Disaggregation of Remotely Sensed Soil Moisture in Heterogeneous Landscapes using Holistic Structure based Models

Subit Chakrabarti, *Student Member, IEEE*,
Jasmeet Judge, *Senior Member, IEEE*, Anand Rangarajan, *Member, IEEE*,
Sanjay Ranka, *Fellow, IEEE*.

Abstract

In this study, a novel machine learning algorithm is proposed to disaggregate coarse-scale remotely sensed observations to finer scales, using correlated auxiliary data at the fine scale. It includes a regularized Cauchy-Schwarz distance based clustering step that assigns soft memberships to each pixel at the fine-scale followed by a kernel regression that computes the value of the desired variable at all the pixels. This algorithm, based on self-regularized regressive models (SRRM), is implemented to disaggregate soil moisture (SM) from 10km to 1km using land cover, precipitation, land surface temperature, leaf area index and some point observations of SM. This was tested using multi-scale synthetic observations in NC Florida for heterogeneous agricultural land covers, with two growing seasons of sweet corn and one of cotton, annually. It was found that the root mean square error (RMSE) for 96 % of the pixels was less than 0.02 $\mathbf{m^3/m^3}$. The Kullback Leibler divergence (KLD) between the true SM and the disaggregated estimates was close to 0, for both vegetated and baresoil land covers. The disaggregated estimates are compared to those generated by the Principle of Relevant Information (PRI) method. The RMSE for the PRI disaggregated estimates are higher than the RMSE for the SRRM methods on each day of the season. The KLD of the disaggregated estimates generated by the SRRM method estimates is at least 3 orders of magnitude lesser than for the PRI disaggregated estimates.

Index Terms

Disaggregation, Microwave brightness temperature, Super-resolution Soil Moisture, Kernel Regression, Clustering, Multi-spectral Remote Sensing.

Preprint submitted to IEEE® Transactions on Image Processing on 28th January 2015.
This work was supported in part by the NASA-Terrestrial Hydrology Program (THP)-NNX13AD04G.
S. Chakrabarti 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, USA; A. Rangarajan and S. Ranka are with the Department of Computer & Information Science & Engineering, University of Florida, Gainesville. E-mail: subitc@ufl.edu

I. Introduction

REMOTELY-SENSED images at high spatial resolutions provide richer detail and improved information extraction capacities. Applications such as species identification, urban change studies, hydrothermal mapping, and crop monitoring benefit from higher resolution data than afforded by current generation radiometric instruments. Spatial scaling of satellite-based RS images has been a super-resolution problem, from an image processing and analysis perspective. The earliest super-resolution algorithm was developed to improve the spatial resolution of LANDSAT images, using multiple under-sampled images with sub-pixel displacements in the frequency domain [1]. These frequency domain formulations are computationally efficient and use shift and aliasing properties of discrete and continuous Fourier transforms to improve spatial resolution. However, they do not incorporate prior knowledge or spatial regularization, that is needed in more complex observational models. In contrast the recent spatial domain approaches assume that the coarse resolution and fine resolution image are linearly related with sparsity constraints. Thus estimators such as maximum likelihood, maximum *a posteriori* and projection onto convex sets have been used to interpolate the coarse scale image in a regularized manner to obtain the high resolution image. For example, a quadratic cost function along with a Huber prior registration model was used for the super-resolution of images from the moderate-resolution imaging spectrometer (MODIS) to exclude clear outlier pixels and preserves edges [2]. Other models have been proposed for the regularization step including Universal Hidden Markov Trees [3] and Multi-fractal models [4]. However, regularization may not guarantee optimality in reconstruction and these algorithms have been found to sacrifice radio-accuracy for visual appeal of the resultant images [5].

Unlike most super-resolution techniques, remote sensing disaggregation algorithms maintain radio-accuracy by generating higher resolution data using local correlations to other high-resolution physical or meteorological data, where coarse resolution data only serves to regularize the estimates. 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 interpolated, power-law or temporal persistence across scales. Methods using this approach usually assume static vegetation and micro-meteorology 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 geo-statistical methods, such as regression, co-kriging and block kriging, and fractal interpolation. The third approach is to employ the Triangle Method [6], [7] and statistical models to extrapolate the dependant 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 auto-correlation [8]. Moreover, these approaches use second order metrics, which do not leverage all the information in the data that is necessary in a highly non-linear regression problems such as disaggregation [9].

A recently implemented disaggregation algorithm [10] based on the principle of relevant information (PRI) addresses the above inadequacies by utilizing the full probability density function of a set of training observations, rather than second order moments, to approximate a transformation function that relates micro-meteorological data recorded in a region to *in-situ* soil moisture (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 and the initial SM values using an information theoretic cost function. Although this results in low disaggregation errors, it is computationally intensive. Additionally, it requires a comprehensive training set for the initial estimate of the multi-dimensional PDF to converge. In this study, a self-regularized regressive model (SRRM) is used to disaggregate SM. It is 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.

SM is a key governing factor in surface and sub-surface hydrological and agricultural models as it regulates the land-atmosphere interactions. Representational models of weather [11]–[13], crop growth [14], ecosystem and carbon cycle processes [15], [16], dust generation [17], trace gas fluxes [18], and agricultural drought [19], [20] require soil moisture data at a fine spatial resolution. Recent satellite missions, including the already functional European Space Agency (ESA) Soil Moisture and Ocean Salinity (SMOS) and the soon to be launched National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive missions, provide for SM retrievals at unprecedented spatial and temporal resolutions of tens of kilometres every 2-3 days, with worldwide coverage. However, models simulating physical processes need SM at even finer scales of 1 km [20]. Disaggregation addresses this discrepancy in scales by generating local fine-resolution data from coarse-resolution data obtained from satellites.

This goal of this study is to implement a novel machine learning algorithm that disaggregates coarse scale remotely sensed products with auxiliary fine-scale data. The primary objectives of this study are to - 1) 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), land cover (LC) and precipitation (PPT); and 2) evaluate the SRRM-based methodology and compare it with the PRI method using a synthetic dataset.

Section II describes the theoretical details of the disaggregation framework based on self-regularized regressive models and provides a brief description of the PRI algorithm for disaggregation. Section III illustrates the steps for the implementation of

the SRRM algorithm and presents the disaggregation results for SM at 1 km and Section IV summarizes the important results, concludes the paper, and outlines the scope for future studies.

II. Disaggregation Framework

The problem of disaggregation is an ill-conditioned problem that is limited physically by the convolution of the point spread function of the imaging system. Under this constraint, generation of fine-scale data from coarse-scale data needs additional information which is spatially correlated to the variable to be disaggregated, 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 method 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 sections. Using spatial regions or dynamic conglomeration of pixels to generate models instead of treating each pixel in a sample-based method reduces the effect of spatial autocorrelation on the disaggregated estimates.

A. Disaggregation Framework based on Self-Regularized Regressive Models (SRRM)

In this study, a number of models are created dynamically based on generalized proximity regions in the high dimensional correlated data. The membership of a pixels to a region, and thus to a model, is soft and constrained to a sum of one across the space of models. The models themselves are trained using a kernel regression based method. This is a novel way to account for correlated features using algorithms that require an IID (independence and identical distribution) assumption [8]. Figure[a] 1 shows a flow diagram of the algorithm and the steps required to generate disaggregated estimates. The overall organization and the datasets involved is shown in Figure 2. The two steps of the algorithm include clustering and kernel regression, as follows.

1) Information theoretic clustering based on the Cauchy-Schwarz Distance: The generalized proximity regions are identified using a regularized variant of a clustering method based on information theory [21]. For two vectors $\mathbf{x}$ and $\mathbf{y}$, the Cauchy-Schwarz inequality is,

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

where $<\mathbf{x},\mathbf{y}>$ is the inner product of vectors $\mathbf{x}$ and $\mathbf{y}$. For probability density functions $p(x)$ and $q(x)$, the inner product is defined as, $<p,q>=\int p(x)q(x)\mathrm{d}x$ over the support for the distributions p and q. Then the Cauchy-Schwarz inequality in a metric space spanned by the PDF's is,

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

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 Equation 2 is maximized. Since the algorithm is gradient based, and logarithm is a monotonically increasing function, only the argument of the logarithm in $D_{CS}=-\log J_{CS}(p,q)$ can be minimized. An estimator for this can be constructed from data-samples and extended to the case of multiple clusters by using a membership vector.

