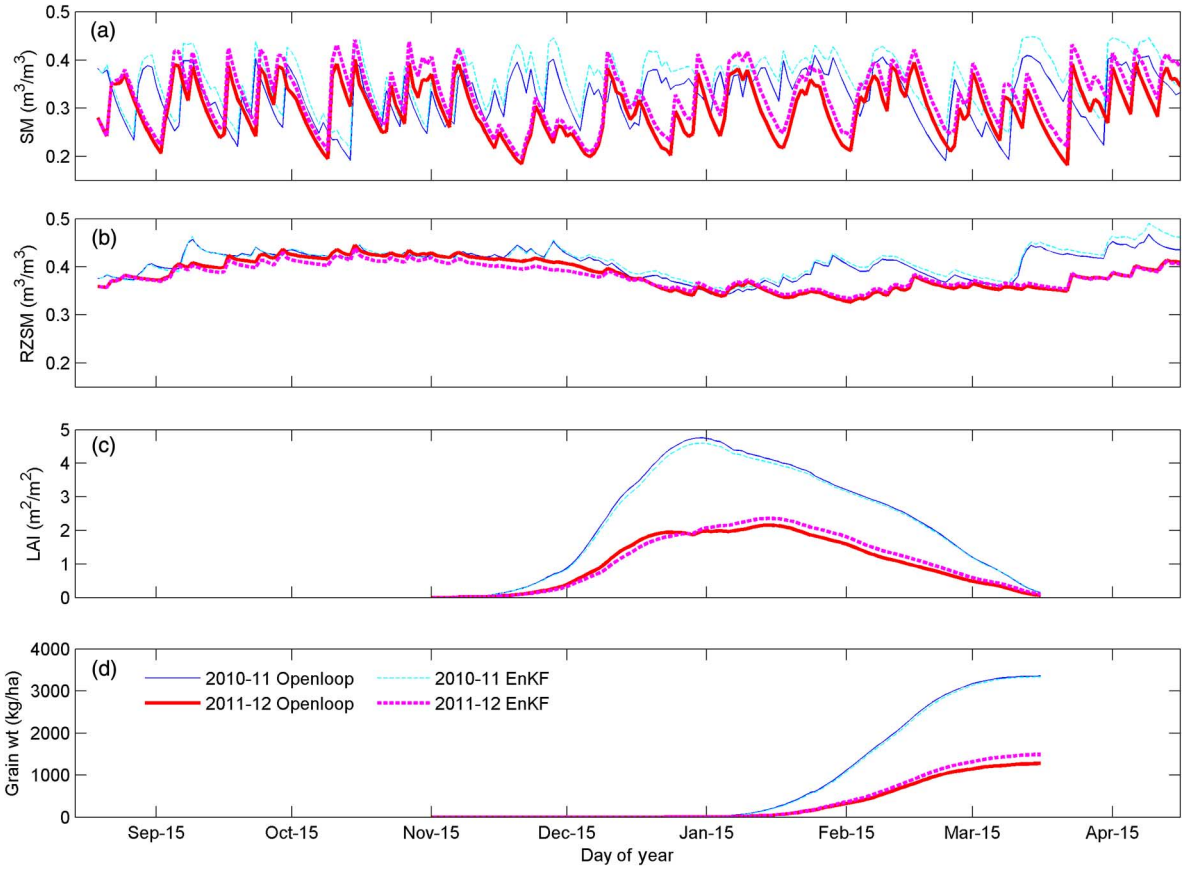



Fig. 10. Regional mean of (a) the volumetric SM ($\mathrm{VSM}_{0-5}, \mathrm{m}^3/\mathrm{m}^3$). (b) The root zone SM $\mathrm{RZSM}, \mathrm{m}^3/\mathrm{m}^3$. (c) LAI. (d) Grain weight (kg) from September to April, 2010–2011 and 2011–1012.

