Utilizing Self-Regularized Regressive Models to Downscale Microwave Brightness Temperatures for Agricultural Land Covers in the SMAPVEX-12 Region

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

***Abstract*—A novel algorithm is developed to downscale microwave brightness temperatures (T_B), obtained at satellite scales of 10–40 to $\leq$1 km, meaningful for agricultural applications. Downscaling T_B directly bypasses the errors induced by inverse modeling encountered while downscaling satellite-based soil moisture products. This algorithm is based upon self-regularized regressive models (SRRM) and uses higher order correlations between auxiliary variables, such as precipitation (PPT), land cover, leaf area index, and land surface temperature, and horizontally polarized T_B observations. It includes information-theoretic clustering based on auxiliary variables to identify areas of similarity, followed by kernel regression that produces downscaled T_B. The algorithm was evaluated using a multiscale synthetic dataset over North Central Florida for one year, including two growing seasons of corn and one growing season of cotton. Compared to the true T_B, the downscaled T_B had a root-mean-square error (RMSE) of 5.76 K with standard deviation (SD) of 2.8 K during the growing seasons and an RMSE of 1.2 K with an SD of 0.9 K during nonvegetated. The SRRM algorithm effectively captured the variability in T_B at 1 km through the auxiliary variables. This algorithm was implemented to downscale SMOS observations available for five days during the SMAPVEX-12 experiment. Spatially averaged root-mean-square difference (RMSD) between the downscaled T_B and the airborne T_B observations from the airborne passive-active L-band sensor was 6.2 K, with Kullback–Leibler divergences of up to 0.91. For the SMAPVEX-12 dataset, better downscaling results are obtained for days when there was no PPT due to regional biases in the remotely sensed PPT from the NASA Tropical Measurement Mission. The RMSDs were lower when *in-situ* PPT data were used.**

***Index Terms*—Microwave radiometry, remote sensing, soil moisture.**

Manuscript received October 14, 2015; revised July 1, 2016 and September 30, 2016; accepted November 23, 2016. Date of publication January 18, 2017; date of current version January 23, 2017. This work was supported in part by the NASA-Terrestrial Hydrology Program-NNX13AD04G. *(Corresponding author: Subit Chakrabarti.)*

S. Chakrabarti and J. Judge are with the Center for Remote Sensing, Department of Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32611 USA (e-mail: subitc@ufl.edu; jasmeet@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/JSTARS.2016.2637927

I. Introduction

LOCAL and regional distribution of soil moisture (SM) heavily influences the global water cycle. Accurate knowledge of SM at spatial scales of 1–5 km is critical for characterizing evapotranspiration and estimating infiltration and run-off in agricultural and hydrological models. Microwave observations at frequencies $<$10 GHz are very sensitive to SM in the top 5–10 cm, due to large differences in dielectric constants of dry and wet soils, and have been widely used to retrieve SM [1], [2]. Current satellite-based microwave missions provide global-scale microwave observations. For instance, the European Space Agency's Soil Moisture Ocean Salinity (SMOS) mission provides passive observations at 25–50 km every two to three days. The NASA Soil Moisture Active Passive (SMAP) mission provides passive observations at 1.41 GHz at the spatial resolution of 36 km with a repeat coverage of two to three days [3]. However, the observations at these resolutions are still too coarse to integrate with crop models [2], reducing their effectiveness for agricultural applications, especially in areas that have heterogeneous land cover (LC) and meteorological conditions [4].

Most studies have downscaled SM derived from microwave observations (T_B) [1], [5]–[8], while very few studies [9] have downscaled satellite T_B observations directly to match model scales. The errors in downscaled SM due to subfootprint variability in LC, were found to be multimodal and, in some scenarios, had an root-mean-square error (RMSE) of 0.09 m^3/m^3, even for simple SM scenarios with limited variability, and ignoring errors due to model inadequacies, input parameter uncertainties, and sensor calibration errors [10]. Downscaling such a biased product will only increase the final error, even if the downscaling method corrects for physical and meteorological heterogeneity. Thus, downscaling T_B directly and then assimilating the downscaled product into hydrology models or crop growth models may significantly improve root zone SM and crop yield estimates [11]. Piles *et al.* [12] downscaled T_B directly into SM by applying the Universal Triangle (UT) method and used a second-degree regression-based linking model to relate coarse resolution SM to T_B from the SMOS mission, and other high-resolution products, aggregated to the resolution of SMOS observations. The fine-scale SM was then estimated using the