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

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

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

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 vector then is equivalent to solving this constrained optimization problem:

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

[a]All Figures and Tables are included at the end of the manuscript for clarity.

where $\mathbf{1}$ is a 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,

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

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

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

where $\Gamma = \mathrm{diag}(2\sqrt{m_{i1}}, \ldots, 2\sqrt{m_{iK}})$ is the magnitude normalizing factor. The Lagrange Multipliers then, after constructing the necessary Lagrange Function is given by

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

The updated vector for the next iteration is,

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

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 all the diagonal elements are positive, the direction of the gradients are the same. This is enforced by adding a small positive constant $\alpha \sim 0.05$ to all elements in the iteration. If $\hat{J}_{CS}$ is represented as $\frac{U}{V}$, then the gradient of $\hat{J}_{CS}^{REG}$ can be calculated as:

$$\begin{aligned}
&\frac{\partial \hat{J}_{CS}^{REG}}{\partial \mathbf{m}_i} = \frac{V \frac{\delta U}{\partial \mathbf{m}_i} - U \frac{\partial V}{\partial \mathbf{m}_i}}{V^2} - \psi \sum_{k=1}^{K} (1 + \log(m_{ik})) \text{ with} \\
&U = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(1 - \mathbf{m}_i^T \mathbf{m}_j\right) G_{\sqrt{2}\sigma}(\mathbf{x}_i, \mathbf{x}_j), \quad V = \sqrt{\prod_{k=1}^{K} v_k}, \\
&\frac{\partial U}{\partial \mathbf{m}_j} = -\sum_{j=1}^{N} \mathbf{m}_j G_{\sigma\sqrt{2}}(\mathbf{x}_i, \mathbf{x}_j) \quad \text{and} \quad \frac{\partial V}{\partial \mathbf{m}_i} = \frac{1}{2} \sum_{k'=1}^{K} \sqrt{\frac{\prod_{\substack{k=1 \\ k \neq k'}}^{K} v_k}{v_{k'}}} \frac{\partial v_{k'}}{\partial \mathbf{m}_i},
\end{aligned} \tag{10}$$

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

Kernel Annealing: The technique of Kernel Annealing is very useful in this algorithm. The performance surface has local minima which can inhibit the performance of this algorithm. So the kernel width is gradually decreased over the course of iterations. The initial value of the kernel is chosen according to the Silverman's rule of thumb given by

$$\sigma_{\mathrm{SIL}} = \sigma_X \left(4N^{-1} (2d + 1)^{-1}\right)^{\frac{1}{d+4}} \tag{11}$$

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

$$r = \frac{\sigma_{\mathrm{SIL}} - \sigma_{\mathrm{LOW}}}{N_{\mathrm{TOT}}} = \frac{3\sigma_{\mathrm{SIL}}}{4N_{\mathrm{TOT}}} \tag{12}$$

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 is a parametric regression technique that adds a scaled regularizing term to the cost function to to increase generalization. The cost function for ridge regression is

$$\mathcal{E}(\mathbf{w}, \mathbf{x}) = \frac{1}{2} \sum_i (y_i - w^{\mathbf{T}} \mathbf{x}_i)^2 + \frac{1}{2} \mu \|\mathbf{w}\|^2 \tag{13}$$

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

$$\frac{\partial \mathcal{E}}{\partial \mathbf{w}} = 0 \implies \mathbf{w} = \left(\sum_i \mathbf{x}_i \mathbf{x}_i^{\mathbf{T}} + \mu \mathbf{I} \right)^{-1} \left(\sum_i y_i \mathbf{x}_j \right) \tag{14}$$

If this computation was carried out in a Reproducing Kernel Hilbert Space (RKHS), then the inner-products can be replaced with a kernel evaluation. Let $\mathcal{H}$ be a Hilbert space with an inner-product metric $< \cdot, \cdot >_{\mathcal{H}}$. 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 $< \mathbf{x}, \mathbf{y} >_{\mathcal{H}} = \kappa(\mathbf{x}, \mathbf{y})$. 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,

$$\mathbf{w} = (\mu \mathbf{I}_D + \Phi \Phi^{\mathrm{T}})^{-1} \Phi \mathbf{y} \tag{15}$$

Where D is the dimension of the feature space. The dimension of the feature space is not well-defined in many cases, so the weights can be rewritten using the identity, $(A^{-1} + B^{\mathrm{T}} C^{-1} B)^{-1} B^{\mathrm{T}} C^{-1} = A B^{\mathrm{T}} (B A B^{\mathrm{T}} + C)^{-1}$,

$$\mathbf{w} = \Phi (\mu \mathbf{I}_N + \Phi^{\mathrm{T}} \Phi)^{-1} \mathbf{y} \tag{16}$$

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 $\mathbf{x}'$,

$$\begin{aligned} \hat{y} &= \mathbf{w}^{\mathrm{T}} \Phi(\mathbf{x}') \\ &= \mathbf{y} (\mu \mathbf{I}_N + \Phi^{\mathrm{T}} \Phi)^{-1} \Phi^{\mathrm{T}} \Phi((x')) \\ &= \underbrace{\mathbf{y} (\mu \mathbf{I}_N + \mathrm{K})^{-1}}_{\mathbf{w}} \kappa(\mathbf{x}, \mathbf{x}') \end{aligned} \tag{17}$$

where 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.

Algorithm 1 Disaggregation using Self-Regularized Regressive Models

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. $\mathrm{N_{DAYS}}$ is the total number of days.

for $i = 0$ to $\mathrm{N_{DAYS}}$ **do**

Step 1: Clustering

for $i = 1$ to 30 **do**

Calculate $\hat{J}_{CS}^{REG}$ and $\frac{\partial \hat{J}_{CS}^{REG}}{\partial \mathbf{m}_i}$ according to Equation 4 and 10.

Update λ_i and $\mathbf{v}_i^+$ according to Equation 8 and 9.

end for

Step 2: Kernel Regression

Calculate $\mathbf{w}$ according to Equation 16 using the training set.

Estimate the disaggregated observations, $\hat{\mathbf{y}}$ for the test set using Equation 17.

Run 10-fold cross-validation for the values of N and the cross-validation constants ψ and μ.

end for

3) Algorithm Summary and Computational Complexity: The SRRM disaggregation method is summarized and shown in Algorithm 1. The complexity of the $\mathrm{D_{CS}}$ based clustering algorithm is $\mathcal{O}(N^2)$ for each iteration. For good convergence 30 iterations are needed, which is much lower than the dimensionality of the data-set, and does not affect the complexity. 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 $\mathcal{O}(MN)$ $(M << N)$. For M being 33% of N, the results are comparable to the original method and takes a fraction of the time. The average complexity of the ridge-regression method is $\mathcal{O}(N^3)$ [22].