The KLD for the PRI SM case is 2.13, whereas that for the DSSAT assimilated SM is 0.6, indicating that the assimilated SM is closer to a Gaussian distribution. Figs. 8 and 9 show the mean inputs and outputs of the downscaling and assimilation algorithm during the 2010–2011 and the 2011–2012 seasons, respectively. For the 2010 season, the regional crop yield estimate is lower in areas with high PPT and SM and is lower in the southern part of the region where the PPT is very low. For the 2011 season, the crop yield estimate is uniformly low throughout the region, as is the region. The temporal variation of the assimilated SM, RZSM, and crop yield, along with the LAI is shown in Fig. 10. The SM is lower for the 2010–2011 season than for the 2011–2012 season, which is expected since 2011–2012 was a drought season. The LAI is also much lower for the 2011 season at that site, indicating poor crop health. Consequently, the crop yield estimates for the 2010–2011 season is also significantly lower than that for the 2011–2012 season.

As shown in Fig. 11, the updated parameters, porosity and drainage rate, have similar values for both seasons. For both seasons, the average porosity is $0.459\ \mathrm{m}^3/\mathrm{m}^3$ that is $0.019\ \mathrm{m}^3/\mathrm{m}^3$ higher than the open loop mean. However, the drainage rate oscillates around the 0.25 value through both the seasons. The drainage rate decreases rapidly from the open loop mean to a minimum of $0.13\ \mathrm{day}^{-1}$. As the plant grows, drainage rate increases to 0.37 and $0.34\ \mathrm{day}^{-1}$ for 2010–2011 and 2011–2012 season, respectively. After harvest, the drainage rate decreases. This would indicate more water can be retained in the bare soil. However, as the plant grows, water is better able to penetrate through the soil.

The means and standard deviations of the regional crop yields for the two seasons obtained from DSSAT, with and without assimilation, are compared with the yields from CONAB and IBGE Table IV. Although assimilated yields are improved during both seasons and match well with the observed yields in the region, with their estimates within one standard deviation, the improvement was more significant during the drought season. This indicated the viability of the downscaling–assimilation framework presented in this study to understand and predict the impacts of agricultural droughts through combination of remotely sensed observations and crop growth models. Moreover, the estimated yields are within one standard deviation of the published crop yield values, indicating robustness of the framework. This methodology can allow partial sparsity in data as well. The parameter $\beta$ is useful in dealing with data sparsity. It can be dynamically adjusted to make the downscaling algorithm respond to unavailability of data or to availability of data with high uncertainty.