assumption that the linking model at the coarse resolution also holds at finer resolutions. The assumption of scale invariance has been found to result in high downscaling errors, particularly during heterogeneous LC conditions [1]. Das *et al.* [13] used the correlations between changes in passive radiometer observations and active radar backscatter using a time-series approach to obtain a merged SM product at 9 km. Some studies have used statistical inversion techniques such as linear inversion with regularization [14], singular value decomposition [15], and gradient descent in Banach spaces [9]. A major drawback of these approaches is the assumption of static vegetation conditions and the recalibration of empirical constant needed for different LCs. In addition, spatial scaling algorithms based upon second-order statistics can lead to significant loss of structural information in the data [2], especially under highly heterogeneous and dynamic conditions when the relationships between auxiliary data and SM are nonlinear. Studies involving heterogeneous and dynamic LC conditions are necessary to understand the validity of downscaling algorithms in agricultural regions.

In this study, a novel algorithm is presented that downscales $\mathrm{T_B}$ directly using higher order relationships between auxiliary data including satellite derived LST, leaf area index (LAI), precipitation (PPT), and LC at finer spatial resolutions. The algorithm, based on self-regularized regressive models (SRRMs), uses a combined clustering and regression approach to improve downscaling accuracies with smaller training sets. The goal of this study is to develop and implement the SRRM-based downscaling algorithm that directly disaggregates coarse-scale $\mathrm{T_B}$ using auxiliary fine-scale remotely sensed data. The primary objectives are to develop an algorithm that downscaled $\mathrm{T_B}$ to 1 km using coarse-scale $\mathrm{T_B}$ and other spatially correlated variables, validate the downscaling algorithm using a year-long multiscale synthetic dataset [16] based in North Central Florida, and implement the algorithm to downscale H-Pol $\mathrm{T_B}$ from SMOS to 1 km for four days in the agricultural regions of the SMAP Validation Experiment-2012 (SMAPVEX-12).

The rest of this paper is structured as follows. Section II describes the theoretical details of the disaggregation framework based on SRRMs. Section III describes the synthetic dataset and the SMAPVEX-12 observations for the algorithm and the SMAPVEX-12 experiment. Section IV discusses the steps for the SRRM-based algorithm, and Section V presents the disaggregation results for $\mathrm{T_B}$ at 1 km. Finally, Section VI summarizes the important results and concludes this paper.

II. Downscaling Framework

Downscaling is an ill-conditioned problem that is limited 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-based algorithm addresses this problem by employing a clustering algorithm to create a number of regions of similarity, which subsequently are used in a kernel regression framework.

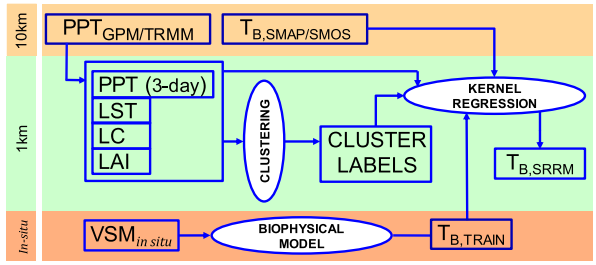

Fig. 1. Flowchart of the SRRM method.