B. The PRI Framework

The disaggregation methodology using PRI includes a transformation process to obtain a probabilistic relationship between the variable to be disaggregated, say $\mathbf{y}$, at 1 km using auxiliary information, say $\mathbf{X}$, the same scale. A discrete formulation of the Bayes rule is used to estimate $\mathbf{y}_{\mathrm{INITIAL}}$ at fine resolution, as given in equation (18), wherein $\mathbf{y}_{\mathrm{TRAIN}}^{i}$ is discretized into k classes, $i \in [1, k]$, and $\mathbf{x}_{j,\mathrm{TRAIN}}^{i_1}$ is discretized into k_j classes in $i_1 \in [1, k_j]$, where j indexes the individual variables that comprise $\mathbf{X}$, say m.

$$p(\mathbf{y}^{i_1}_{\text{INITIAL}}|\mathbf{X}^{i_1}_{\text{TRAIN}}) = \frac{p(\mathbf{X}^{i_1}_{\text{TRAIN}}|\mathbf{y}^{i}_{\text{TRAIN}})p(\mathbf{y}^{i}_{\text{TRAIN}})}{p(\mathbf{X}^{i_1}_{\text{TRAIN}})}$$
$$\mathbf{y}^{i}_{\text{INITIAL}} = \arg\max_{\mathbf{y}^{i}_{\text{TRAIN}}} \frac{p(\mathbf{X}^{i_1}_{\text{TRAIN}})p(\mathbf{y}^{i}_{\text{TRAIN}})}{p(\mathbf{X}^{i_1}_{\text{TRAIN}})}$$
$$p(\mathbf{X}^{i_1}_{\text{TRAIN}}) = \sum_{i=1}^{k} p(\mathbf{X}^{i_1}_{\text{TRAIN}}|\mathbf{y}^{i}_{\text{TRAIN}})p(\mathbf{y}^{i}_{\text{TRAIN}}) \quad (18)$$

In the second step, $\mathbf{y}_{\text{INITIAL}}$ is merged with the observations at the coarser resolutions, $\mathbf{y}_{\text{COARSE}}$ to obtain improved estimates at fine resolution,

$$\arg\max_{\mathbf{m}} J(\mathbf{m}) = H(\mathbf{m}) + \beta KL(p_{\mathbf{m}}||p_{\mathbf{y}_{\text{INITIAL}}}) \quad (19)$$

where $J(\mathbf{m})$ is the cost function, $p_{\mathbf{y}_{\text{INITIAL}}}$ is the PDF of the original data, and $p_{\mathbf{m}}$ is the PDF at each iteration. $H(\mathbf{m})$ is the entropy, and KL is the KL divergence. $\mathbf{m}$ is initialized to $\mathbf{y}_{\text{COARSE}}$ at the first iteration. The β is a user-defined weighting parameter that balances the redundancy and information preservation in $J(\mathbf{m})$. 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 study, 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. A detailed description of the PRI algorithm can be found in (Chakrabarti et. al. 2014).

III. Experimental Description and Results

A. Multiscale synthetic dataset

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 Tranfer (DSSAT) model, described in [23]. A 50×50 km^2 region, equivalent to 25 SMAP pixels, was chosen in North Central Florida (see Figure 3) for the simulations. The region encompassed the UF/IFAS Plant Science Research and Education Unit, Citra, FL, where a series of season-long field experiments, called the Microwave, Water and Energy Balance Experiments (MicroWEXs), have been conducted for various agricultural land covers over the last decade [24]–[26] used in this study. Simulated observations of LST & 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 study 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 precipitation, relative humidity, air temperature, downwelling solar radiation, and wind speed were obtained from eight Florida Automated Weather Network (FAWN) stations [27] located within the study region (see Figure 3). The observations were spatially interpolated using splines to generate the meteorological forcings at 200 m. Long-wave radiation were estimated following Brutsaert [28].

The model simulations were performed over each agricultural field rather than all the pixels, to reduce computation time. Based upon land cover information at 200 m, contiguous, homogeneous regions of sweet-corn and cotton were identified, as shown in Figure 4. A realization of the LSP-DSSAT model was 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 forcings at the centroid, as shown in Figure 4. Linear averaging is typically sufficient to illustrate the effects of resolution degradation [29]. 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 were used as truth to evaluate the downscaling methodology. To simulate rain-fed systems, all the water input from both precipitation and irrigation were combined together, and the "PPT" in this study represents these combined values, representing a rain-fed system.

B. Disaggregation Framework based on SRRM

The simulation period, from Jan 1 (DoY 1) to Dec 31 (DoY 365), 2007, consisted of two growing seasons of sweet corn and one season of cotton, as shown in Table[a] I. The LST, PPT, and LAI observations at 1 km were obtained by adding noise to account for satellite observation errors, instrument measurement errors, and micro-meteorological variability, following [30]–[32]. Errors with zero mean and standard deviations of 5K, 1 mm, 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 method uses LST, 3-day PPT, LAI, LC at 1 km and SM at 10 km every 3 days as input. In the first step, the information-theoretic cost function described in Section II-A is used to cluster the field 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. The optimal number of iterations that produces a usable clustering result is determined by evaluating the root mean square error (RMSE) for Day 222, characterized by maximum input heterogeneity, in disaggregated SM after every iteration, for upto 200 iterations. At the end of this step, each pixel has a vector of n numbers, $(m_1, m_2, \ldots, m_N)$ that sum upto 1 describing its membership to each of the n clusters.

In the second step, N models, $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_N$ are developed using LST, 3-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 33% of the pixels that make up the field. 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,

$$\mathrm{SM}_i^{1\mathrm{km}} = \mathbf{m}^{\mathrm{T}} \cdot \left(\hat{f}_1(\mathbf{x'}_i), \hat{f}_2(\mathbf{x'}_i), \ldots, \hat{f}_N(\mathbf{x'}_i)\right) \tag{20}$$

The SRRM method is evaluated by plotting the RMSE and standard deviation of the errors over the entire season. The RMSE was also plotted for the entire time-period for each land-cover. Moreover, the disaggregated SM is plotted versus the true SM. To evaluate how close the density function of the disaggregated estimates is to the density function of the true SM, the Kulback Liebler-Divergence (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 like the Hellinger distance or χ^2-distance can also be used. In addition, 5 days were selected from the season to understand the effect of the heterogeneity in inputs on the error in disaggregated SM. Variabilities in precipitation, ranging from uniformly wet to uniformly dry, and in land cover, ranging from bare soil to vegetated with both cotton and sweetcorn, 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 land cover conditions provide an index of dynamic errors that can be expected.

Figure 5 shows the spatially averaged RMSE between disaggregated SM and the observations at 1 km on DoY 222 for different iterations of the clustering algorithms. During DoY 222, both the land cover and micro-meteorological conditions were heterogeneous, providing the worst case-scenario for convergence of the algorithm. All parameters, except the number of clusters are cross-validated for each individual iteration. The number of clusters is cross-validated once, using 50 iterations of the clustering algorithm. The error oscillates with a mean amplitude of 1.2×10^{-4} $\mathrm{m}^3/\mathrm{m}^3$ after 30 iterations. In this study, 30 iterations of the clustering algorithm is used.

The averaged spatial RMSE for each day of the year in the simulation period is shown in Figure 6. A Z-test was performed to evaluate how close the average disaggregated SM is to the mean than 0.04 $\mathrm{m}^3/\mathrm{m}^3$ for meaningful use in hydrological models [33]. This null hypothesis was found to be true for every day of the simulation period. Figure 7 shows the fraction of days for which the null-hypothesis is 1 with most of the days have an RMSE of less than 0.02 $\mathrm{m}^3/\mathrm{m}^3$. Figure 9 shows the disaggregated SM versus true SM at 1 km. Most of the points for sweet-corn and cotton pixels and all of the points for cotton lie within 0.04 $\mathrm{m}^3/\mathrm{m}^3$. Figure 8 shows the errors for each DoY segregated by type of land cover. Baresoil pixels during periods of vegetation have the highest RMSE. This is due to the effect of sub-pixel vegetation at 250 m in a pixel classified as a baresoil pixel at 1 km, 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. Baresoil pixels at 1 km without any vegetation at 250 m have the lowest KLD. Baresoil pixels at the end of the season, which are affected by remnant crops and baresoil pixels at 1 km which have some vegetation at 250 m, have a higher KLD, but very close to 0. Vegetated pixels at 1 km have a higher KLD as well. The boundary pixels classified as bare-soil have vegetation at the 250 m scale which contributes to these errors.

For the five selected days, the inputs, the first SM estimate, and PRI disaggregated SM are shown in Figures 10-13. Both DoY 39, shown in Figure 10 and DoY 354, shown in Figure 11 are during bare soil land cover 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 precipitation in the region (Figure 11e), the error for DoY 354 is higher than 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 Figures 11a, and 12a, for corn and cotton, when the LST is fairly uniform across the region, while PPT is heterogeneous due to precipitation 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 Figure 14.