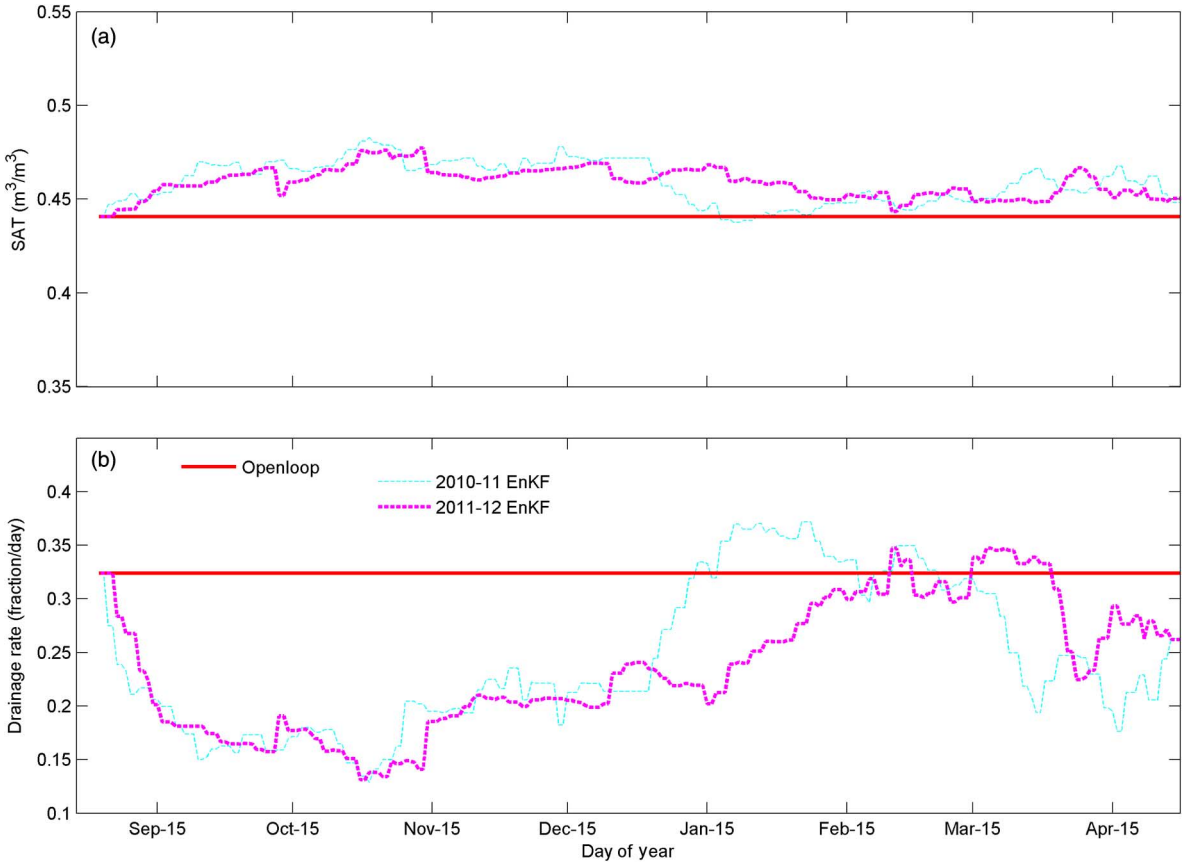



Fig. 11. Plot of ENKF parameters. (a) Saturated soil water content. (b) Drainage rate, for the two seasons.

TABLE IV
OPEN LOOP, ASSIMILATED, AND CONAB AND IBGE CROP YIELDS

| Season | Openloop mean (kg) | Openloop STD (kg) | EnKF mean (kg) | EnKF STD (kg) | CONAB yield (kg) | IBGE yield (kg) |
|---|---|---|---|---|---|---|
| 2010–2011 | 3361 | 660 | 3342 | 673 | 2845 | 2875 |
| 2011–2012 | 1280 | 574 | 1499 | 805 | 1560 | 1430 |

## VI. SUMMARY AND CONCLUSIONS

In this study, the SMOS SM product at 25 km is downscaled to 1 km and assimilated into the DSSAT crop growth model to understand the impact of agricultural drought on crop yields. The PRI downscaling algorithm utilizes the full pdf of the data instead of just the first and second moments and provides reasonable estimates even when some of the auxiliary data is not available. The downscaled estimate was validated by comparing it with the *in situ* measurement and was found to have an RMSD of $0.045\ \mathrm{m}^3/\mathrm{m}^3$ with a maximum difference of $0.17\ \mathrm{m}^3/\mathrm{m}^3$. The downscaled SM was assimilated using an EnKF technique with simultaneous updates in the states and parameters. The assimilated crop yield was improved during the two seasons, with a higher improvement during the second season, which was affected by agricultural drought. The crop yield at the validation site provided by CONAB and IBGE matched well with the observed crop yields.

## ACKNOWLEDGMENT

The authors acknowledge computational resources and support provided by the University of Florida High-Performance Computing Center for all the model simulations conducted in this study. The authors thank Dr. Clyde Fraisse, Agricultural and Biological Engineering Department, University of Florida, FL, USA; Mr. Marcello Moreira, Cargill Inc., Brazil; Dr. Jose Yi, Cargill Inc.; and Dr. Carlos Gustavo Tornquist for facilitating the implementation of this study in the region.

