Disaggregation of Remotely Sensed Soil Moisture in Heterogeneous Landscapes Using Holistic Structure-Based Models

Subit Chakrabarti, *Student Member, IEEE*, Jasmeet Judge, *Senior Member, IEEE*, Tara Bongiovanni, Anand Rangarajan, *Member, IEEE*, and Sanjay Ranka, *Fellow, IEEE*

***Abstract*—In this paper, a novel machine learning algorithm is presented for disaggregation of satellite soil moisture (SM) based on self-regularized regressive models (SRRMs) using high-resolution correlated information from auxiliary sources. It includes regularized clustering that assigns soft memberships to each pixel at a fine scale followed by a kernel regression that computes the value of the desired variable at all pixels. Coarse-scale remotely sensed SM was disaggregated from 10 to 1 km using land cover (LC), precipitation, land surface temperature, leaf area index, and *in situ* observations of SM. This algorithm was evaluated using multiscale synthetic observations in NC Florida for heterogeneous agricultural LCs. It was found that the rmse for 96% of the pixels was less than 0.02 m^3/m^3. The clusters generated represented the data well and reduced the rmse by up to 40% during periods of high heterogeneity in LC and meteorological conditions. The Kullback–Leibler divergence (KLD) between the true SM and the disaggregated estimates is close to zero, for both vegetated and bare-soil LCs. The disaggregated estimates were compared with those generated by the principle of relevant information (PRI) method. The rmse for the PRI disaggregated estimates is higher than the rmse for the SRRM on each day of the season. The KLD of the disaggregated estimates generated by the SRRM is at least four orders of magnitude lower than those for the PRI disaggregated estimates, whereas the computational time needed was reduced by three times. The results indicate that the SRRM can be used for disaggregating SM with complex nonlinear correlations on a grid with high accuracy.**

***Index Terms*—Clustering, disaggregation, kernel regression, microwave remote sensing, multispectral remote sensing, soil moisture (SM).**

Manuscript received September 14, 2015; revised January 9, 2016; accepted February 29, 2016. Date of publication April 14, 2016; date of current version June 1, 2016. This work was supported in part by the National Aeronautics and Space Administration's Terrestrial Hydrology Program under Grant NNX13AD04G.

S. Chakrabarti is with the Department of Electrical Engineering, University of Florida, Gainesville, FL 32601 USA, with the Center of Remote Sensing, Department of Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, Florida, FL 32611-0570 USA, and also with the Computational Neuro Engineering Laboratory, University of Florida, Gainesville, FL 32611 USA (e-mail: subitc@ufl.edu).

J. Judge and T. Bongiovanni are with the Center for Remote Sensing, Department of Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, FL 32611-0570 USA (e-mail: jasmeet@ufl.edu; tigert@ufl.edu).

A. Rangarajan and S. Ranka are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: anand@cise.ufl.edu; ranka@cise.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/TGRS.2016.2547389

I. Introduction

SOIL moisture (SM) is a key governing factor in surface and subsurface hydrological and agricultural models as it regulates land–atmosphere interactions. It has also been recognized as an essential climate variable by the Global Climate Observing System [1]. Representational models of weather [2]–[4], crop growth[5], ecosystem and carbon cycle processes [6], [7], dust generation [8], trace gas fluxes [9], and agricultural drought [10], [11] require SM data at a fine spatial resolution. Recent satellite missions, including the European Space Agency Soil Moisture and Ocean Salinity (SMOS) and the National Aeronautics and Space Administration Soil Moisture Active Passive (SMAP) missions [12], provide for SM retrievals at unprecedented spatial resolutions of tens of kilometers every 2–3 days, with worldwide coverage. However, models simulating physical processes for agricultural regions need SM at even finer scales of 1 km [11]. Disaggregation addresses this discrepancy in scales by generating local fine-resolution data from coarse-resolution data obtained from satellites.

Most of the disaggregation techniques broadly fall into three approaches. The first approach is based on the assumption that spatial disaggregation follows a known hierarchical model such as fractal interpolation, power law, or temporal persistence across scales. Methods using this approach usually assume static vegetation and micrometeorology for a given area, due to the difficulties associated with parametrizing weather and land cover (LC) data across temporal and spatial scales in such models. However, the static assumption in this approach introduces large errors in realistic applications. The second approach uses empirical models based on statistical and geostatistical methods, such as regression, cokriging and block kriging, and fractal interpolation. The third approach employs statistical models based on the triangle method [13]–[15] to extrapolate the dependent data within the hypothetical triangle formed by the observed data. The robustness of the statistical methods over heterogeneous vegetation and weather conditions remain mostly untested. Treating each pixel as a sample instead of using spatial information to regularize the disaggregation results in salt-and-pepper noise due to spatial autocorrelation [16]. Moreover, these approaches use second-order metrics, which do not leverage all the information in the data that is necessary in a highly nonlinear regression problem such as disaggregation [17].

A recently implemented disaggregation algorithm [18] based on the principle of relevant information (PRI) addresses the given inadequacies by utilizing the full probability density function (pdf) of a set of training observations, rather than second-order moments, to approximate a transformation function that relates micrometeorological data recorded in a region to *in situ* SM. It uses the transformation function to generate an initial set of SM values for the rest of the data set. The disaggregated SM is obtained by iterating between the coarse-scale SM values and the initial SM values using an information-theoretic (IT) cost function. The PRI method was compared with the widely used disaggregation algorithm based on a second-order regression using the triangle method [14]. It was found to have lower disaggregation errors, particularly for complex noise models added to the coarse-resolution SM. Notably, the Kullback–Leibler distance between the true and disaggregated SM was 50% lower for the PRI method, compared with the triangle method. This is because methods based on the second-order triangular or quadrilateral regressions do not have separate steps for error-bias and error-variance controls and rely on the data being well posed to achieve a balance between error bias and error variance. Although the PRI method results in low disaggregation errors, training a fully Bayesian transformation function is computationally intensive. Additionally, it requires a comprehensive training set for the initial estimate of the multidimensional pdf to converge. In this paper, a self-regularized regressive model (SRRM) is used to disaggregate SM. It is expected to be less computationally intensive as it uses auxiliary features correlated to SM to perform clustering of pixels and subsequently trains a single model for each cluster. Furthermore, it requires fewer samples for training.

The goal of this paper is to develop and implement a novel machine learning algorithm to disaggregate coarse-scale remotely sensed SM using auxiliary fine-scale data. The primary objectives of this paper are to: 1) develop an algorithm to identify contiguous regions of similarity in gridded images and, subsequently, for each region, use kernel regression to estimate a disaggregation model for each region; 2) implement this algorithm to estimate SM at 1 km using SM at 10 km and other spatially correlated variables in the region such as land surface temperature (LST), leaf area index (LAI), LC, and precipitation (PPT); and 3) evaluate the SRRM-based methodology and compare it with the PRI method using a synthetic data set.

Section II describes the theoretical details of the disaggregation framework based on SRRMs and provides a brief description of the PRI algorithm for disaggregation. Section III illustrates the steps for the implementation of the SRRM and presents the disaggregation results for SM at 1 km, and Section IV summarizes the important results, concludes this paper, and outlines the scope for future studies.

II. Disaggregation Framework

Disaggregation is an ill-conditioned problem that is limited physically by the convolution of the point-spread function of the imaging system. This constrains the generation of fine-scale data from coarse-scale data. Additional spatially correlated information is needed to regularize the fine-scale estimates. Methods that use regression to bridge the difference in scales have to use regularization to address the multiplicity of solutions. The SRRM addresses this problem by using a clustering algorithm to create a number of regions of similarity, which, subsequently, are used in a kernel regression framework. This is described in more detail in the following. Using spatial regions or dynamic conglomerations of pixels to generate models instead of treating each pixel in a sample-based method also reduces the effect of spatial autocorrelation on the disaggregated estimates.

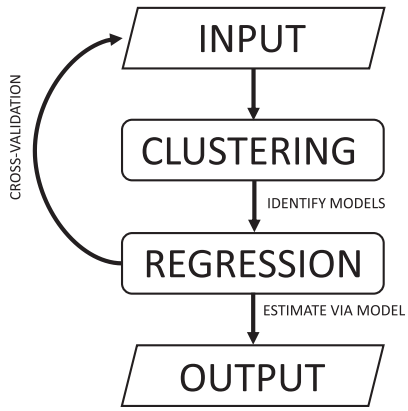

Fig. 1. Flowchart of the SRRM-based algorithm.

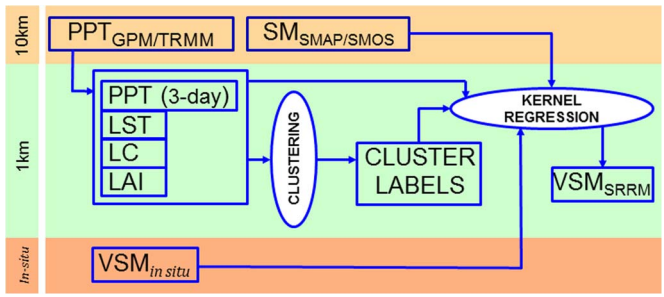

Fig. 2. Flow diagram of the self-regularized kernel regression models.

A. Disaggregation Framework Based on SRRMs

In this paper, contiguous regions are identified in multidimensional correlated data using clustering, and subsequently, a regression model is trained for each cluster for disaggregation. The membership vector of every pixel to a region, and thus to a model, is soft and constrained to sum to one across the space of models. The models themselves are trained using a kernel-regression-based method. It is a novel way to account for correlated features using algorithms that require an independent and identical distribution assumption [16]. Fig. 1 shows a flow diagram of the algorithm for generating disaggregated estimates. The overall organization and the data sets involved is shown in Fig. 2. The two steps of the algorithm include clustering and kernel regression, as follows.