C. Comparison between SRRM & PRI

The PRI method uses LST, 3-day PPT, LAI, LC, and SM at 1 km every 3 days as input to obtain the transformation function. To disaggregate SM, in Equation 18, $\mathbf{X}$ is set to $\{\mathrm{LST}, \mathrm{PPT}, \mathrm{LAI}, \mathrm{LC}\}$ and $\mathbf{y}_{\mathrm{TRAIN}}$ is set to $\{\mathrm{SM}_{IN-SITU}\}$. In this study, 33% of the data set is used for training the parametric Bayesian model. For the second step, in Equation 19, the SM observations at 10 km are set as $\mathbf{y}_{\mathrm{COARSE}}$ and first estimates of SM at 1 km from the transformation function are set as $\mathbf{y}_{\mathrm{INITIAL}}$. The value of M after the cost function $\mathrm{J}(\mathbf{m})$ is the disaggregated SM estimates.

The disaggregated estimates using the SRRM algorithm were compared with the PRI disaggregated estimates using the RMSE and the KLD of the estimated densities of the disaggregated observations. The spatial errors are also compared for the selected five days during the simulation period, representing different micro-meteorological and land cover conditions.

Figure 6 shows that the RMSE of the disaggregated observations using the self-regularized regressive models was less than the RMSE using the PRI algorithm. The trends observed when disaggregating SM using the PRI algorithm, such as the error being higher during periods of vegetation is preserved in the disaggregation results using self-regularized regressive models. 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 algorithm are shown by the PRI algorithm. However, the KLD for the SRRM estimates are 3 orders of magnitude lesser than for PRI estimates.

Figures 10-13 compares the disaggregated estimates using the PRI algorithm to the disaggregated estimates using the self-regularized regressive models. Equation 19, 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. Using multiple regressive models with soft boundaries ensures that sharpness is maintained with low RMSE.

IV. Conclusion

In this study, we implemented and evaluated a downscaling methodology based upon SRRM models that preserves the high variability in SM due to heterogeneous meteorological and vegetation conditions. The SRRM method 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 RS products, *viz.* PPT, LST, LAI, and LC. The kernel regression was implemented on the clusters using *in-situ* SM. 96 % of the pixels across the whole season was found to have a disaggregation error of less than $0.02\ \mathrm{m}^3/\mathrm{m}^3$. The KLD values for disaggregated SM at 1 km for the SRRM method was equal to 0, for all land covers. In contrast, the PRI method has KLD values several orders higher in magnitude. The averaged spatial error is also markedly lower for the SRRM method compared to the PRI method.

It is envisioned that the SRRM-based method implemented and evaluated in this study may be applied using satellite-based higher resolution remote sensing data. 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.

Acknowledgement

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.

References

[1] R. Tsai and T. Huang, "Recursive reconstruction of high-resolution image from noisy undersampled multiframes," *Adv. Comput. Vs. Image Process.*, vol. 1, pp. 317–339, 1984.

[2] H. Shen, M. Ng, P. Li, and L. Zhang, "Super-resolution reconstruction algorithm to MODIS remote sensing images," *The Computer Journal*, vol. 50, no. 1, pp. 90–100, 2009.

[3] F. Li, X. Jia, D. Fraser, and A. Lambert, "Super resolution for remote sensing images based on a universal hidden Markov tree model," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1270–1278, 2009.

[4] J. Hu, J. Gao, and X. Wang, "Multifractal analysis of sunspot time series: the effects of the 11-year cycle and Fourier truncation," *Journal of Statistical Mechanics: Theory and Experimentsis*, vol. 2009, no. 02, p. O02066, 2009.

[5] C. Hughes and M. Ramsey, "A radiometrically-accurate super-resolution approach to thermal infrared image data," *International Journal of Image and Data Fusion*, vol. 2009, pp. 1–23, 2012.

[6] M. Piles, A. Camps, M. Vall-llossera, I. Corbella, R. Panciera, C. Rudiger, Y. Kerr, and J. Walker, "Downscaling SMOS-derived soil moisture using MODIS visible/infrared data," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 9, pp. 3156–3166, 2011.

[7] S. Sanchez-Ruiz, M. Piles, N. Sanchez, J. Martinez-Fernandez, M. Vall-llossera, and A. Camps, "Downscaling SMOS-derived soil moisture using visible/infrared data," *Journal of Hydrology*, vol. 516, pp. 273–283, 2014.

[8] Z. Jiang, S. Shekhar, X. Zhou, J. Knight, and J. Corcoran, "Focal-test-based spatial decision tree learning: A summary of results," *Proceedings of the IEEE 13th International Conference on Data Mining (ICDM)*, pp. 320–329, 2013.

[9] J. Principe, D. Xu, and J. Fisher, *Information theoretic learning, in unsupervised adaptive filtering*. New York: Wiley, 2010.

[10] 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 Sensing*, vol. 53, no. 1, pp. 85–101, 2014.

[11] M. Fennessy and J. Shukla, "Impact of initial soil wetness on seasonal atmospheric prediction," *J. Clim.*, vol. 12, pp. 3167–3180, 2009.

[12] H. Douville and F. Chauvin, "Relevance of soil moisture for seasonal climate predictions: A preliminary study," *Clim. Dyn.*, vol. 16, pp. 719–736, 2000.

[13] The GLACE Team, R. Koster1, P. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C. Lu, S. Malyshev, B. McAvaney, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y. Sud, C. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada, "Regions of strong coupling between soil moisture and precipitation," *Science*, vol. 305, no. 5687, pp. 1138–1140, 2004.

[14] 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, pp. 259–270, 2002.

[15] J. Yuste, M. Nagy, I. Jenkins, I. Janssens, A. Carrara, and R. Ceulemans, "Soil respiration in a mixed temperate forest and its contribution to total ecosystem respiration," *Tree Physiol.*, vol. 25, no. 05, pp. 609–619, 2005.

[16] A. Friend, A. Arneth, N. Kiang, M. Lomass, J. Ogee, C. Rodenbeck, S. Running, J. Santaren, S. Sitch, N. Viovy, F. Woodwards, and S. Zaehle, "FLUXNET and modelling the global carbon cycle," *Global Change Biology*, vol. 13, pp. 613–633, 2007.

[17] 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. Hydrospheres Space Sci.*, vol. 17, no. 01, pp. 149–157, 1999.

[18] 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, 1999.

[19] B. Narasimhan and R. Srinivasan, "Development and evaluation of soil moisture deficit index (SMDI) and evapotranspiration deficit index (ETDI) for agricultural drought monitoring," *Agricultural and Forest Meteorology*, vol. 133, no. 1-4, pp. 69–88, 2005.

[20] 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*, vol. 7, no. 9, pp. 3867–3879, 2013.

[21] R. Jennsen, D. Erdogmus, K. Hild, J. Principe, and T. Eltoft, "Optimizing the Cauchy-Schwarz PDF distance for information theoretic, non-paremetric clustering," in *Proceedings of the 5th international conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 1. Proc. EMMCVPR 2005, 2005, pp. 34–45.

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

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

[24] T. Bongiovanni, R. DeRoo, J. Judge, Y. Goykhman, X. Duan, D. Preston, R. Shrestha, C. Slatton, M. Moghaddam, and A. England, "Field observations during the eighth microwave, water, and energy balance experiment (MicroWEX-8): from June 16 through August 24, 2009." http://edis.ifas.ufl.edu/ae476, Center for Remote Sensing, University of Florida, Tech. Rep., 2009.

[25] J. Casanova, F. Yan, M. Jang, J. Fernandez, J. Judge, C. Slatton, K. Calvin, T. Lin, O. Lanni, and L. W. Miller, "Field observations during the fifth microwave, water, and energy balance experiment (MicroWEX-5): from March 9 through May, 2006. Circular no. 1514," http://edis.ifas.ufl.edu/AE407, Center for Remote Sensing, University of Florida, Tech. Rep., 2006.