## REFERENCES

[1] M. Anderson, C. Hain, B. Wardlow, A. Pimstein, J. Mecikalski, and W. Kustas, "Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental united states," *GISsc. Remote Sens.*, vol. 45, no. 1, pp. 16–46, 2008.

[2] J. Brown, B. Wardlow, T. Tadesse, M. Hayes, and B. Reed, "The vegetation drought response index (VEGDRI): A new integrated approach for monitoring drought stress in vegetation," *J. Clim.*, vol. 24, no. 8, pp. 2025–2044, 2011.

[3] Y. Gu, E. Hunt, B. Wardlow, J. Basara, J. Brown, and J. Verdin, "Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data," *Geophys. Res. Lett.*, vol. 35, no. 22, pp. 401–405, 2008.

[4] T. McKee, N. Doesken, and J. Kleist, "Drought monitoring with multiple time scales," presented at the Proc. AMS Conf. Appl. Clim., Dallas, Texas, 1995.

[5] E. Mitchell, D. Lohmann, P. Houser, E. Wood, J. Schaake, A. Robock, B. Cosgrove, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. Pinker, J. Tarpley, D. Lettenmaier, C. Marshall, J. Entin, M. Pan, W. Shi, V. Koren, J. Meng, B. Ramsay, and A. Bailey, "The multi-institution North American land data assimilation system (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system," *J. Geophys. Res.*, vol. 109, no. D7, pp. 2156–2202, 2004.

[6] B. Narasimhan and R. Srinivasan, "Development and evaluation of soil moisture deficit index (SMDI) and evapotranspiration deficit index (ETDI) for agricultural drought monitoring," *Agric. Forest. Meteorol.*, vol. 133, no. 1-4, p. 6988, 2005.

[7] W. T. Crow, S. V. Kumar, and J. D. Bolten, "On the utility of land surface models for agricultural drought monitoring," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 9, pp. 3156–3166, 2011.

[8] A. Yagci, L. Di, and M. Deng, "The effect of land cover change on vegetation greenness-based satellite agricultural drought indicators: A case study in the southwest climate division of Indiana, USA," *Int. J. Remote Sens.*, vol. 34, no. 20, pp. 6947–6968, 2013.

[9] J. Brown, S. Pervez, and B. Wardlow, "Monitoring drought stress across multiple vegetation types within the upper colorado river basin," presented at the Proc. World Clim. Res. Prog. Open Sci. Conf., Denver, Colorado, 2011.

[10] J. Brown, S. Pervez, and B. Wardlow, "Assessment of 2006 and 2007 drought patterns in the vegetation drought response index across Nebraska," presented at the Proc. 17th Pecora Mem. Remote Sens. Symp. Denver, Colorado, 2008.

[11] Y. Kerr, P. Waldteufel, J. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, Aug. 2001.

[12] D. Entekhabi, E. Njoku, P. O'Neill, M. Spencer, T. Jackson, J. Entin *et al.*, "The soil moisture active/passive mission (SMAP)," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 3, 2008, pp. III–1–III–4.

[13] L. Brocca, F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer, "Improving runoff prediction through the assimilation of the ASCAT soil moisture product," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 10, pp. 1881–1893, 2012.

[14] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Adv. Water Resour.*, vol. 33, pp. 69–80, 2010.

[15] D. Liu and R. Pu, "Downscaling thermal infrared radiance for subpixel land surface temperature retrieval," *Sensors*, vol. 1, pp. 2695–2706, 2008.

[16] N. S. Chauhan, S. Miller, and P. Ardanuy, "Spaceborne soil moisture estimation at high resolution: A microwaveoptical/IR synergistic approach," *Int. J. Remote Sens.*, vol. 24, no. 22, pp. 4599–4622, 2003.