1) IT Clustering Based on the Cauchy–Schwarz Distance: Commonly used clustering methods, such as the K-means [19], assume hyperspherical or hyperelliptical clusters [20]. With gridded remotely sensed data, prior assumptions about cluster

shapes are not advisable and lead to noise in the clustering result, as shown in Fig. 3(a). Instead, in the IT clustering method, the generalized proximity regions are identified using a regularized variant of a clustering method based on information theory [20]. The clusters are constructed using the pdfs of the data, resulting in clusters that are representative of the input data, as shown in Fig. 3(b). For any two vectors $\mathbf{x}$ and $\mathbf{y}$, the Cauchy–Schwarz inequality is

$$-\log\left(\frac{|\langle \mathbf{x},\mathbf{y}\rangle|}{\sqrt{\|\mathbf{x}\|^2\|\mathbf{y}\|^2}}\right)\geq 0 \tag{1}$$

where $\langle \mathbf{x},\mathbf{y}\rangle$ is the inner product of vectors $\mathbf{x}$ and $\mathbf{y}$. For pdfs $p(x)$ and $q(x)$, the inner product is defined as $\langle p,q\rangle = \int p(x)q(x)dx$ over the support for the distributions p and q. Then, the Cauchy–Schwarz inequality in a metric space spanned by the pdf is

$$-\log\left(\frac{\left|\int p(x)q(x)dx\right|}{\sqrt{\int p^2(x)dx\int q^2(x)dx}}\right)\geq 0. \tag{2}$$

If $p(x)$ is calculated using pixels lying in cluster C_1 and $q(x)$ is calculated using pixels lying in cluster C_2, the maximum separation is obtained between clusters when the left-hand side of (2), the Cauchy–Schwarz distance D_{CS}, is maximized. Since the logarithm is a monotonically increasing function, only the argument of the logarithm in $D_{\mathrm{CS}} = -\log J_{\mathrm{CS}}(p,q)$ can be equivalently minimized using gradient-descent-based optimization. An estimator $\hat{J}_{\mathrm{CS}}$ of $J_{\mathrm{CS}}(p,q)$ can be constructed from data samples and extended to the case of multiple clusters by using a membership vector, i.e.,

$$\hat{J}_{\mathrm{CS}}(\mathbf{m}_1,\ldots,\mathbf{m}_N) = \frac{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1-\mathbf{m}_i^T\mathbf{m}_j\right)G_{\sigma\sqrt{2}}(\mathbf{x}_i,\mathbf{x}_j)}{\sqrt{\prod_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}m_{ik}m_{jk}G_{\sigma\sqrt{2}}(\mathbf{x}_i,\mathbf{x}_j)}} \tag{3}$$

where $\mathbf{m}_i$ is a soft K-dimensional vector, where the kth element expresses the degree of membership to the kth cluster. K is the total number of clusters that has to be supplied as input. $G_{\sigma\sqrt{2}}(\cdot,\cdot)$ is derived from the convolution of two Gaussian kernels, which is defined as $G_{\sigma\sqrt{2}}(\mathbf{x}_i,\mathbf{x}_j) = \exp(-(\|\mathbf{x}_i - \mathbf{x}_j\|_2^2/2\sigma^2))$. A regularized version can be used as an objective function of clustering, as shown in the following:

$$\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}(\mathbf{m}_1,\ldots,\mathbf{m}_N) = \frac{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1-\mathbf{m}_i^T\mathbf{m}_j\right)G_{\sigma\sqrt{2}}(\mathbf{x}_i,\mathbf{x}_j)}{\sqrt{\prod_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}m_{ik}m_{jk}G_{\sigma\sqrt{2}}(\mathbf{x}_i,\mathbf{x}_j)}} - \psi\sum_{i=1}^{N}\sum_{k=1}^{K}m_{ik}\log(m_{ik}). \tag{4}$$

The second term of the objective function is an estimate of the Shannon entropy of the membership vectors and serves to regularize the membership vectors such that the model selection is sufficiently sparse. Getting the correct membership

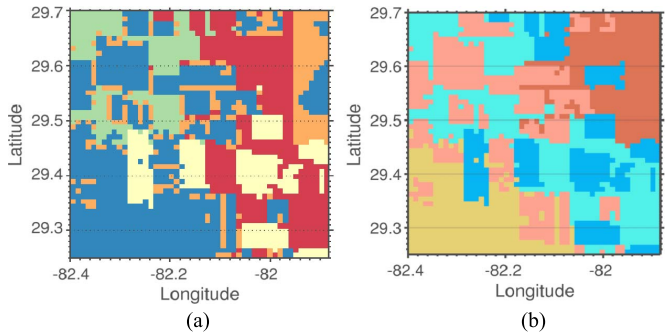

Fig. 3. Clustering result obtained from (a) D_{CS}-based clustering algorithm and (b) K-means clustering algorithm.

vector then is equivalent to solving this constrained optimizatio problem, i.e.,

$$\min_{\mathbf{m}_1,\ldots,\mathbf{m}_N}\ \hat{J}_{\mathrm{CS}}^{\mathrm{REG}}(\mathbf{m}_1,\ldots,\mathbf{m}_N) \quad \text{subject to}\ \ \mathbf{m}_j^T\mathbf{1}_{K\times 1}-1=0,\quad j=1,\ldots,N \tag{5}$$

where $\mathbf{1}_{K\times 1}$ is a $K\times 1$ vector whose elements are all one. Consider $m_{ik}=v_{ik}^2, k=1,\ldots,K$, which corresponds to a form that can be optimized by using Lagrange multipliers. The Lagrangian can be expressed as

$$L = \hat{J}_{\mathrm{CS}}^{\mathrm{REG}}(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_N) + \sum_{i=1}^{N}\lambda_i\left(\mathbf{v}_i^T\mathbf{v}_i-1\right). \tag{6}$$

The optimization problem (6) amounts to adjusting vectors $\mathbf{v}_i, i=1,\ldots,N$ such that

$$\frac{\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}}{\partial\mathbf{v}_i} = \left(\frac{\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}}{\partial\mathbf{m}_i}^T\frac{\partial\mathbf{m}_i}{\partial\mathbf{v}_i}\right)^T = \Gamma\frac{\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}}{\partial\mathbf{m}_i}\to 0 \tag{7}$$

where $\Gamma = \mathrm{diag}(2\sqrt{m_{i1}},\ldots,2\sqrt{m_{iK}})$ is the magnitude normalizing factor. The memberships are forced to be positive by adding a constant of small magnitude $\alpha\sim 0.05$ to all elements of Γ. The Lagrange multipliers then, after constructing the necessary Lagrange function is given by

$$\lambda_i = \frac{1}{2}\sqrt{\frac{\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}}{\partial\mathbf{v}_i}^T\frac{\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}}{\partial\mathbf{v}_i}}. \tag{8}$$

The updated vector for the next iteration is

$$\mathbf{v}_i^+ = -\frac{1}{2\lambda_i}\frac{\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}}{\partial\mathbf{v}_i}. \tag{9}$$

The square of the membership vectors are initialized as $\mathbf{v}_i = |\mathcal{N}(0;\gamma^2\mathbf{I})|$, where $\mathcal{N}$ denotes the Gaussian distribution, and γ is a very small number.

Stochastic Approximation of the Gradient and Computational Complexity: If $\hat{J}_{\mathrm{CS}}$ is represented as (U/V), then the gradient of $\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}$ can be calculated as

$$\frac{\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}}{\partial\mathbf{m}_i} = \frac{V\frac{\delta U}{\partial\mathbf{m}_i}-U\frac{\partial V}{\partial\mathbf{m}_i}}{V^2} - \psi\sum_{k=1}^{K}(1+\log(m_{ik})) \tag{10}$$

where U and V are defined as

$$U = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1 - \mathbf{m}_i^T\mathbf{m}_j\right) G_{\sqrt{2}\sigma}(\mathbf{x}_i, \mathbf{x}_j) \quad \text{and} \quad V = \sqrt{\prod_{k=1}^{K} v_k} \tag{11}$$

and the gradients of U and V are defined as

$$\begin{aligned}\frac{\partial U}{\partial \mathbf{m}_j} &= -\sum_{j=1}^{N} \mathbf{m}_j G_{\sigma\sqrt{2}}(\mathbf{x}_i, \mathbf{x}_j) \\ \frac{\partial V}{\partial \mathbf{m}_i} &= \frac{1}{2}\sum_{k'=1}^{K} \sqrt{\frac{\prod_{\substack{k=1\\k\neq k'}}^{K} v_k}{v_{k'}}} \frac{\partial v_{k'}}{\partial \mathbf{m}_i}\end{aligned} \tag{12}$$

where $v_k = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{m}_{ik}\mathbf{m}_{jk} G_{\sqrt{2}\sigma}(\mathbf{x}_i, \mathbf{x}_j)$ and $\partial v_{k'}/\partial m_i = [0, \ldots, 2\sum_{j=1}^{N} \mathbf{m}_j(k') G_{\sqrt{2}\sigma}(\mathbf{x}_i, \mathbf{x}_j), \ldots, 0]^T$.

Kernel Annealing: The objective function in (4) has local minima that can inhibit the performance of this algorithm. To ensure that the clustering solution is global and not just local minimum, the kernel width is gradually decreased in this algorithm over the course of iterations. The initial value of the kernel is chosen according to the Silverman's rule of thumb [21] given by

$$\sigma_{\mathrm{SIL}} = \sigma_X \left(4N^{-1}(2d+1)^{-1}\right)^{\frac{1}{d+4}} \tag{13}$$

where d is the dimensionality of the data, N is the number of samples, and $\sigma_X^2 = d^{-1}\sum_i \sum_{X_{ii}}$, with $\sum_{X_{ii}}$ being the diagonal values of the sample covariance matrix. The lower value of the kernel size is set to $\sigma_{\mathrm{LOW}} = (\sigma_{\mathrm{SIL}}/4)$. Thus, the annealing rate is

$$r = \frac{\sigma_{\mathrm{SIL}} - \sigma_{\mathrm{LOW}}}{N_{\mathrm{TOT}}} = \frac{3\sigma_{\mathrm{SIL}}}{4N_{\mathrm{TOT}}}. \tag{14}$$

2) Regularized Kernel Regression: A kernel-based regression technique that uses a training set of pixels and fits a function to it, by minimizing the representational error, is used to generate the disaggregated estimates. Ridge regression [22] is a parametric regression technique that adds a scaled regularizing term to the cost function. This improves the stability of the regression as the added L_2-norm term in the cost function results in smaller eigenvalues. The cost function is

$$\mathcal{E}(\mathbf{w}, \mathbf{x}) = \frac{1}{2}\sum_i \left(y_i - w^{\mathbf{T}}\mathbf{x}_i\right)^2 + \frac{1}{2}\mu\|\mathbf{w}\|^2. \tag{15}$$