[26] T. Lin, J. Judge, K. Calvin, J. Casanova, M. Jang, O. Lanni, L. W. Miller, and F. Yan, "Field observations during the third microwave, water, and energy balance experiment (MicroWEX-3): from June 16 through December 21, 2004. Circular no. 1481," http://edis.ifas.ufl.edu/ae361, Center for Remote Sensing, University of Florida, Tech. Rep., 2004.

[27] (2011) Florida automated weather network. [Online]. Available: http://fawn.ifas.ufl.edu/

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

[29] 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. in Water Res.*, vol. 26, no. 2, pp. 137–149, 2003.

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

[31] J. Privette, R. Myneni, Y. Knyazikhin, M. Mukelabai, G. Roberts, Y. Tian, Y. Wang, and S. Leblanc, "Early spatial and temporal validation of MODIS LAI product in the Southern African Kalahari," *Remote Sens. Env.*, vol. 83, pp. 232–243, 2002.

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

[33] 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 Sensing*, vol. 49, no. 5, pp. 1504–1512, 2011.

LIST OF TABLES

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

TABLE II
DAYS SELECTED FOR EVALUATING PRI ESTIMATES. THESE DAYS CAPTURE VARIABILITY IN PRECIPITATION/IRRIGATION (PPT) AND LAND COVER (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

TABLE III
RMSE, SD, AND KL DIVERGENCE OVER THE 50×50 KM^2 REGION FOR THE DISAGGREGATED ESTIMATES OF SM OBTAINED AT 1 KM USING THE PRI AND UT METHODS.
A - BARESOIL PIXELS WITH VEGETATED SUB-PIXELS AT 250 M TILL DOY 332, B - BARESOIL PIXELS AFTER DOY 332
C - BARESOIL PIXELS WITHOUT ANY VEGETATED SUB-PIXELS AT 250 M TILL DOY 332

Land Cover	$\mathrm{KLD_{SRRM}}$	$\mathrm{KLD_{PRI}}$
Corn	1.8615×10^{-17}	0.0234
Cotton	2.4828×10^{-04}	0.0283
Baresoil$^{\mathrm{A}}$	5.6222×10^{-5}	0.1036
Baresoil$^{\mathrm{B}}$	5.628×10^{-6}	0.0120
Baresoil$^{\mathrm{C}}$	2.5948×10^{-6}	0.0114

LIST OF FIGURES

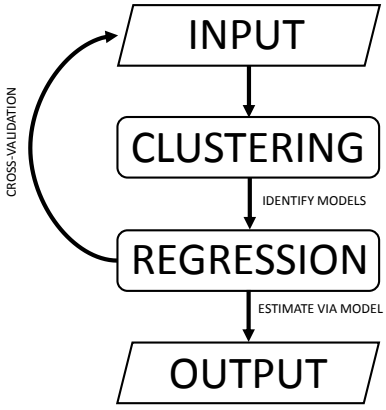

Fig. 1. Flowchart of the SRRM Method.

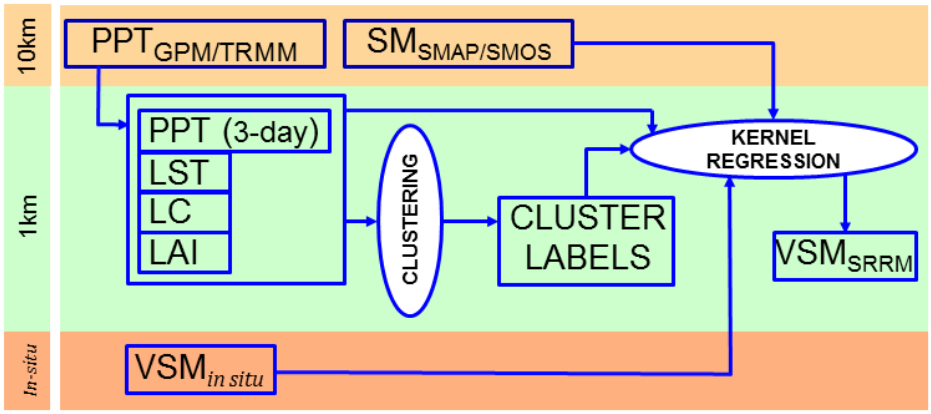

Fig. 2. Flow diagram of the Self-regularized Kernel Regression models.

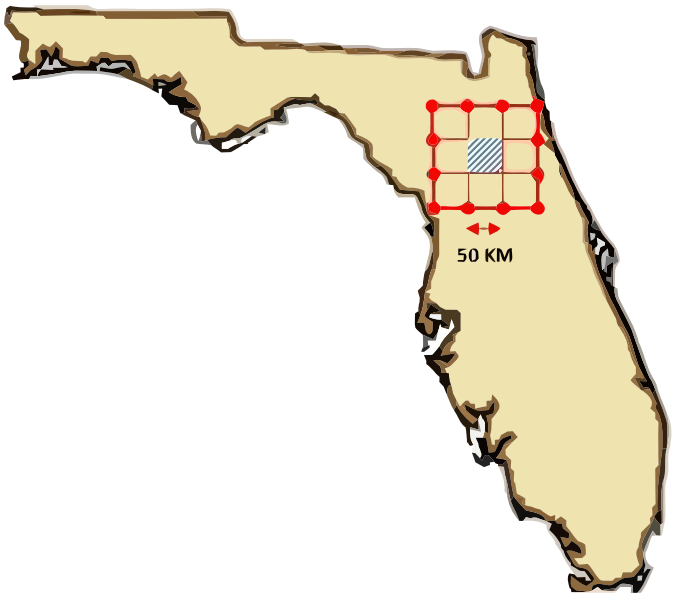

Fig. 3. Study region in North Central Florida. LSP-DSSAT-MB simulations were performed over the shaded 50×50 km^2 region.

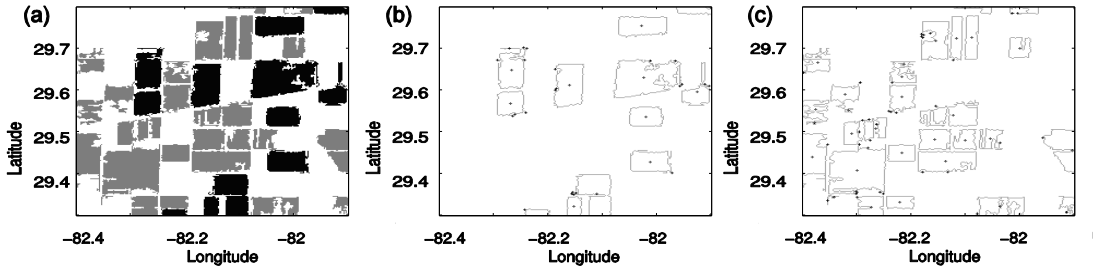

Fig. 4. (a) Land cover at 200m during cotton and corn seasons. White, gray, and black shades represent baresoil, cotton, and sweet-corn regions, respectively. Homogeneous crop fields along with centers for (b) sweet-corn and (c) cotton.

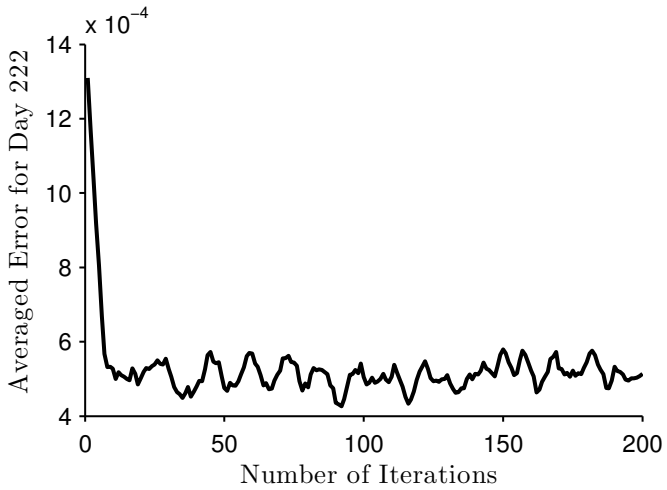

Fig. 5. Root mean Square error in disaggregated soil moisture at 1 km versus number of iterations of the D_{CS} clustering algorithm.