[17] P. Kumar, "A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 182–197, Jan. 1999.

[18] R. Gillies, T. Carlson, J. Cui, W. Kustas, and K. Humes, "A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the normalized difference vegetation index (NDVI) and surface radiant temperature," *Int. J. Remote Sens.*, vol. 18, no. 15, pp. 3145–3166, 1997.

[19] T. Carlson, W. Capehart, and R. Gillies, "A new look at the simplified method for remote sensing of daily evapotranspiration," *Remote Sens. Environ.*, vol. 54, no. 2, pp. 161–167, 1995.

[20] O. J. Merlin, A. Chehbouni, R. Panciera, and Y. H. Kerr, "Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3935–3946, 2008.

[21] M. Piles, A. Camps, M. Vall-llossera, I. Corbella, R. Panciera, C. Rudiger *et al.*, "Downscaling SMOS-derived soil moisture using MODIS visible/infrared data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3156–3166, Sep. 2011.

[22] M. Anderson, J. Norman, J. Mecikalski, J. Otkin, and W. Kustas, "A climatological study of evapotranspiration and moisture stress across continental united states based on thermal remote sensing," *J. Geophys. Res.*, vol. 112, no. D10, pp. 2156–2202, 2007.

[23] O. Merlin, J. P. Walker, J. Kalma, E. Kim, J. Hacker, R. Panciera, R. Young, G. Sumerrell, J. Hornbuckle, M. Hafeez, and T. Jackson, "The NAFE'06 data set: Towards soil moisture retrieval at intermediate resolution," *Adv. Water Resour.*, vol. 31, no. 5, pp. 1444–1455, 2008.

[24] M. Drusch, E. F. Wood, and H. Gao, "Observation operators for the direct assimilation of satellite retrieved soil moisture into land surface models," *Geophys. Res. Lett.*, vol. 32, p. L15403, 2005.

[25] R. Reichle and R. Coster, "Bias reduction in short records of satellite soil moisture," *Geophys. Res. Lett.*, vol. 31, no. 19, p. L19501, 2004.

[26] S. Rao, A. Martins, and J. Principe, "Mean shift: An information theoretic perspective," *Pattern Recognit. Lett.*, vol. 30, no. 3, pp. 222–230, 2009.

[27] J. Principe, *Information Theoretic Learning: Renyis Entropy and Kernel Perspectives*. New York, NY, USA: Springer-Verlag, 2010.

[28] S. Chakrabarti, T. Bongiovanni, J. Judge, K. Nagarajan, and I. J. C. Principe, "Downscaling satellite based soil moisture using high resolution remote sensing products and information theory," *IEEE Trans. Geosci. Remote Sens.*, to be published.

[29] A. Monsivais-Huertero, W. Graham, J. Judge, and D. Agrawal, "Effect of simultaneous state-parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation using ENKF," *Adv. Water Resour.*, vol. 33, no. 4, pp. 468–484, 2010.

[30] T. Bongiovanni, K. Nagarajan, J. W. Jones, A. Monsivais-Huertero, and J. Judge, "Improving crop biomass through asynchronous assimilation of LAI and soil moisture during multiple growing seasons of corn," presented at the Fall Meet. Amer. Geophys. Union, San Francisco, CA, USA, Dec. 13–17, 2010.

[31] J. Jones, G. Hoogenboom, C. Poter, K. Boote, W. Hunt, P. Wilkens, U. Singh, A. Gijsman, and J. Ritchie, "The DSSAT cropping system model," *Eur. J. Agron.*, vol. 18, no. 3–4, pp. 235–265, 2003.

[32] A. Gijsman, G. Hoogenboomb, W. J. Partonc, and P. C. Kerridgea, "Modifying DSSAT crop models for low-input agricultural systems using a soil organic matter residue module from CENTURY," *Agron. J.*, vol. 94, no. 3, pp. 462–474, 2002.