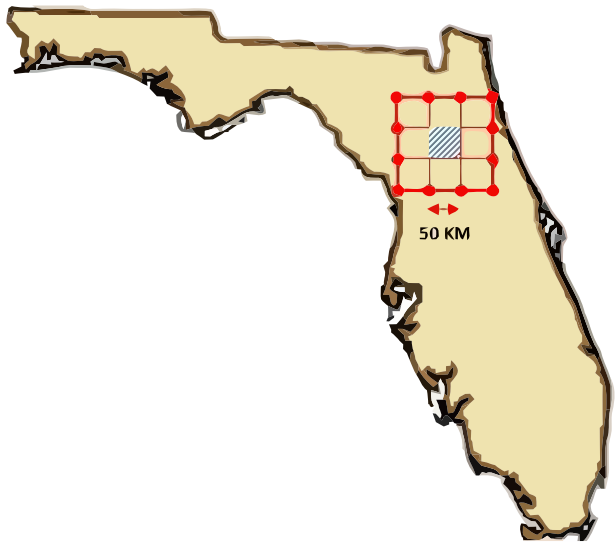

Fig. 2. Study region in North Central Florida. LSP-DSSAT-MB simulations were performed over the shaded 50×50 km^2 region.

Fig. 1 shows the overall organization and datasets used in the algorithm.

A. Self-Regularized Regressive Models

The first step of the algorithm clusters the study area into proximity regions using the auxiliary features. In this study, the clustering algorithm uses information-theoretic measures of inter- and intracluster similarity [17]. Fractional membership vectors are used to decrease downscaling errors along cluster boundaries. For example, if $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N\}$ is a matrix containing auxiliary features for N pixels, the Cauchy–Schwarz cost function, $\hat{J}_{\mathrm{CS}}$, estimates optimal memberships of the pixels to clusters, $\mathbf{m}$, in an unsupervised manner. The cost function is regularized using the weighted Shannon entropy of the membership vector, such that the membership vectors are sufficiently sparse, as shown in the following equation:

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

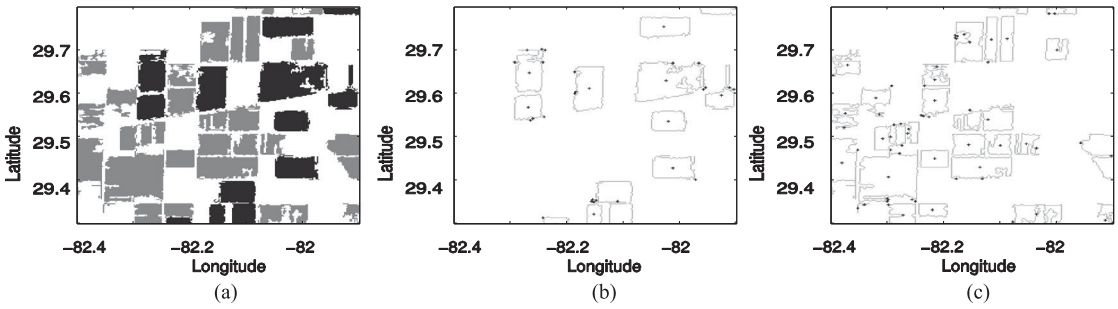

Fig. 3. (a) Land cover at 200 m during cotton and corn seasons. White, gray, and black shades represent bare soil, cotton, and sweet corn regions, respectively. Homogeneous crop fields along with centers for (b) sweet corn and (c) cotton.

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

where K is the number of clusters, $G_{\sigma\sqrt{2}}$ is the Gaussian kernel with standard deviation (SD) σ, and ν is the regularization weight. The regularization weight ν in (1) counterbalances the first term of the equation that increases the separation between the clusters discovered against the second term of the equation that promotes sparsity in the membership vector. The optimal value of the membership vector can be obtained from the following constrained optimization problem:

$$\begin{aligned} &\min_{\mathbf{m}_1,\ldots,\mathbf{m}_N} \hat{J}_{\mathrm{CS}}^{\mathrm{REG}}(\mathbf{m}_1,\ldots,\mathbf{m}_N) \\ &\text{subject to } \mathbf{m}_j^{\mathrm{T}}\mathbf{1}-1=0, \quad j=1,\ldots,N. \end{aligned} \tag{2}$$

To compute optimum values of $\mathbf{m}$, and thus the membership of each pixel to the K clusters, a Lagrange multiplier formulation can be used along with a stochastic gradient descent scheme, the details of which are shown in [18].