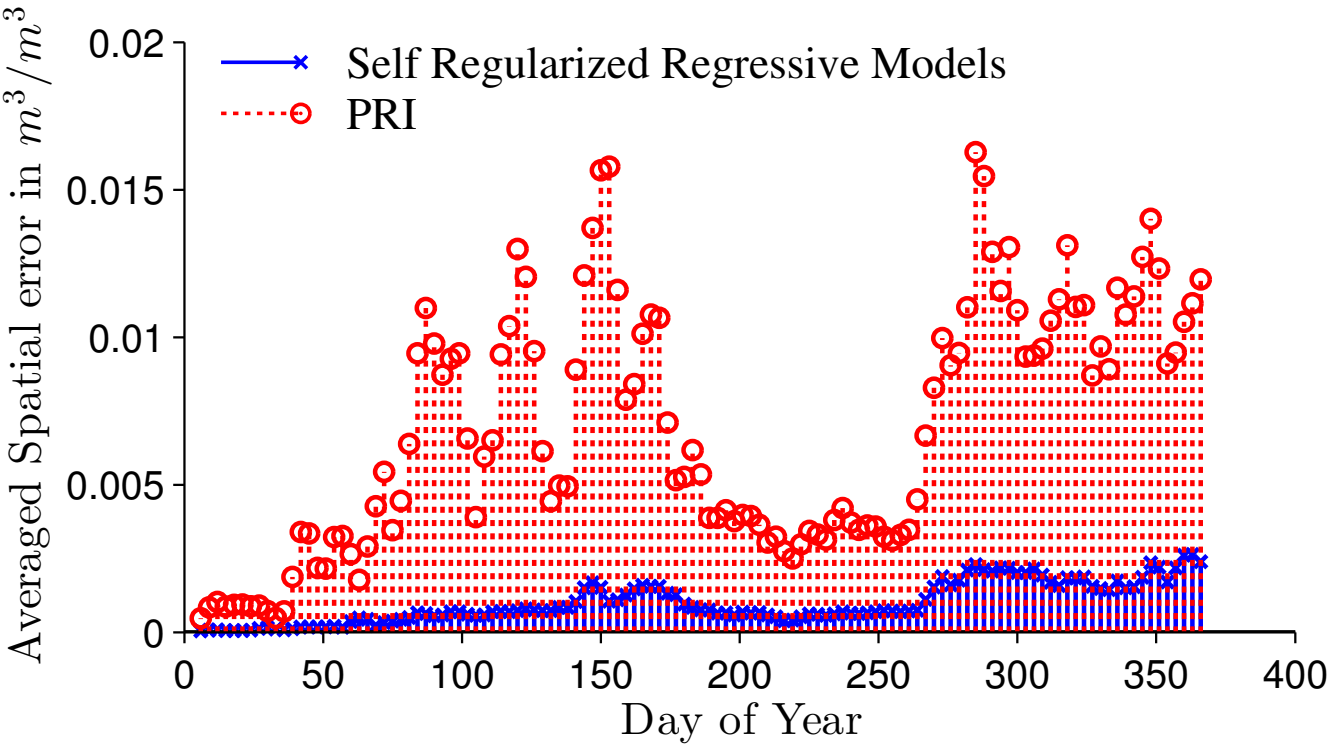

Fig. 6. Spatially averaged root mean square error in disaggregated Soil Moisture at 1 km for each day of the year in the simulation period.

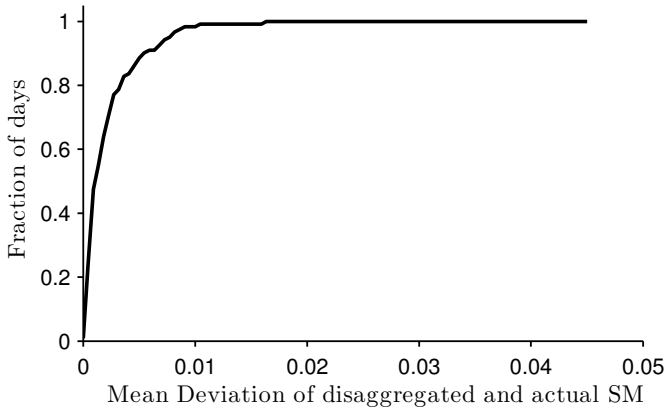

Fig. 7. Spatially averaged root mean square error in disaggregated Soil Moisture at 1 km for each day of the year in the simulation period.

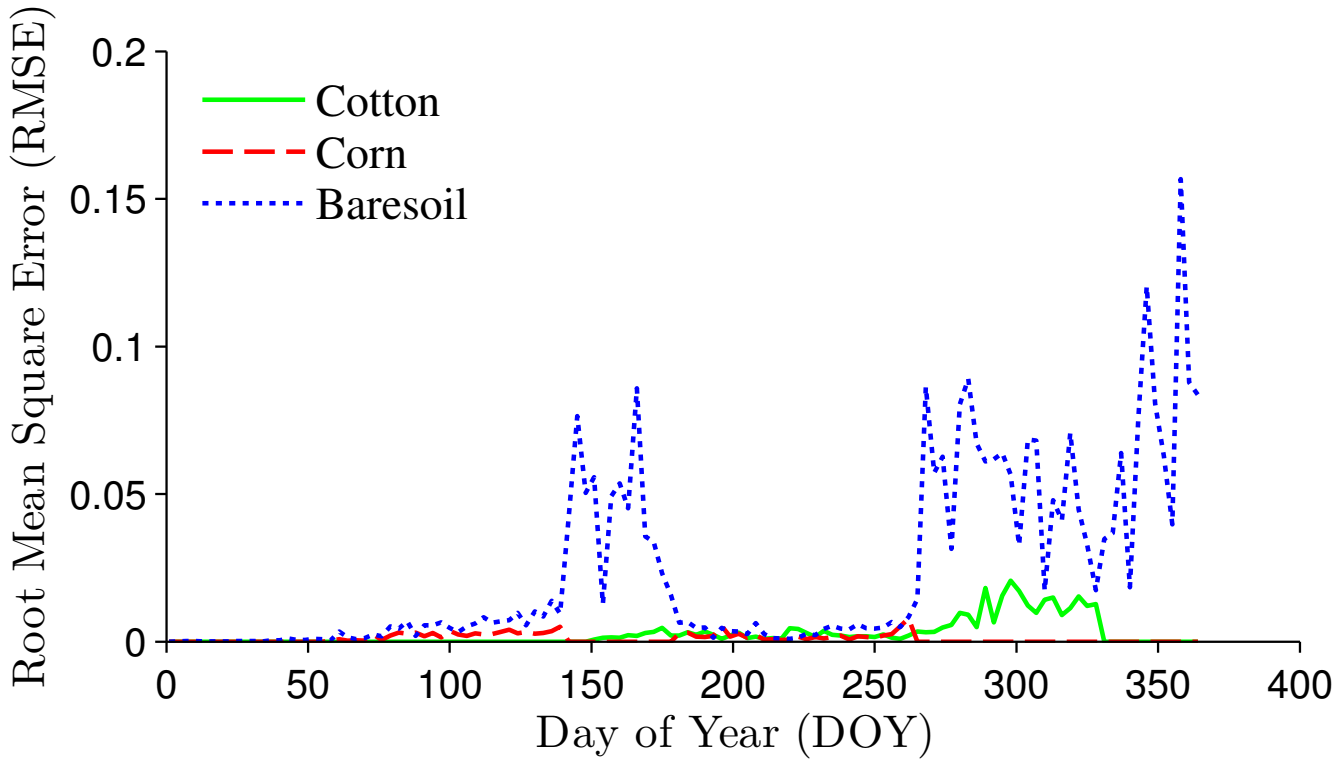

Fig. 8. Spatially averaged root mean square error in disaggregated Soil Moisture at 1 km for each day of the year in the simulation period.