[33] E. Sadler, B. Gerwig, D. Evans, W. Busscher, and P. Bauer, "Sit-specific modeling of corn yield in the SE coastal plain." *Agric. Syst.*, vol. 94, no. 3, pp. 462–474, 2000.

[34] K. Nagarajan, C. Krekeler, K. C. Slatton, and W. D. Graham, "A scalable approach to fusing spatiotemporal data to estimate streamflow via a bayesian network," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3720–3732, Oct. 2010.

[35] B. Li, D. Toll, X. Zhan, and B. Cosgrove, "Improving estimated soil moisture fields through assimilation of AMSR-E soil moisture retrievals with an Ensemble Kalman filter and a mass conservation constraint," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 1, pp. 105–119, 2012.

[36] Y. Kaheil, M. K. Gill, M. McKee, L. A. Bastidas, and E. Rosero, "Downscaling and assimilation of surface soil moisture using ground truth measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1375–1384, May 2008.

[37] G. Cawley and N. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *J. Mach. Learn. Res.*, vol. 11, no. 3, pp. 2079–2107, 2010.

[38] C. Tucci and R. T. Clarke, "Environmental issues in the La Plata Basin," *Int. J. Water Resour. Develop.*, vol. 14, no. 2, pp. 157–173, 1998.

[39] J. Oades and A. Waters, "Aggregate hierarchy in soils," *Australian J. Soils Res.*, vol. 29, no. 6, pp. 815–828, 1991.

[40] N. Batjes, "ISRIC-WISE global data set of derived soil properties on a 0.5 by 0.5 degree grid," ISRIC—World Soil Information, Wageningen, The Netherlands, Tech. Rep. 2005/08, 2005.

[41] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2001.

[42] B. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman and Hall, 1986.

[43] A. Renyi, "Some fundamental questions of information theory," *Field Crop Res.*, vol. 2, no. 174, pp. 526–552, 1976.

[44] D. Erdogmus, "Information theoretic learning: Renyi's entropy and its applications to adaptive system training," Ph.D. dissertation, Univ. Florida, Gainesville, FL, USA, 2002.

[45] B. Ruz-Nogueira, K. Booteb, and F. Saua, "Calibration and use of CROP-GRO-soybean model for improving soybean management under rainfed conditions," *Agric. Syst.*, vol. 68, no. 2, pp. 151–173, 2001.

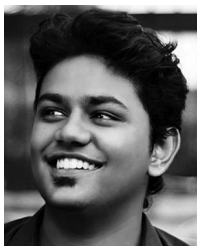
[46] S. Jagtap and J. Jones, "Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production," *Agric. Ecosyst. Environ.*, vol. 93, pp. 73–85, 2002.

[47] F. Sau, K. Boote, and B. Ruz-Nogueiraa, "Evaluation and improvement of CROPGRO-soybean model for a cool environment in galicia, northwest Spain," *Fields Crops Res.*, vol. 61, no. 3, pp. 273–291, 1999.

[48] F. Wang, C. Fraisse, N. Kitchen, and K. Sudduth, "Site-specific evaluation of the CROPGRO-soybean model on Missouri claypan soils," *Agric. Syst.*, vol. 76, no. 3, p. 985–1005, 2003.

[49] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA, USA: MIT Press, 1974.