The weights can be calculated by differentiating the error cost function with respect to the weights and setting it to zero, i.e.,

$$\frac{\partial \mathcal{E}}{\partial \mathbf{w}} = 0 \implies \mathbf{w} = \left(\sum_i \mathbf{x}_i\mathbf{x}_i^{\mathbf{T}} + \mu\mathbf{I}\right)^{-1}\left(\sum_i y_i\mathbf{x}_j\right). \tag{16}$$

For computation in a reproducing kernel Hilbert space, then, the inner products can be replaced with a kernel evaluation. Let $\mathcal{H}$ be a Hilbert space with an inner product metric $\langle\cdot,\cdot\rangle_{\mathcal{H}}$. Then, according to the representer theorem, a kernel function $\kappa(\mathbf{x}, \mathbf{y})$ exists on $\mathbb{R}^{\mathrm{N}} \times \mathbb{R}^{\mathrm{N}}$ such that $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathcal{H}} = \kappa(\mathbf{x}, \mathbf{y})$. Now, if $\Phi : \mathbb{R}^{\mathrm{N}} \to \mathbb{R}^{\mathrm{N}}$ is a mapping that transforms the feature vector in the original vector space to $\mathcal{H}$, then the weights can de redefined as

$$\mathbf{w} = \left(\mu\mathbf{I}_D + \Phi\Phi^T\right)^{-1}\Phi\mathbf{y} \tag{17}$$

where D is the dimension of the feature space. The dimension of the feature space is not well defined in many cases; therefore, the weights can be rewritten using the identity $(A^{-1} + B^TC^{-1}B)^{-1}B^TC^{-1} = AB^T(BAB^T + C)^{-1}$, i.e.,

$$\mathbf{w} = \Phi(\mu\mathbf{I}_N + \Phi^T\Phi)^{-1}\mathbf{y}. \tag{18}$$

The weight vector $\mathbf{w}$ can be calculated using a training set of observations, where $\mathbf{y}$ is known. This can then be used to calculate the estimated value for a new data point $\mathbf{x}'$, i.e.,

$$\begin{aligned}\hat{y} &= \mathbf{w}^T\Phi(\mathbf{x}') \\ &= \mathbf{y}(\mu\mathbf{I}_N + \Phi^T\Phi)^{-1}\Phi^T\Phi\left((x')\right) \\ &= \underbrace{\mathbf{y}(\mu\mathbf{I}_N + \mathbf{K})^{-1}}_{\mathbf{w}}\kappa(\mathbf{x}, \mathbf{x}')\end{aligned} \tag{19}$$

where $\mathbf{K}$ is the Gram matrix of inner products of all the training data points. This does not address the constant that must be present in the regression. To solve this problem, the feature vector is augmented by adding a constant feature 1 to all samples.

3) Algorithm Summary and Computational Complexity: The SRRM disaggregation is summarized and shown in Algorithm 1. A tenfold cross-validation was used to determine the number of clusters N, the kernel size for the clustering ψ, and the regularization weight for the regression μ. The performance of the algorithm was less sensitive to the kernel size for regression than the other parameters and was set to the standard deviation of $\mathbf{y}$ at coarse scale.

Algorithm 1 Disaggregation Using SRRMs

Require: Initialize membership vectors, $\mathbf{v}_i \leftarrow |\mathcal{N}(0; \gamma^2\mathbf{I})|$ and number of clusters, N for each day of the data set. N_{DAYS} is the total number of days.
for $i = 0$ to N_{DAYS} **do**
  Step 1: Clustering
  for $i = 1$ to 30 **do**
    Calculate $\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}$ and $\partial\hat{J}_{\mathrm{CS}}^{\mathrm{REG}}/\partial\mathbf{m}_i$ according to Equations (4) and (10).
    Update λ_i and $\mathbf{v}_i^+$ according to Equations (8) and (9).
  end for
  Step 2: Kernel Regression
  Calculate $\mathbf{w}$ according to Equation (18) using the training set.
  Estimate the disaggregated observations, $\hat{\mathbf{y}}$ for the test set using Equation (19).
  Run 10-fold cross-validation for the values of N and the cross-validation constants ψ and μ.
end for

The complexity of the D_{CS} based clustering algorithm is $\mathcal{O}(N^2)$ for each iteration. For good convergence, 30 iterations are needed. This is much lower than the dimensionality of the data set and does not affect the complexity of the algorithm. To reduce the computational load, a stochastic sampling method is used. For this, the gradient is approximated by using M samples out of all N. The complexity then becomes $\mathcal{O}(MN)$ $(M \ll N)$ per iteration. M can be much lesser than N, and the results are comparable to the original method, taking a fraction of the time. The average complexity of the ridge regression method is $\mathcal{O}(N^3)$ [23].

B. PRI Framework

The disaggregation methodology using PRI includes a transformation process to obtain a probabilistic relationship between the variable to be disaggregated $\mathbf{y}$ at 1 km using auxiliary information $\mathbf{X}$ at the same scale. A discrete formulation of the Bayes rule is used to estimate $\mathbf{y}_{\mathrm{initial}}$ at fine resolution, as given in (20), wherein $\mathbf{y}^i_{\mathrm{train}}$ is discretized into k classes, $i \in [1, k]$, and $\mathbf{x}^{i_1}_{j,\mathrm{train}}$ is discretized into k_j classes in $i_1 \in [1, k_j]$, where j indexes the individual variables that comprise $\mathbf{X}$, m, i.e.,

$$p\left(\mathbf{y}^{i_1}_{\mathrm{initial}} \left| \mathbf{X}^{i_1}_{\mathrm{train}}\right.\right) = \frac{p\left(\mathbf{X}^{i_1}_{\mathrm{train}} \left| \mathbf{y}^{i}_{\mathrm{train}}\right.\right) p\left(\mathbf{y}^{i}_{\mathrm{train}}\right)}{p\left(\mathbf{X}^{i_1}_{\mathrm{train}}\right)}$$

$$\mathbf{y}^{i}_{\mathrm{initial}} = \arg\max_{\mathbf{y}^{i}_{\mathrm{train}}} \frac{p\left(\mathbf{X}^{i_1}_{\mathrm{train}}\right) p\left(\mathbf{y}^{i}_{\mathrm{train}}\right)}{p\left(\mathbf{X}^{i_1}_{\mathrm{train}}\right)}$$

$$p\left(\mathbf{X}^{i_1}_{\mathrm{train}}\right) = \sum_{i=1}^{k} p\left(\mathbf{X}^{i_1}_{\mathrm{train}} | \mathbf{y}^{i}_{\mathrm{train}}\right) p\left(\mathbf{y}^{i}_{\mathrm{train}}\right). \quad (20)$$

In the second step, $\mathbf{y}_{\mathrm{initial}}$ is merged with the observations at the coarser resolutions $\mathbf{y}_{\mathrm{coarse}}$ to obtain improved estimates at fine resolution, i.e.,

$$\arg\max_{\mathbf{m}} L(\mathbf{m}) = H(\mathbf{m}) + \beta \mathrm{KL}\left(p_{\mathbf{m}} \| p_{\mathbf{y}_{\mathrm{initial}}}\right) \quad (21)$$

where $I(\mathbf{m})$ is the cost function, $p_{\mathbf{y}_{\mathrm{INITIAL}}}$ is the pdf of the original data, and $p_{\mathbf{m}}$ is the pdf at each iteration. $I(\mathbf{m})$ is the entropy, and KL is the KLD. $\mathbf{m}$ is initialized to $\mathbf{y}_{\mathrm{COARSE}}$ at the first iteration. The β is a user-defined weighting parameter that balances the redundancy and information preservation in $I(\mathbf{m})$. As the value of β increases, the cost function gives more emphasis to KL, thus preserving more information about the data at the cost of extremely high redundancy reduction. In this paper, an intermediate value of $\beta = 2$ was chosen so that the PRI image would approximate the mean level of $\mathbf{y}$ at coarse scales but will also embed the level of detail provided by the initial estimates of $\mathbf{y}$ at 1 km, to obtain morphed estimates of $\mathbf{y}$ at 1 km. The computational complexity of the PRI algorithm is given as $\mathcal{O}(N^3 \prod k_j)$, where k_j are the number of bins used to estimate the pdf of the features for the transformation function. A detailed description of the PRI algorithm can be found in [18].

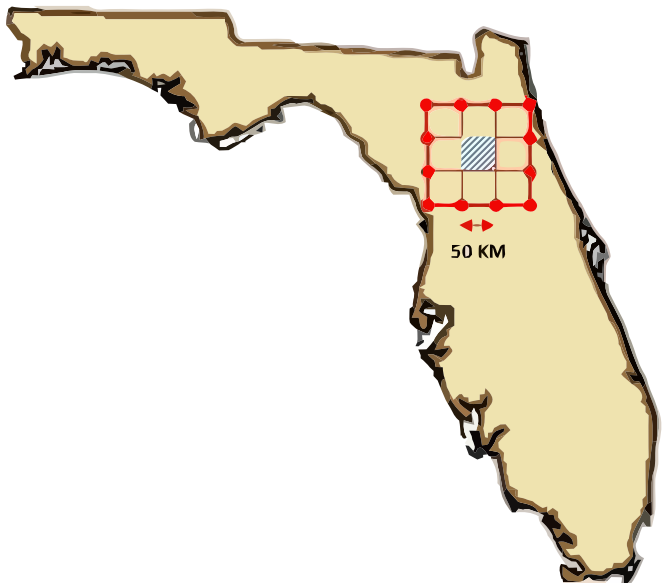

Fig. 4. Study region in North Central Florida. LSP-DSSAT-MB simulations were performed over the shaded 50×50 km^2 region.

III. Experimental Description and Results

A. Multiscale Synthetic Data Set

The proposed algorithm for disaggregation was tested using data generated by a simulation framework consisting of the Land Surface Process (LSP) model and the Decision Support System for Agrotechnology Transfer (DSSAT) model, described in [24]. A 50×50 km^2 region, which is equivalent to approximately 25 SMAP pixels at 9-km spatial resolution, was chosen in North Central Florida (see Fig. 4) for the simulations. The region encompassed the University of Florida's Institute of Food and Agricultural Sciences Plant Science Research and Education Unit, Citra, FL, USA, where a series of season-long field experiments, called the Microwave, Water and Energy Balance Experiments (MicroWEXs), have been conducted for various agricultural LCs over the last decade [25]–[27]. Simulated observations of LST and LAI were generated at 200 m for a period of one year, from January 1, 2007 through December 31, 2007. Topographic features, such as slope, were not considered in this paper because the region is typically characterized by flat and smooth terrains with no run-off due to soils with high sand content. The soil properties were assumed constant over the study region.