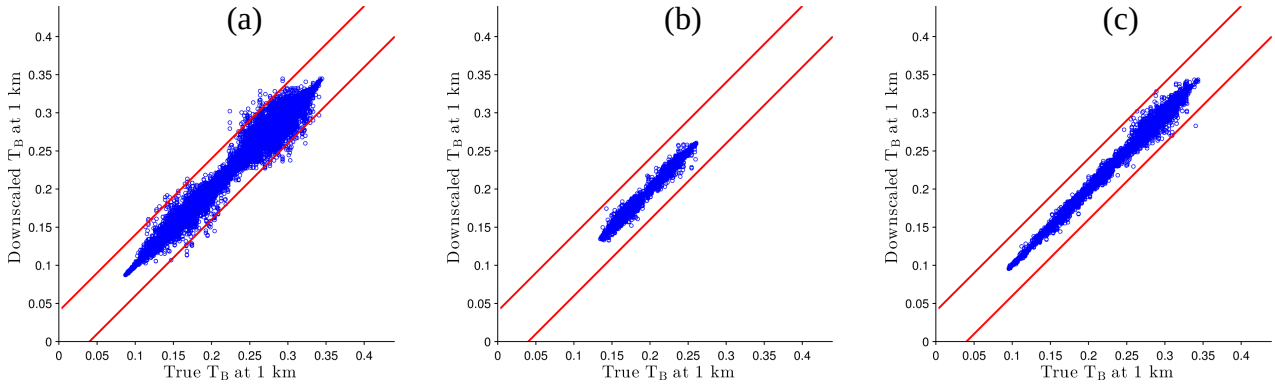

Fig. 9. Disaggregated Soil Moisture vs. True Soil Moisture at 1 km during the whole season for (a)baresoil pixels (b)corn pixels, and (c)cotton pixels. Lines corresponding to 4 % soil-moisture are shown for each plot.

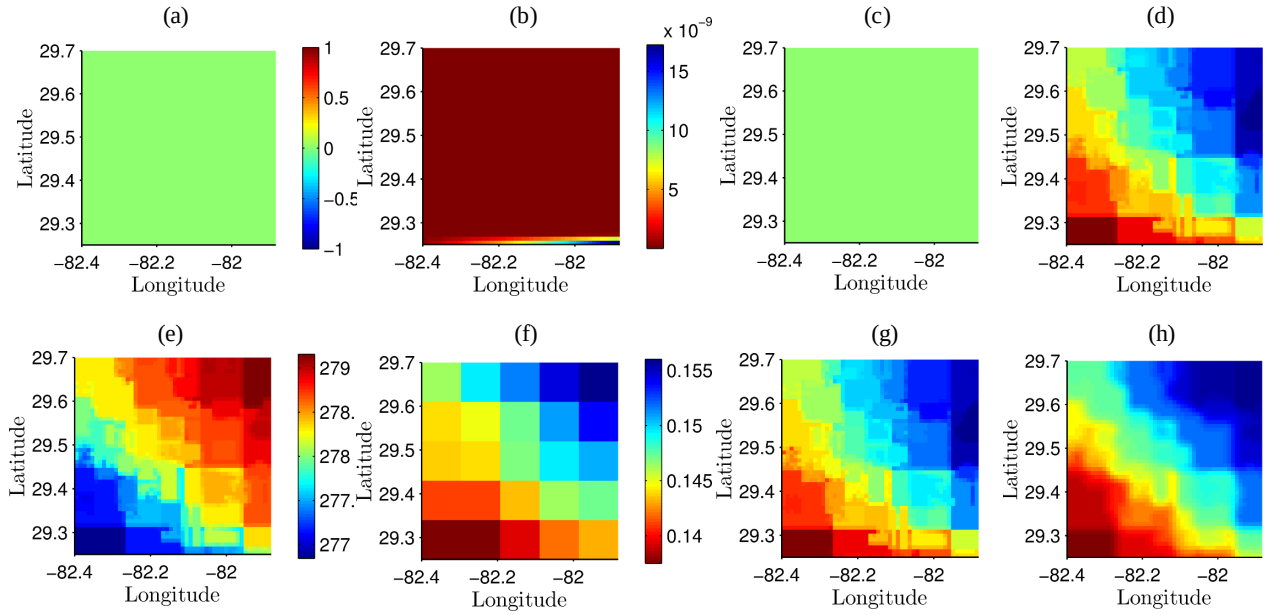

Fig. 10. DoY 39 - (a) LAI at 1 km, (b) PPT at 1 km, (c) LC at 1 km (green represents baresoil), (d) true SM at 1 km, (e) LST at 1 km, (f) SM observations at 10 km, (g) disaggregated SM using SRRM method, (h) disaggregated SM using PRI method.

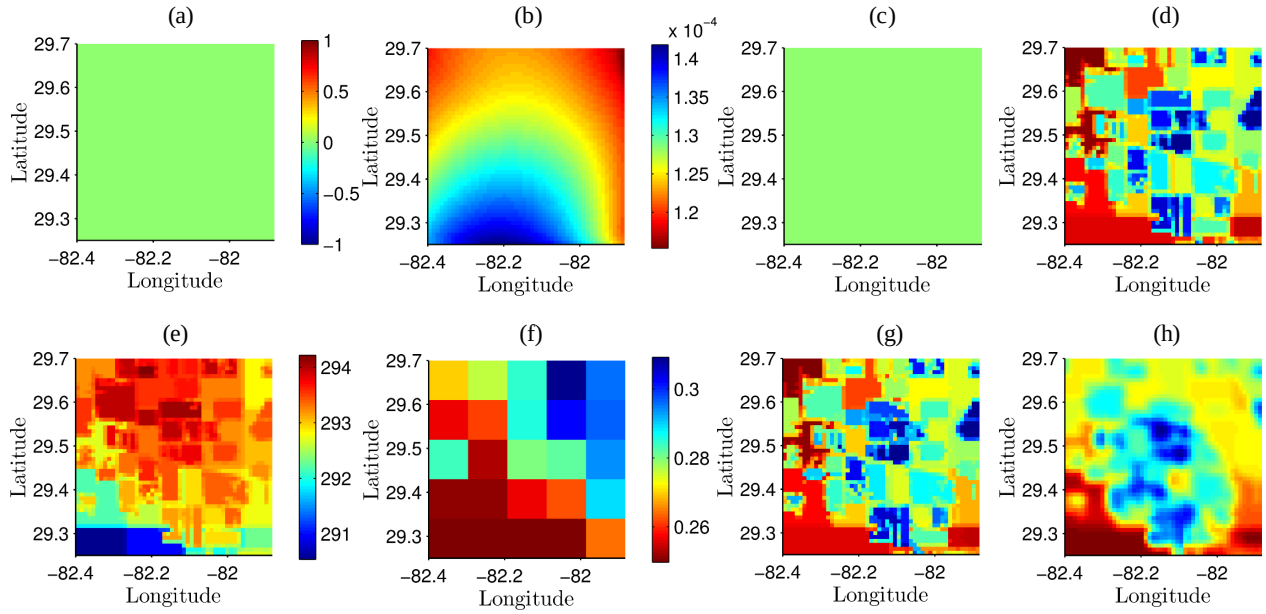

Fig. 11. DoY 354 - (a) LAI at 1 km, (b) PPT at 1 km, (c) LC at 1 km (green represents baresoil), (d) true SM at 1 km, (e) LST at 1 km, (f) SM observations at 10 km, (g) disaggregated SM using SRRM method, (h) disaggregated SM using PRI method.