[50] H. Moradkhani, S. Sorooshina, H. Gupta, and P. Houser, "Dual state-parameter estimation of hydrological models using Ensemble Kalman filter," *Adv. Water Resour.*, vol. 28, no. 2, pp. 135–147, 2005.
[51] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dynam.*, vol. 53, no. 4, pp. 343–367.
[52] L. Zotarelli, N. Zatorre, B. Alves, C. Jantalia, J. Franchini, S. Urquiaga, and R. Boddey, "Influence of no-tillage and frequency of a green manure legume in crop rotations on balancing n outputs and preserving soil organic c stocks," *Field Crop Res.*, vol. 132, pp. 185–195, 2012.
[53] E. Habib, W. Krajewski, and A. Kruger, "Sampling errors of tipping-bucket rain gauge measurements," *J. Hydrol. Eng.*, vol. 6, no. 2, pp. 159–166, 2001.
[54] G. Ciach, "Local random errors in tipping-bucket rain gauge measurements," *J. Atmos. Oceanic Technol.*, vol. 20, pp. 752–759, 2003.
[55] L. F. Alliprandini, C. Abatti, P. Bertagnolli, J. Cavassim, H. Gabe, and A. Kurek, "Understanding soybean maturity groups in brazil: Environment, cultivar classification, and stability," *Crop Sci.*, vol. 49, no. 3, pp. 801–808, 2009.



**Subit Chakrabarti** (S'09) received the B.Tech. degree in applied electronics and instrumentation engineering from Haldia Institute of Technology, Haldia, West Bengal, India, in 2012. He is currently pursuing the Ph.D. degree in electrical and computer engineering at the University of Florida, Gainesville, FL, USA.

His research interests includes applying machine learning techniques for spatial scaling of satellite-based observations for agricultural applications at the Center for Remote Sensing, University of Florida.

**Tara Bongiovanni** received the B.S. and M.E. degrees in agricultural and biological engineering from the University of Florida, Gainesville, FL, USA, in 2009 and 2012, respectively.

She is currently working with the Center of Remote Sensing, University of Florida. Her research interests include modeling crop growth and development and improving modeled estimates of biomass and yield through assimilating remote sensing observations and also Geographic Information Systems (GIS) work for drought forecasting models. She also supervises crop vegetation sampling, data collection, sensor monitoring, and data archiving for the Center.

**Jasmeet Judge** (S'94–M'00–SM'05) received the Ph.D. degree in electrical engineering and atmospheric, oceanic, and space sciences from the University of Michigan, Ann Arbor, MI, USA, in 1999.

She is currently the Director of the Center for Remote Sensing and an Associate Professor in the Agricultural and Biological Engineering Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA. Her research interests include microwave remote sensing applications to terrestrial hydrology for dynamic vegetation; modeling of energy and moisture interactions at the land surface and in the vadose zone; spatial and temporal scaling of remotely sensed observations in heterogenous landscapes; and data assimilation. She is the Chair of the National Academies Standing Committee on Radio Frequencies. She also serves the American Geophysical Union as the past Chair of the Remote Sensing Technical Committee in the Hydrology Section.

Dr. Judge is a Member of the Frequency Allocations in Remote Sensing Technical Committee in the IEEE-GRSS.

**Lincoln Zotarelli** received the M.S. and Ph.D. degrees in agronomy from the Federal University Rural of Rio de Janeiro, Rio de Janeiro, Brazil, in 2000 and 2005, respectively.

He was an Assistant Professor with the Agronomy Department, State University of Londrina, Londrina, Brazil, from 2003 to 2005. He is currently an Assistant Professor with the Horticultural Sciences Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA. His research interests include crop management with emphasis in efficient water and nutrient use in agricultural systems; nitrogen, and carbon cycling in agroecosystems.

**Cimelio Bayer** received the Ph.D. degree in soil science from the Federal University of Rio Grade do Sul, Porto Alegre, Brazil, in 1996.

He is a Professor of Management of Tropical Soils with Federal University of Rio Grande do Sul, Southern Brazil. He is currently the Coordinator of the Agronomy Committee of the National Council for Scientific and Technologic Development (CNPq) and the Vice Coordinator of the Agrarian Committee of Rio Grande do Sul Research Foundation (Fapergs). He is responsible for two 30-year old field experiments in Southern Brazil and for carrying out collaborative research projects with scientists in Brazil, South America, Canada, and in the European Community. His research interests include soil organic matter dynamics under conventional and alternative agricultural practices, and the effects on physical, chemical, and biological qualities of tropical and subtropical soils.