Fifteen-minute observations of PPT, relative humidity, air temperature, downwelling solar radiation, and wind speed were obtained from eight Florida Automated Weather Network stations [28] located within the study region (see Fig. 4). The observations were spatially interpolated using splines to generate the meteorological forcings at 200 m. Long-wave radiation was estimated following [29].

The model simulations were performed over each agricultural field rather than all the pixels, to reduce computation time. Based upon LC information at 200 m, contiguous and homogeneous regions of sweet corn and cotton were identified, as shown in Fig. 5. A realization of the LSP-DSSAT model was

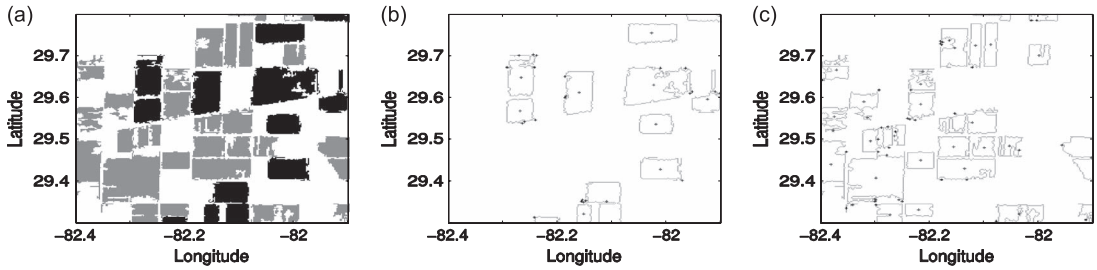

Fig. 5. (a) LC at 200 m during cotton and corn seasons. White, gray, and black shades represent bare-soil, cotton, and sweet-corn regions, respectively. Homogeneous crop fields along with centers for (b) sweet corn and (c) cotton. See [24].

TABLE I
PLANTING AND HARVEST DATES FOR SWEET CORN AND COTTON DURING THE 2007 GROWING SEASON

Crop	Planting DoY	Harvest DoY
Sweet Corn	61	139
	183	261
Cotton	153	332

used to simulate LST, LAI, and PPT at the centroid of each homogeneous region, using the corresponding crop module within DSSAT. The model simulations were performed using the 200-m forcing at the centroid, as shown in Fig. 5. Linear averaging is typically sufficient to illustrate the effects of resolution degradation [30]. The model simulations at 200 m were spatially averaged to obtain PPT, LST, LAI, SM, and T_B at 1 and 10 km. The SM obtained at 1 km was divided into the training and test sets that were used as truth to evaluate the disaggregation methodology and serve as simulated "*in situ*" measurements to train the algorithm, respectively. PPT, LAI, and LST are typically chosen due to their high correlations with SM [14], [18], [31]. Other geophysical descriptors, such as slope and soil texture, were not used in this paper because of their limited utility in a flat and primarily sandy region, such as that in North Central Florida. To simulate rain-fed systems, all the water input from both PPT and irrigation were combined together, and the "PPT" in this paper represents these combined values.

B. Disaggregation Framework Based on SRRM

The simulation period, from January 1 (DoY 1) to December 31 (DoY 365), 2007, consisted of two growing seasons of sweet corn and one season of cotton, as shown in Table I. The LST, PPT, and LAI observations at 1 km were obtained by adding white Gaussian noise to account for satellite observation errors, instrument measurement errors, and micrometeorological variability, following [32]–[34]. Errors with zero mean and standard deviations of 5 K, 1 mm/h, 0.03 m^3/m^3, and 0.1 for LST, PPT, SM, and LAI, respectively, were added to the values at 10 km.

The SRRM uses LST, three-day PPT, LAI, LC at 1 km, and SM at 10 km every three days as input. In the first step, the IT cost function described in Section II-A is used for clustering using the inputs at 1 km and the x and y coordinates of each pixel scaled to a range of 0 and 1. This step of the algorithm uses two parameters: the number of clusters N and a regularization constant μ. Both the number of clusters and the regularization constant is determined by cross-validating against the absolute mean error in SM at the end of the second step for each day.

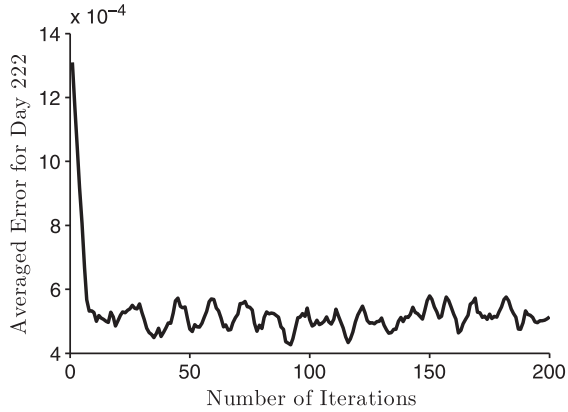

Fig. 6. RMSE in disaggregated SM at 1 km versus number of iterations of the D_{CS} clustering algorithm.

The optimal number of iterations that produced a usable clustering result was determined by the minimum rmse for a day when both the LC and micrometeorological conditions were heterogeneous, i.e., DoY 222, providing the worst-case scenario for convergence of the clustering algorithm. At the end of this step, each pixel has a vector of N numbers, $(m_1, m_2, \ldots, m_N)$ that sum to 1 describing its membership to each of the N clusters. Fig. 6 shows the spatially averaged rmse between disaggregated SM and the observations at 1 km on DoY 222 for different iterations of the clustering algorithms. All parameters, except the number of clusters, were cross-validated for each individual iteration. The number of clusters was cross-validated once, using 50 iterations of the clustering algorithm. For the cross-validation, the training set was randomly divided into ten equal parts. Nine parts were used for training, and one part was used for evaluating the algorithm. This methodology, known as tenfold cross-validation, is repeated ten times with different randomly selected partitions to approximate the average errors that the SRRM would incur. The error oscillates with a mean amplitude of 1.2×10^{-4} m^3/m^3 after 30 iterations. In this paper, 30 iterations of the clustering algorithm are used.

In the second step, N models, $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_N$ are developed using LST, three-day PPT, LAI, LC, SM at 1 km, and SM at 10 km as inputs to the regularized kernel regression algorithm described in Section II-A, using training set. The

TABLE II
DAYS SELECTED FOR EVALUATING SRRM ESTIMATES. THESE DAYS CAPTURE VARIABILITY IN PPT/IRRIGATION AND LC

DoY	PPT	LC
39	Dry	Bare
135	Dry, Irrigated	Sweet Corn
156	Wet	Cotton
222	Dry, Irrigated	Sweet Corn and Cotton
354	Wet	Bare

training set was consisted of randomly selected 33% of the pixels or 500 out of the 2500 pixels that make up the region. The remaining pixels were used as the test set. The hard membership of each pixel i for model development purposes is determined by the maximum value in its membership vector $\mathbf{m}^i = (m_1^i, m_2^i, \ldots, m_N^i)$. The disaggregated value of SM is computed for each point in the test, represented as a vector, $\mathbf{x}_i' = (\mathrm{LST}_i^{1\,\mathrm{km}}, \mathrm{PPT}_i^{1\,\mathrm{km}}, \mathrm{LAI}_i^{1\,\mathrm{km}}, \mathrm{LC}_i^{1\,\mathrm{km}}, \mathrm{SM}_i^{10\,\mathrm{km}})$ by

$$\mathrm{SM}_i^{1\,\mathrm{km}} = \mathbf{m}^T \cdot \left(\hat{f}_1\left(\mathbf{x}_i'\right), \hat{f}_2\left(\mathbf{x}_i'\right), \ldots, \hat{f}_N\left(\mathbf{x}_i'\right)\right). \quad (22)$$

The SRRM is evaluated using the rmse and standard deviation of the errors over the entire season. The rmse over for the entire time period is assessed for each LC. Moreover, the disaggregated SM is compared with the true SM. To evaluate how close the density function of the disaggregated estimates is to the density function of the true SM, the KLD between the density of the estimated observations and the true SM is calculated for different LC's over the season. The KLD is a member of the class of well-known f-divergences that convey distances in probability space. Any other f-divergence such as the Hellinger distance or χ^2-distance can also be used. A sensitivity analysis was conducted to determine how each auxiliary variable separately contributed to errors in downscaled SM. A single auxiliary variable, LST, LAI, or PPT, was allowed to vary for each LC, whereas the others were set to their mean values. The relative rmses RMSE_r, $\mathrm{RMSE}_r = \Delta\mathrm{RMSE}/\mathrm{RMSE}_{\mathrm{org}}$, were investigated, where ΔRMSE is the change in rmse when a single auxiliary variable is used compared with when all the auxiliary variables are used $(\mathrm{RMSE}_{\mathrm{org}})$ for each day in 2007. The daily RMSE_r averaged over each LC, bare soil, corn, and cotton in 2007 is also studied.

In addition, five days were selected from the season to understand the effect of the heterogeneity in inputs on the error in disaggregated SM. Variabilities in PPT, ranging from uniformly wet to uniformly dry, and in LC, ranging from bare soil to vegetated with both cotton and sweet corn, were used as criteria for selecting the days, as shown in Table II. Quantitative analyses of spatial variations in SM observed under dynamic vegetation and heterogeneous LC conditions provide an index of dynamic errors that can be expected. The utility of using multiple models in the region, i.e., one model for each cluster, was also investigated by comparing the disaggregation results to when the entire data set is considered as a single cluster and only one model is used for disaggregation, on DoY 222 of the study.

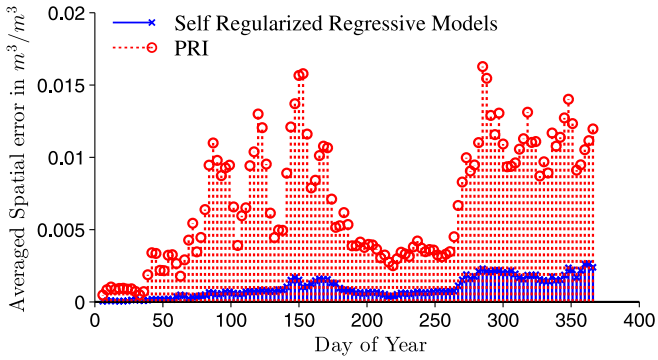

Fig. 7. Spatially averaged rmse in disaggregated SM at 1 km for each day of the year in the simulation period using the SRRM and PRI method.