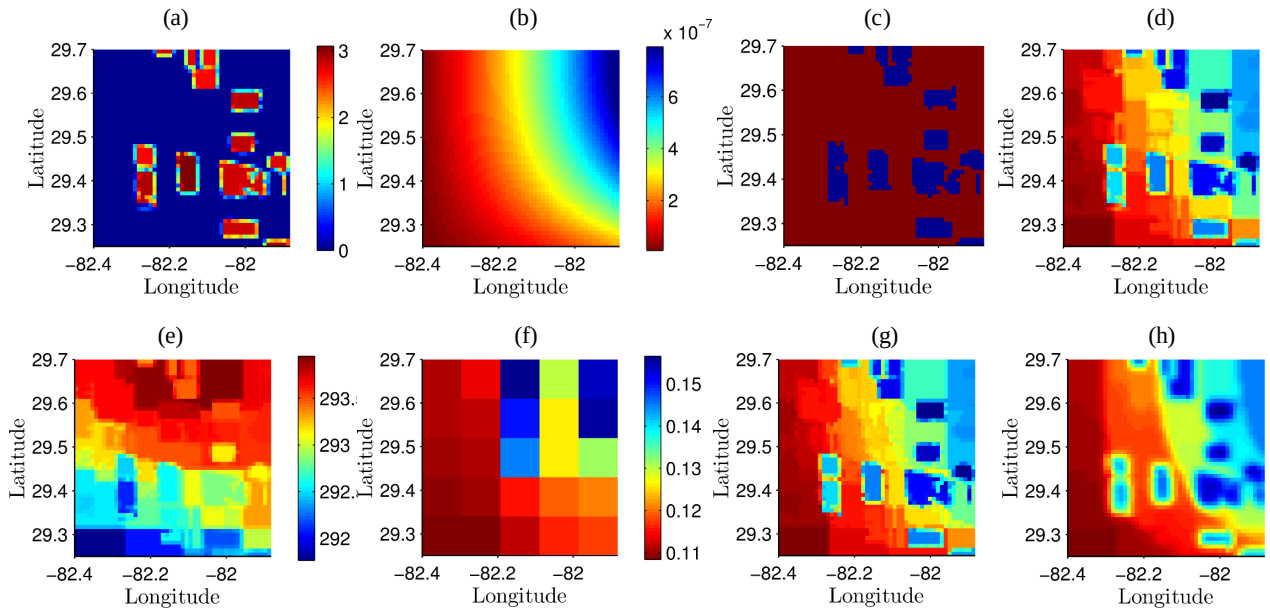

Fig. 12. DoY 135 - (a) LAI at 1 km, (b) PPT at 1 km, (c) LC at 1 km (brown represents baresoil and blue represents sweet-corn), (d) true SM at 1 km, (e) LST at 1 km, (f) SM observations at 10 km, (g) disaggregated SM using SRRM method, (h) disaggregated SM using PRI method.

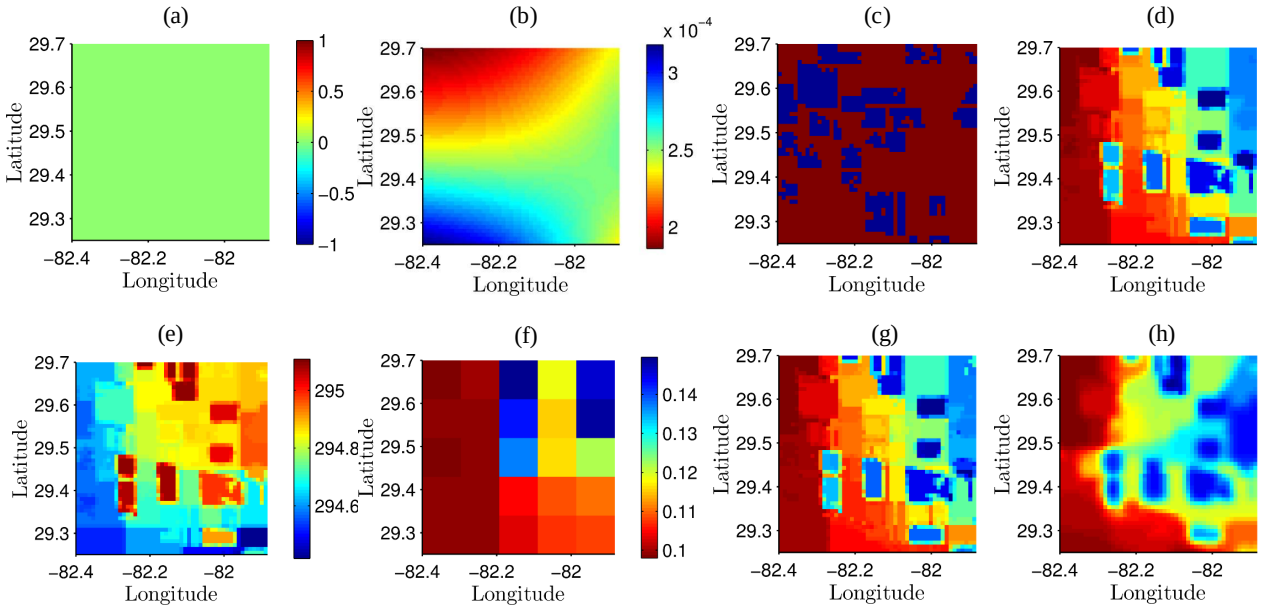

Fig. 13. DoY 156 - (a) LAI at 1 km, (b) PPT at 1 km, (c) LC at 1 km (brown represents baresoil and blue represents cotton), (d) true SM at 1 km, (e) LST at 1 km, (f) SM observations at 10 km, (g) disaggregated SM using SRRM method, (h) disaggregated SM using PRI method.

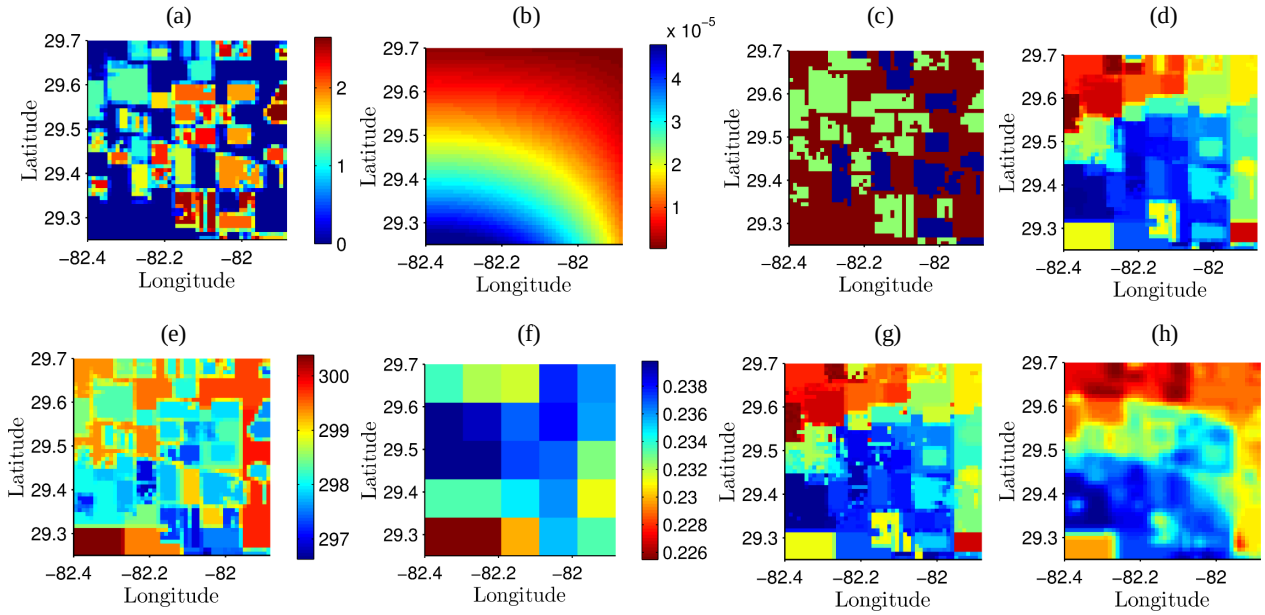

Fig. 14. DoY 222 - (a) LAI at 1 km, (b) PPT at 1 km, (c) LC at 1 km (brown represents baresoil, green represents cotton and blue represents sweet-corn), (d) true SM at 1 km, (e) LST at 1 km, (f) SM observations at 10 km, (g) disaggregated SM using SRRM method, (h) disaggregated SM using PRI method.