In the second step, ridge regression [19], a kernel-based parametric regression technique, is used to generate the downscaled estimates. A training set of pixels is used in the regression to fit a nonlinear function from the set of the auxiliary data and coarse-scale $\mathrm{T_B}$ to fine-scale $\mathrm{T_B}$. The cost function of ridge regression is

$$\mathcal{E}(\mathbf{w},\mathbf{x}) = \frac{1}{2}\sum_i (y_i - \mathbf{w}^{\mathbf{T}}\mathbf{x}_i)^2 + \frac{1}{2}\mu\|\mathbf{w}\|^2 \tag{3}$$

where y_i is the true $\mathrm{T_B}$, $\mathbf{w}^{\mathbf{T}}\mathbf{x}_i$ is the downscaled $\mathrm{T_B}$, $\mathbf{w}$ is the weight vector, and μ is the regularization constant. The cost function includes an $\mathrm{L_2}$-norm regularizing term, which improves the stability of the regression by inducing smaller eigenvalues. The weights can be calculated by differentiating the error cost function with respect to the weights and setting it to zero. The inner products can be replaced with a kernel evaluation if this computation is conducted in a reproducing kernel Hilbert space (RKHS). 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 $\Phi : \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}^{\mathrm{N}}$ is a mapping that transforms the feature vector in the original vector space to $\mathcal{H}$, then the weights can be redefined as

$$\mathbf{w} = (\mu\mathbf{I}_D + \Phi\Phi^{\mathrm{T}})^{-1}\Phi\mathbf{y}. \tag{4}$$

Here, D is the dimension of the transformed feature space, and Φ maps the points from the original feature space to the RKHS. The estimated value of the downscaled $\mathrm{T_B}$, $\hat{y}$, for a data point not in the training set $\mathbf{x}'$ is

$$\hat{y} = \mathbf{w}^{\mathrm{T}}\Phi(\mathbf{x}') = \mathbf{y}(\mu\mathbf{I}_N + \mathrm{K})^{-1}\kappa(\mathbf{x},\mathbf{x}') \tag{5}$$

where K is the Gram matrix of inner products of all the training data points. The feature vector is augmented by adding a constant feature 1 to all samples to address the regression bias. More details about the algorithm can be found in [18] and [19].

III. DATASETS

A. Multiscale Synthetic Dataset

The downscaling algorithm in the study used dataset from the simulation framework consisting of a soil–vegetation–atmosphere transfer model, the land surface process (LSP) model, coupled with a crop growth model, and the decision support system for agrotechnology transfer (DSSAT) model, described in [16]. A 50 $\times$ 50 km^2 region, equivalent to approximately 25 SMAP pixels with a spatial resolution 9 km/pixel, was chosen in North Central Florida (see Fig. 2) for the simulations. The region encompassed the UF/IFAS 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, have been conducted for various agricultural LCs over the past decade [20]–[22]. Simulated observations of LST and LAI were generated at 200 m for one year, from January 1, 2007 through December 31, 2007. Topographic features, such as slope, are constant as 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.

For the LSP-DSSAT model simulations, 15-min observations of PPT, relative humidity, air temperature, downwelling solar radiation, and wind speed were obtained from eight Florida Automated Weather Network stations [23] located within the