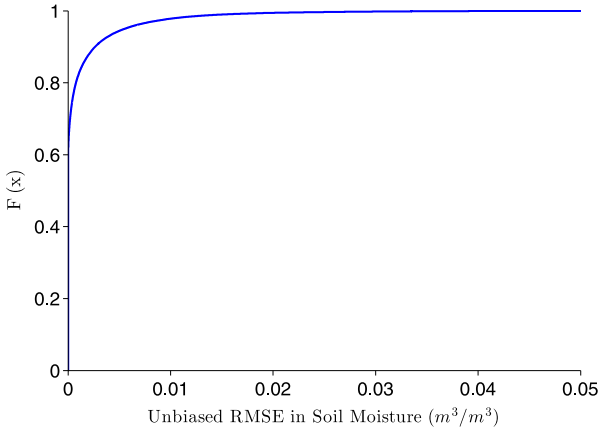

Fig. 8. Cumulative distribution function (CDF) of the errors in disaggregated SM at 1 km.

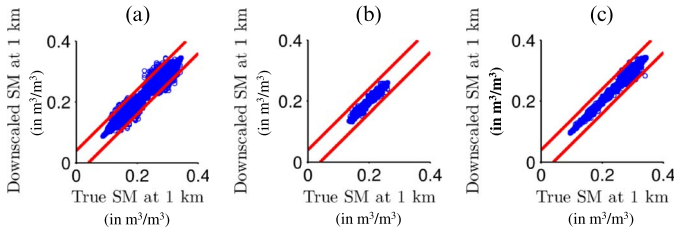

Fig. 9. Disaggregated SM and True soil moisture at 1 km during 2007 for (a) bare-soil, (b) corn, and (c) cotton pixels. The red lines correspond to 0.04 $\mathrm{m}^3/\mathrm{m}^3$ error in disaggregated SM.

The spatially averaged rmse for each DoY in the simulation period is shown in Fig. 7. A Z-test was performed to evaluate whether the disaggregated SM at 1 km is within a standard deviation of $\pm 0.04\ \mathrm{m}^3/\mathrm{m}^3$ from the true SM at 1 km, for meaningful use in hydrological models [35]. This null hypothesis was found to be true for every day of the simulation period. Fig. 8 shows the cumulative density function (cdf) of the errors in disaggregated SM. About 98% of the days have an rmse of less than 0.02 $\mathrm{m}^3/\mathrm{m}^3$ in the disaggregated SM. Fig. 9 shows the disaggregated SM versus true SM at 1 km. The algorithm does not introduce any bias, and the data points are scattered around the $\hat{y} - y = 0$ line, with a positive variance. Most of the points for sweet-corn pixels and all of the points for cotton lie within 0.04 $\mathrm{m}^3/\mathrm{m}^3$. Fig. 10(a) shows the errors for each

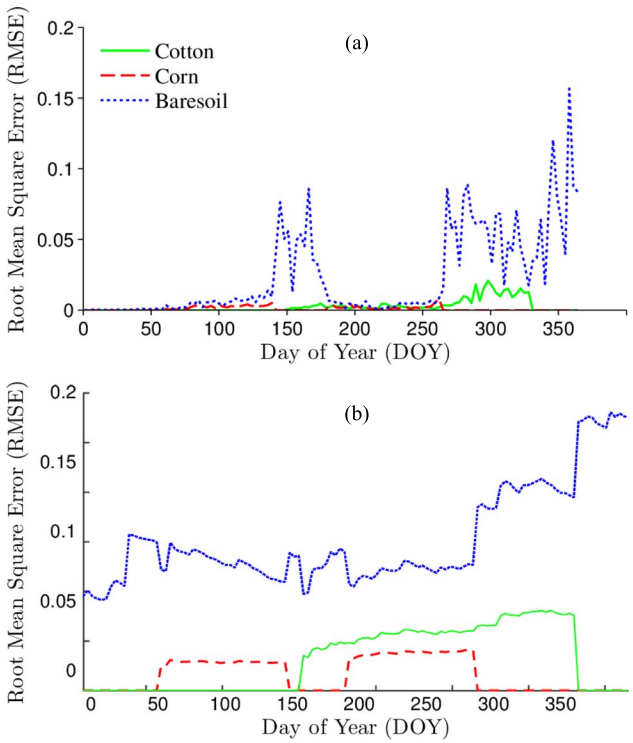

Fig. 10. Spatially averaged rmse in disaggregated SM at 1 km for each day of year in 2007 for bare-soil, corn, and cotton LCs using the (a) SRRM and (b) PRI method.

TABLE III
KLD OVER THE 50×50 km^2 REGION FOR THE DISAGGREGATED ESTIMATES OF SM OBTAINED AT 1 km USING THE SRRM AND PRI METHOD

Land Cover	$\mathrm{KLD_{SRRM}}$	$\mathrm{KLD_{PRI}}$
Corn	1.8615×10^{-17}	0.0234
Cotton	2.4828×10^{-04}	0.0283
Baresoil[b]	5.6222×10^{-5}	0.1036
Baresoil[c]	5.628×10^{-6}	0.0120
Baresoil[d]	2.5948×10^{-6}	0.0114

[b]Baresoil pixels with vegetated sub-pixels at 250 m till DoY 332
[c]Baresoil pixels after DoY 332
[d]Baresoil pixels without any vegetated sub-pixels at 250 m till DoY 332

DoY segregated by type of LC. Bare-soil pixels during periods of vegetation have the highest rmse. This is due to subpixel vegetation at 250 m within a pixel classified as a bare-soil pixel, when the vegetation fraction is < 0.5 at 1 km. Table III shows the KLD between the densities of the disaggregated estimates and the true SM. Bare-soil pixels at 1 km without any vegetation at 250 m have the lowest KLD. Bare-soil pixels at the end of the season that are affected by remnant crops and bare-soil pixels at 1 km with partial vegetation cover at 250 m have a higher KLD, but very close to 0. Vegetated pixels at 1 km contain a higher KLD as well. The boundary pixels classified as bare soil have vegetation at the 250-m scale contributing to these errors.

Among the three scenarios considered for the sensitivity analysis, rmses in downscaled SM are the lowest when just LST is used for disaggregation. This suggests that SM is more strongly coupled to LST than LAI or PPT. This is expected since the spatial patterns apparent in LST images also appear in the SM image, particularly bare-soil pixels, as shown in Fig. 11(b), with a weaker and more complex relationship in corn and cotton, as shown in Fig. 11(c) and (d), respectively. LAI shows higher and similar effects on errors in disaggregated SM for during the mid and late growing seasons of corn and cotton crops. The use of PPT to disaggregate SM results in a lower rmse immediately following a major rainfall event. At other times, its sensitivity to SM is comparable to LST, for bare-soil pixels, and to LAI, for vegetated pixels.

For the five selected days, the inputs, clustering results, the first SM estimate, and PRI disaggregated SM are shown in Figs. 12–15. The clustering results indicate that the implicit inclusion of spatial coordinate information adequately constrains the clusters from becoming too small, whereas the LST, LAI, PPT, and LC ensure that the clusters are simultaneously representative of the LC and meteorological conditions in the region. When fields are significantly smaller than the resolution of auxiliary variables, for example, in developing countries, the implicit inclusion of coordinates might not result in a clustering that accurately follows field boundaries, although it would still separate out regions with different meteorological conditions. This would reduce the accuracy of disaggregated SM at the field edges, and postprocessing based on finer scale LC will be needed in such scenarios. DoY 39 shown in Fig. 12 and DoY 354 shown in Fig. 13 are the bare-soil LC before and after the growing seasons, respectively. The disaggregated estimates for both days are very close to the true SM at 1 km, but due to crop residue and slightly heterogeneous PPT in the region [see Fig. 13(b)], the error for DoY 354 is higher than for DoY 39. It was found that heterogeneity in any one input is enough to capture vegetation patterns in the disaggregated estimate using Kernel regressive models as shown in Figs. 13(a), and 14(a), for corn and cotton, when the LST is fairly uniform across the region, whereas PPT is heterogeneous due to PPT patterns. On DoY 222, even when there was maximum heterogeneity in LC with corn, cotton, and bare soil, the error in SM is minimal as shown in Fig. 16. The effects of noise amplitude in the coarse-scale SM on the disaggregated SM were also investigated on DoY 222. Independent Gaussian noise with zero mean and standard deviations ranging from 0 to 0.1 m^3/m^3 was added to the coarse-scale SM and the spatially averaged unbiased rmse in disaggregated SM is shown in Fig. 17. The errors grow sublinearly, i.e., with a slope lower than 1, whereas the uncertainty in SM is < 0.06 m^3/m^3. When the uncertainty in coarse SM is > 0.06 m^3/m^3, the errors grow with a slope of 1.14 showing that the errors in the disaggregated SM have a higher magnitude than the uncertainties added to coarse SM.

The novelty and efficacy of this disaggregation algorithm lies in the utilization of multiple models using clusters. This is evident in Fig. 18(a); a regression model based on a single cluster fails to fit the coarse SM and auxiliary data with a sufficient degree of accuracy, resulting in speckle noise in the disaggregated SM. Instead, Fig. 18(b) shows that using multiple cluster-based models is an elegant solution that adequately fits the coarse SM and the auxiliary data, and provides disaggregated estimates of SM with low rmse.

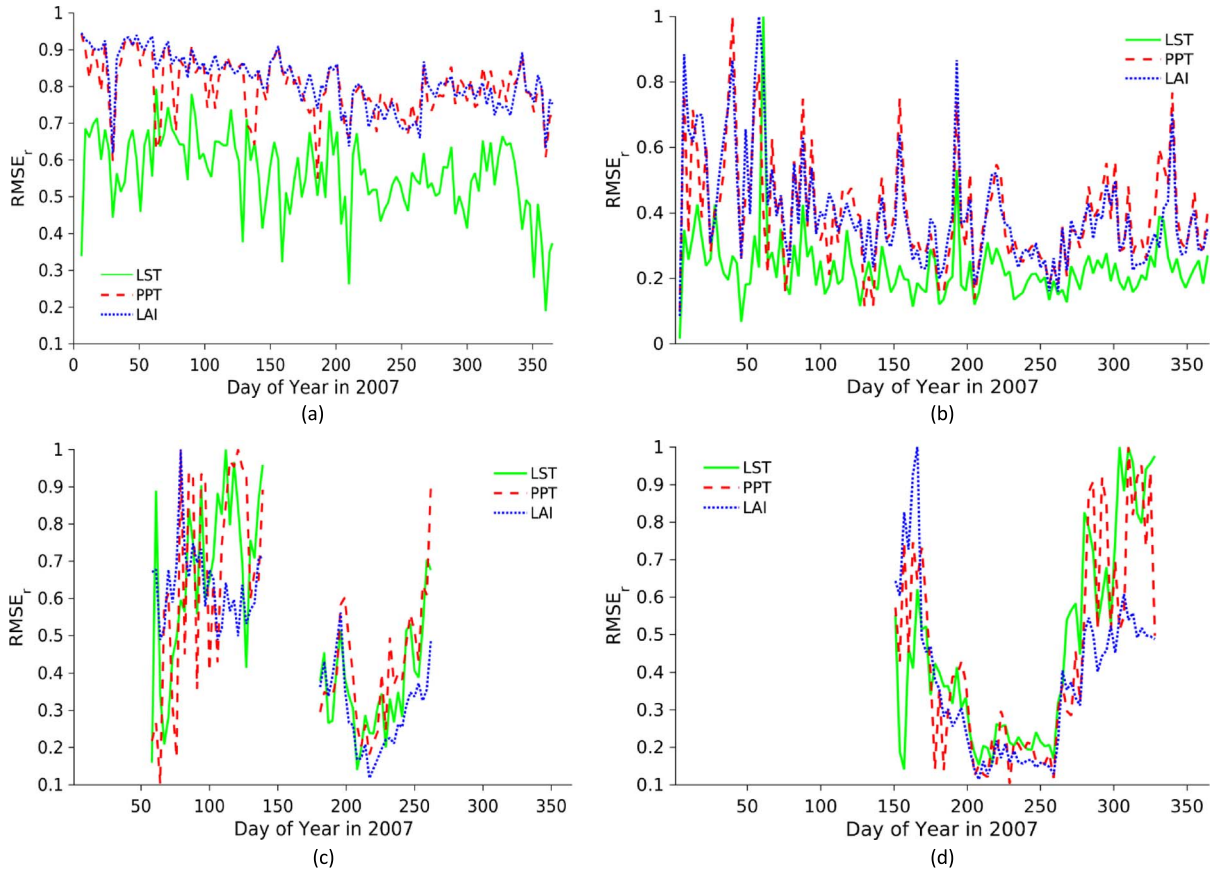

Fig. 11. (a) Relative change in rmse (RMSE_r) when only LST, LAI, or PPT is used as input for disaggregation in (a) the whole region, (b) bare soil, (c) corn, and (d) cotton for each day of 2007.

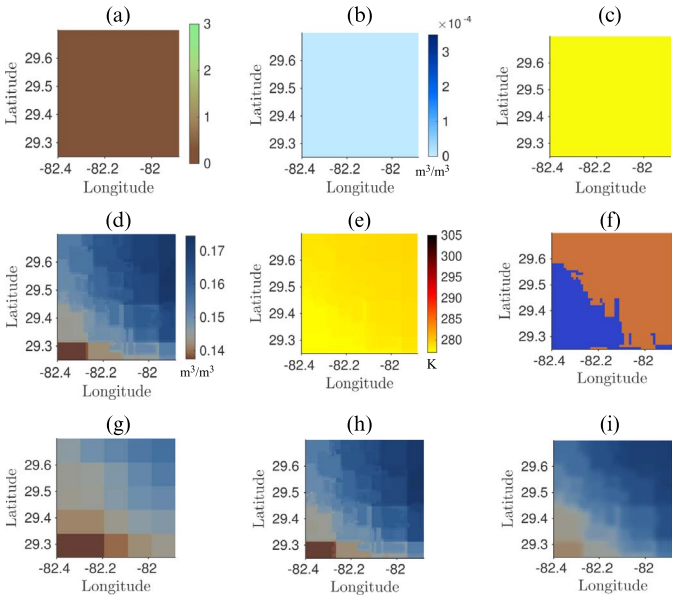

Fig. 12. DoY 39. (a) LAI at 1 km. (b) PPT at 1 km. (c) LC at 1 km (yellow represents bare soil). (d) True SM at 1 km. (e) LST at 1 km. (f) Clustering result at 1 km. (g) SM observations at 10 km. (h) Disaggregated SM using SRRM. (i) Disaggregated SM using the PRI method.

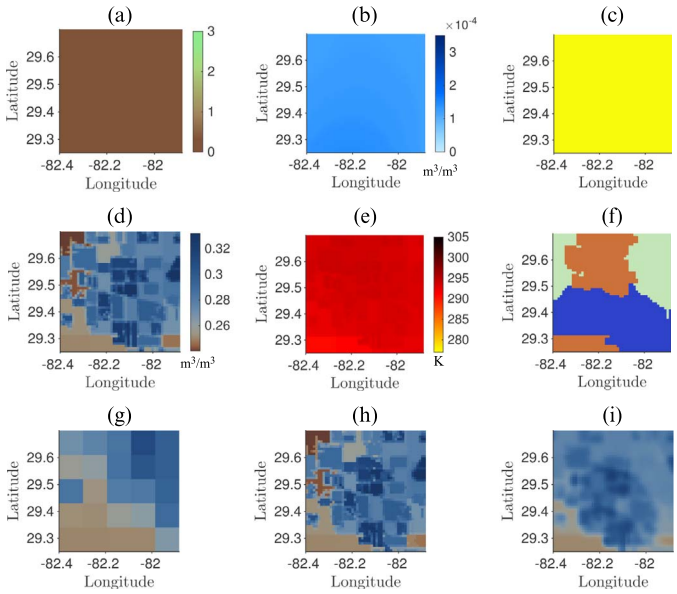

Fig. 13. DoY 354. (a) LAI at 1 km. (b) PPT at 1 km. (c) LC at 1 km (yellow represents bare soil). (d) True SM at 1 km. (e) LST at 1 km. (f) Clustering result at 1 km. (g) SM observations at 10 km. (h) Disaggregated SM using SRRM. (i) Disaggregated SM using the PRI method.

C. Comparison Between SRRM and PRI

The PRI method uses LST, three-day PPT, LAI, LC, and SM at 1 km every three days as input to obtain the first estimate of the SM. To disaggregate SM, in (20), $\mathbf{X}$ is set to {LST, PPT, LAI, LC} and $\mathbf{y}_{\mathrm{train}}$ is set to $\{\mathrm{SM}_{\mathrm{insitu}}\}$. In this paper, 33% of the data set, selected randomly, is used for training the parametric Bayesian model. For the second step, in (21), the SM observations at 10 km are set as $\mathbf{y}_{\mathrm{coarse}}$, and the first

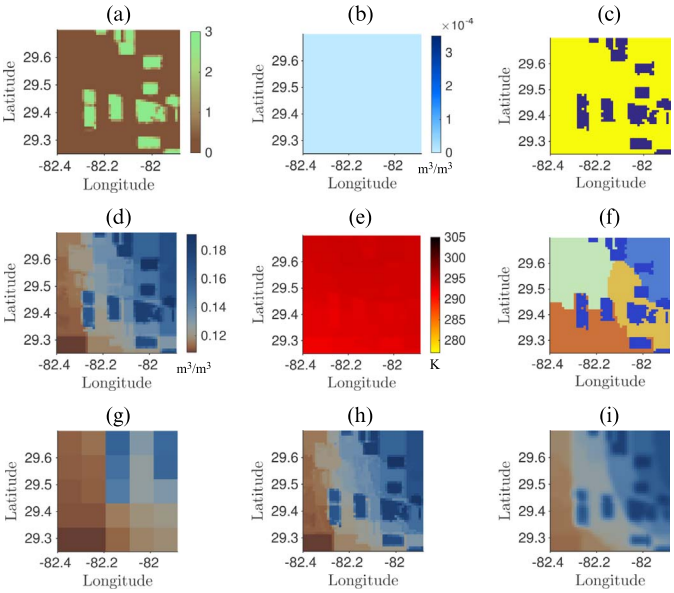

Fig. 14. DoY 135. (a) LAI at 1 km. (b) PPT at 1 km. (c) LC at 1 km (yellow represents baresoil, blue represents corn). (d) True SM at 1 km. (e) LST at 1 km. (f) Clustering result at 1 km. (g) SM observations at 10 km. (h) Disaggregated SM using SRRM. (i) Disaggregated SM using the PRI method.

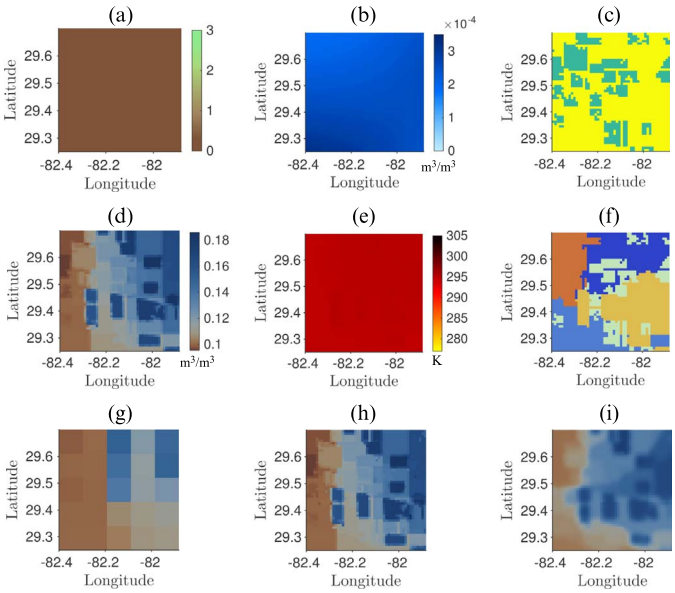

Fig. 15. DoY 156. (a) LAI at 1 km. (b) PPT at 1 km. (c) LC at 1 km (yellow represents baresoil, green represents cotton). (d) True SM at 1 km. (e) LST at 1 km. (f) Clustering result at 1 km. (g) SM observations at 10 km. (h) Disaggregated SM using the SRRM method. (i) Disaggregated SM using the PRI method.

estimates of SM at 1 km from the transformation function are set as $\mathbf{y}_{\text{initial}}$. The value of $\mathbf{m}$ after the cost function $I(\mathbf{m})$ is minimized is the disaggregated SM estimates.

The disaggregated estimates using the SRRM were compared with the PRI estimates using the rmse and the KLD of the estimated densities of the disaggregated observations. The rmse over for the entire time period is assessed for each LC using the SRRM and PRI algorithms that are compared. The spatial errors are also compared for the selected five days during the simulation period, representing different micrometeorological

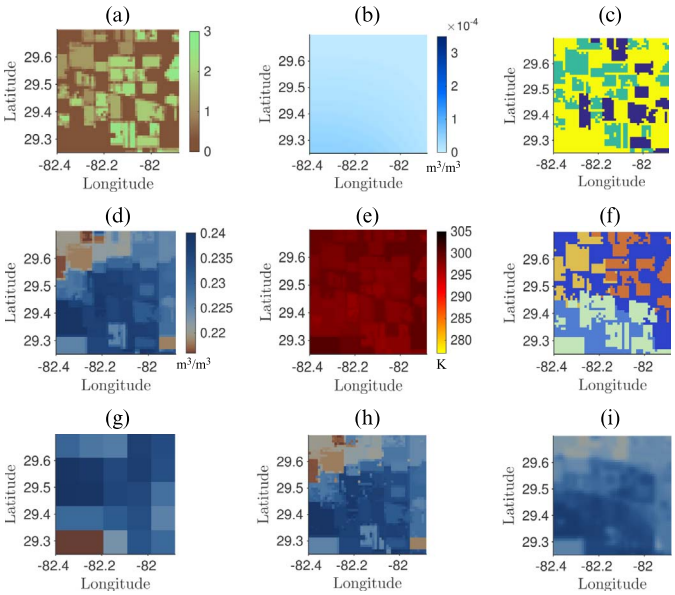

Fig. 16. DoY 222. (a) LAI at 1 km. (b) PPT at 1 km. (c) LC at 1 km (yellow represents bare soil, blue represents corn, and green represents cotton). (d) True SM at 1 km. (e) LST at 1 km. (f) Clustering result at 1 km. (g) SM observations at 10 km. (h) Disaggregated SM using the SRRM method. (i) Disaggregated SM using the PRI method.

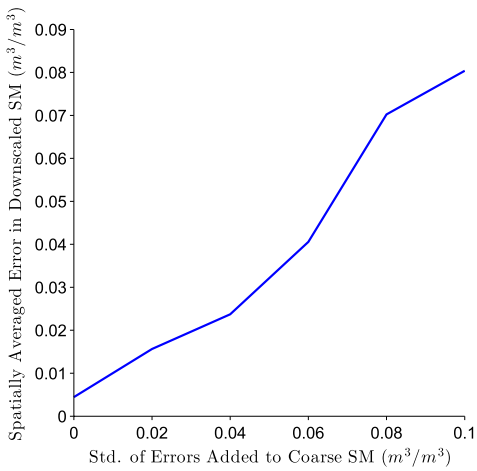

Fig. 17. Standard deviation of noise added to coarse-scale SM as a function of unbiased rmse in disaggregated SM.

and LC conditions. Finally, the running time of the SRRM and PRI algorithms are compared to understand the effects of the difference in algorithm complexity of the two algorithms.

Fig. 7 shows that the rmse of the disaggregated observations using the SRRMs was less than the rmse using the PRI algorithm. The trends observed when the PRI algorithm is used, such as higher errors during periods of vegetation, are preserved when SRRM is used. However, variations in the difference observed between the SRRM- and PRI-based SM can be explained by its different correlations to LC and micrometeorological conditions. The use of separate models in the SRRM enables low rmses even under highly heterogeneous LC. In contrast, the rmses increase by a larger magnitude during heterogeneous LC periods for the PRI algorithm because it uses a single disaggregation model for the whole study region.

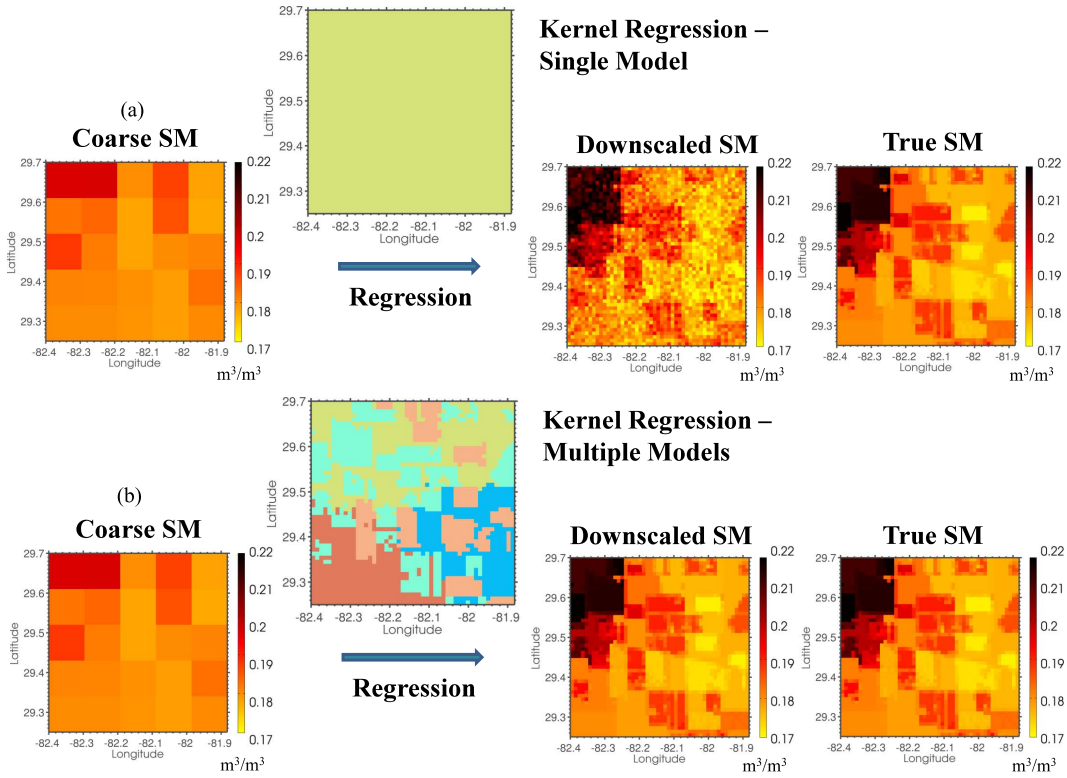

Fig. 18. SM at 10 km, true SM at 1 km and disaggregated SM at 1 km using (a) a single cluster for the study region and (b) multiple clusters following the SRRM algorithm.

Table III compares the KLD between the disaggregated estimates generated by the SRRM and PRI algorithms, and true SM at 1 km. The general trends of KLD over different LC conditions followed by the SRRM are similar to those observed for the PRI. However, the errors for each LC are individually lower for the SRRM compared with the PRI method, as shown in Fig. 10. This is further validated by the KLD for the SRRM estimates that is three orders of magnitude less than for PRI estimates.

Figs. 12–15 compare the disaggregated SM estimates using the SRRM to those using the PRI method. The estimates using PRI does not have sharply defined regions, unlike those observed in the disaggregated SM using SRRM. The sharpness of disaggregated result could arise either from noise or from spatial discontinuities in the inputs due to physical discontinuities in meteorological or LC conditions. Any disaggregation algorithm must maintain the latter while suppressing the former. Equation (21), with $\beta = 2$, maximizes a cost function that blurs the disaggregated SM so that the median error over all pixels is minimized at the cost of a greater variance in error. In the SRRM, the use of multiple models based on clusters ensures that the spatial discontinuity is maintained in the disaggregated SM when it is caused by a physical discontinuity. If the discontinuity originates from additive noise, under certain assumptions, the kernel regression suppresses the discontinuity in the disaggregated SM. The assumptions are that the noise is spatially uncorrelated and has a wide pdf. The results in [36] show that both assumptions are reasonable.

The average execution time of the PRI was about 1.56 h/disaggregation day and that of the SRRM was about 32 min/disaggregation day. This is expected because the complexity of PRI is $\mathcal{O}(N^3 \prod k_j)$ where k_j are the number of bins used to estimate the pdf. For an adequate estimate of the pdfs, $\prod k_j \approx N$ and the complexity approaches $\mathcal{O}(N^4)$, that is an order of magnitude higher than the complexity of the SRRM-based algorithm.

Thus, the SRRM achieves low mean errors using nonlinear regression and low error variances using multiple regressive models with soft boundaries. This ensures that sharpness is maintained along with low rmses. Given exhaustive training data, the PRI algorithm will have similar performance as the SRRM as shown in [18]. However, for operational use of the methodologies at field scale in regions with highly varied LC or micrometeorology with low volume of training data, SRRM provides sharp images of disaggregated SM with a faster run time, less complexity, and lower rmses compared with the PRI.

IV. Conclusion

In this paper, a disaggregation methodology based upon SRRM has been developed, implemented, and evaluated that preserves the high variability in SM due to heterogeneous meteorological and vegetation conditions. The SRRM preserves heterogeneity by utilizing a clustering algorithm to create a number of regions of similarity, which, subsequently, are used in a kernel regression framework. The clusters were computed using remote sensing products, i.e., PPT, LST, LAI, and LC. The kernel regression was implemented on the clusters using *in situ* SM. Across the whole season, 96% of the pixels were found to have a disaggregation error of less than 0.02 m^3/m^3. The KLD values for disaggregated SM at 1 km for the SRRM was equal to 0, for all LCs. In contrast, the PRI method has KLD values that are several orders of magnitude higher, and has a threefold higher execution time. The averaged spatial error is also markedly lower for the SRRM compared with the PRI method.

It is envisioned that the SRRM that is implemented and evaluated in this paper may be applied using satellite images. For example, the PPT data may be obtained from the Global Precipitation Measurement missions and the LAI, LST, and LC products are available from the MODIS sensor aboard Aqua and Terra satellites.

ACKNOWLEDGMENT

The authors would like to thank the High-Performance Computing Center of the University of Florida for all the model simulations conducted in this study, and the anonymous reviewers for their valuable suggestions and comments.

REFERENCES

[1] S. Bojinski *et al.*, "The concept of essential climate variables in support of climate research, applications, and policy," *Bull. Amer. Meteorol. Soc.*, vol. 95, no. 9, pp. 1431–1443, Sep. 2014.

[2] M. Fennessy and J. Shukla, "Impact of initial soil wetness on seasonal atmospheric prediction," *J. Clim.*, vol. 12, no. 11, pp. 3167–3180, Nov. 2009.

[3] H. Douville and F. Chauvin, "Relevance of soil moisture for seasonal climate predictions: A preliminary study," *Clim. Dyn.*, vol. 16, no. 10, pp. 719–736, Oct. 2000.

[4] The GLACE Team *et al.*, "Regions of strong coupling between soil moisture and precipitation," *Science*, vol. 305, no. 5687, pp. 1138–1140, Aug. 2004.

[5] F. Tubiello, C. Rosenzweig, R. Goldberg, S. Jagtap, and J. Jones, "Effects of climate change on U.S. crop production: Simulation results using two different GCM scenarios. Part I: Wheat, potato, maize, and citrus," *Clim. Res.*, vol. 20, no. 3, pp. 259–270, 2002.

[6] J. Yuste *et al.*, "Soil respiration in a mixed temperate forest and its contribution to total ecosystem respiration," *Tree Physiol.*, vol. 25, no. 5, pp. 609–619, May 2005.

[7] A. Friend *et al.*, "FLUXNET and modelling the global carbon cycle," *Global Change Biol.*, vol. 13, no. 3, pp. 613–633, Mar. 2007.

[8] F. Fecan, B. Marticorena, and G. Bergametti, "Parametrization of the increase of the aeolian erosion threshold wind friction velocity due to soil moisture for arid and semi-arid areas," *Ann. Geophys.-Atmos. Hydrosp. Space Sci.*, vol. 17, no. 1, pp. 149–157, Jan. 1999.

[9] G. Holtgrieve, P. Jewett, and P. Matson, "Variations in soil N cycling and trace gas emissions in wet tropical forests," *Oecologia*, vol. 146, no. 4, pp. 584–594, Jan. 1999.

[10] 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, pp. 69–88, Nov. 2005.

[11] S. Chakrabarti, T. Bongiovanni, J. Judge, L. Zotarelli, and C. Bayer, "Assimilation of SMOS soil moisture for quantifying drought impacts on crop yield in agricultural regions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3867–3879, Sep. 2013.

[12] D. Entekhabi *et al.*, "The soil moisture active/passive mission (SMAP)," in *Proc. IEEE IGARSS*, 2008, vol. 3, pp. III-1–III-4.

[13] M. Piles, A. Camps, M. Vall-llossera, and M. Talone, "Spatial-resolution enhancement of SMOS data: a deconvolution-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2182–2192, Jul. 2009.

[14] M. Piles *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.

[15] S. Sanchez-Ruiz *et al.*, "Downscaling SMOS-derived soil moisture using visible/infrared data," *J. Hydrol.*, vol. 516, pp. 273–283, Aug. 2014.

[16] Z. Jiang, S. Shekhar, X. Zhou, J. Knight, and J. Corcoran, "Focal-test-based spatial decision tree learning: A summary of results," in *Proc. IEEE 13th ICDM*, 2013, pp. 320–329.

[17] J. Principe, D. Xu, and J. Fisher, *Information Theoretic Learning, in Unsupervised Adaptive Filtering*. New York, NY, USA: Wiley, 2010.

[18] S. Chakrabarti, T. Bongiovanni, J. Judge, K. Nagarajan, and J. C. Principe, "Downscaling satellite-based soil moisture in heterogeneous regions using high-resolution remote sensing products and information theory: A synthetic study," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 85–101, Jan. 2014.

[19] A. Likas, N. Vlassis, and J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003.

[20] R. Jennsen, D. Erdogmus, K. Hild, J. Principe, and T. Eltoft, "Optimizing the Cauchy-Schwarz PDF distance for information theoretic, non-paremetric clustering," in *Proc. 5th Int. Conf. EMMCVPR*, 2005, vol. 1, pp. 34–45.

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

[22] B. Kibria, "Performance of some new ridge regression estimators," *Commun. Stat.—Simul. Comput.*, vol. 32, no. 1, pp. 419–435, Jan. 2003.

[23] B. Schlkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: The MIT Press, 2001.

[24] K. Nagarajan and J. Judge, "Spatial scaling and variability of soil moisture over heterogenous land cover and dynamic vegetation conditions," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 880–884, Jul. 2013.

[25] T. Bongiovanni *et al.*, "Field observations during the eighth microwave, water, and energy balance experiment (MicroWEX-8): From June 16 through August 24, 2009," Center Remote Sens., Univ. Florida, Gainesville, FL, USA, Tech. Rep., 2009. [Online]. Available: http://edis.ifas.ufl.edu/ae476

[26] J. Casanova *et al.*, "Field observations during the fifth microwave, water, and energy balance experiment (MicroWEX-5): From March 9 through May, 2006," Center Remote Sens., Univ. Florida, Gainesville, FL, USA, Tech. Rep. Circular 1514, 2006. [Online]. Available: http://edis.ifas.ufl.edu/AE407

[27] T. Lin *et al.*, "Field Observations During the Third Microwave, Water, and Energy Balance Experiment (MicroWEX-3): From June 16 Through December 21, 2004," Center Remote Sens., Univ. Florida, Gainesville, FL, USA, Tech. Rep. Circular 1481, 2004. [Online]. Available: http://edis.ifas.ufl.edu/ae361

[28] Florida Automated Weather Network, 2011. [Online]. Available: http://fawn.ifas.ufl.edu/

[29] W. H. Brutsaert, "On a derivable formula for long-wave radiation from clear skies," *Water Res. Res.*, vol. 11, no. 5, pp. 742–744, Oct. 2010.

[30] W. Crow and E. Wood, "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," *Adv. Water Res.*, vol. 26, no. 2, pp. 137–149, Feb. 2003.

[31] O. Merlin *et al.*, "Self-calibrated evaporation-based disaggregation of SMOS soil moisture: An evaluation study at 3 km and 100 m resolution in Catalunya Spain," *Remote Sens. Environ.*, vol. 130, pp. 25–38, Mar. 2013.

[32] C. Huang, X. Li, and L. Lu, "Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter," *Remote Sens. Environ.*, vol. 112, pp. 1320–1336, Apr. 2008.

[33] J. Privette *et al.*, "Early spatial and temporal validation of MODIS LAI product in the Southern African Kalahari," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 232–243, Nov. 2002.

[34] W. Crow and E. F. Wood, "The value of coarse-scale soil moisture observations for regional surface energy balance modeling," *J. Hydrometeorol.*, vol. 3, no. 4, pp. 467–482, Aug. 2002.

[35] N. Das, D. Entekhabi, and E. Njoku, "An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1504–1512, May 2011.

[36] J. Munoz-Sabater, P. de Rosnay, C. Jimenez, L. Isaksen, and C. Albergel, "SMOS brightness temperature angular noise: Characterization, filtering, and validation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5827–5839, Sep. 2014.

Subit Chakrabarti (S'08) was born in Vadodara, India, in 1990. He received the B.Tech. degree in electronics and instrumentation engineering from the West Bengal University of Technology, Kolkata, India, in 2008. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering, University of Florida, Gainesville, FL, USA.

He is currently a Research Assistant with the Center of Remote Sensing, Department of Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, and with the Computational Neuro-Engineering Laboratory, University of Florida. His research interests include applying machine learning techniques for spatial scaling of satellite-based observations for agricultural applications.

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 with the Department of Agricultural and Biological Engineering, 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 heterogeneous landscapes, and data assimilation.

Dr. Judge is the Chair of the National Academies' Standing Committee on Radio Frequencies and a member of the Frequency Allocations in Remote Sensing Technical Committee of the IEEE Geoscience' Remote Sensing Society. She was the Chair of the Remote Sensing Technical Committee in the Hydrology Section of the American Geophysical Union.

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.

Her master's research focused on modeling crop growth and development and on improving modeled estimates of biomass and yield through assimilating remote sensing observations. She is currently working with the Center of Remote Sensing, Department of Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, where she supervises crop vegetation sampling, data collection, sensor monitoring, and data archiving. Her research interests include the use of geographic information systems for decision support models, such as processing field and remotely sensed data.

Anand Rangarajan (S'84–M'91) received the Ph.D. degree from the University of Southern California, Los Angeles, CA, USA.

He was an Assistant Professor with the Department of Diagnostic Radiology and the Department of Electrical Engineering, Yale University, New Haven, CT, USA. He is currently an Associate Professor with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA. His research interests include machine learning, computer vision, remote sensing, medical and hyperspectral imaging, and the scientific study of consciousness.

Sanjay Ranka (F'02) received the B.Tech. degree in computer science from the Indian Institute of Technology, Kanpur, India, and the Ph.D. degree in computer science from the University of Minnesota Twin Cities, Minneapolis, MN, USA.

He is currently a Professor with the Department of Computer Information Science and Engineering, University of Florida. He recently has served as the Chief Technology Officer at Paramark, where he developed real-time optimization software for optimizing marketing campaigns. He is the author or coauthor of four books and 225 journal and refereed conference articles. His current research interests include various issues on data science, such as energy-efficient computing, high-performance computing, data mining, and informatics.

Dr. Ranka served as the Editor for the IEEE TRANSACTION ON COMPUTERS and IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. He currently serves as the Associate Editor-in-Chief of the *Journal of Parallel and Distributed Computing* and as an Associate Editor for ACM/IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, *Sustainable Computing: Systems and Informatics, Knowledge and Information Systems*, and *International Journal of Computing*. He received a Best Student Paper Runner Up Award at the 2014 International Conference on Bioinformatics and Computational Biology and the 2015 IEEE International Geoscience and Remote Sensing, the Best Student Paper Award at the 2010 ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, and the Best Paper Runner Up Award at the 2009 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. He is a Fellow of the American Association for the Advancement of Science.