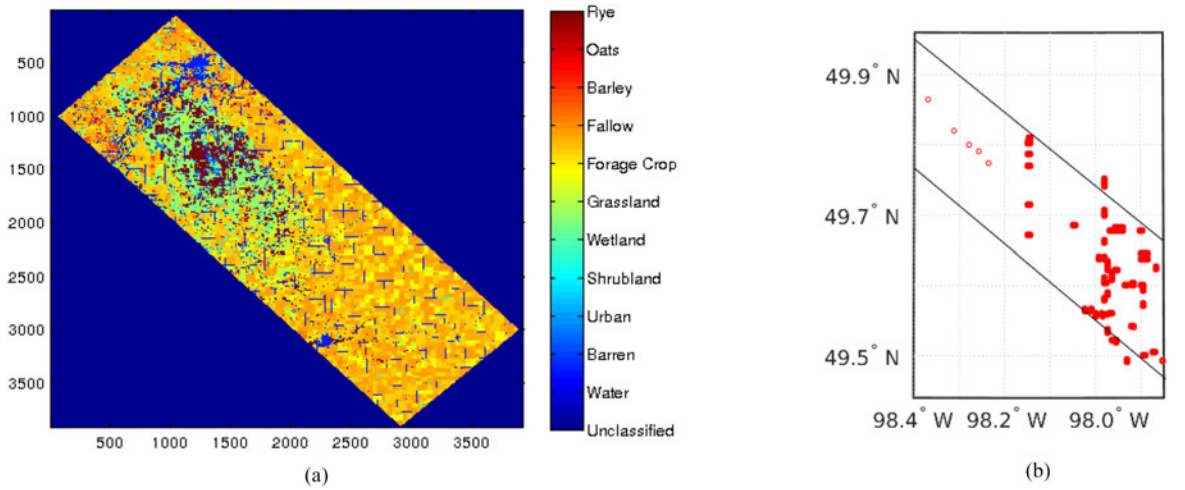

Fig. 4. (a) Land Cover Map for SMAPVEX-12. (b) *In-situ* stations used during the SMAPVEX-12. The black lines denote the boundaries of the agricultural site.

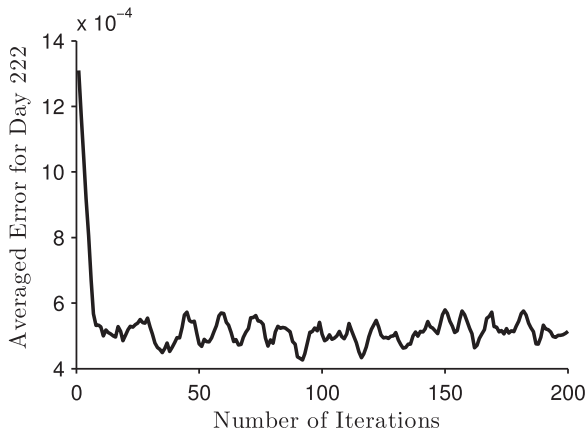

Fig. 5. RMSE in disaggregated T_B at 1 km and number of iterations of the D_{CS} clustering algorithm.

study region. The observations were spatially interpolated using splines to generate the meteorological forcings at 200 m. Long-wave radiation was estimated following [24]. The model simulations were performed over each contiguous homogeneous region of sweet corn, bare soil, and cotton, as shown in Fig. 3, rather than all the pixels, to reduce computation time. A realization of the LSP-DSSAT model was used to simulate LST and LAI 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 Fig. 3.

The T_B for this study is simulated using the widely used τ–ω [25] model. In the model, the vegetated surface is a single isothermal layer of vegetation with diffuse boundaries [16]. The soil medium is assumed to be nonisothermal semi-infinite layered dielectric medium with a rough surface at the upper boundary. Using a zeroth-order radiative transfer approach, the total T_B of a terrain is the sum of contributions from soil, vegetation, and from sky. In this study, similar to [1], the T_{sky} was set to 5 K, r_p was obtained by integrating the bistatic scattering coefficients from the IEM model, the RMS height was 0.62 cm, and the correlation length was 8.72 cm [25]. Other model parameters are set similar to [16]. The SM and temperature profiles, the vegetation biomass, and plant height provided by the LSP-DSSAT model were used by the MB model to estimate H-Pol T_B at L-band.

The model simulations at 200 m were spatially averaged to obtain PPT, LST, LAI, and T_B at 1 and 10 km. Linear averaging is typically sufficient to illustrate the effects of resolution degradation [26]. About 33% of the T_B obtained at 1 km served as simulated *in-situ* measurements and were used for training. The rest of the values were used as the "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.

The simulation period in 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 and the T_B observations at 10 km were obtained by adding white Gaussian noise to the model simulations account for satellite observation errors, instrument measurement errors, and micrometeorological variability, following [27]–[29]. The errors added had zero mean and SDs of 5 K, 1 mm/h, 5 K, and 0.1 for LST, PPT, LAI at 1 km, and T_B at 10 km, respectively.

B. SMAPVEX-12

The downscaling methodology is implemented in the SMAPVEX-12 region. The experiment was conducted from June 6 to July 17, 2012 in a region south of Winnipeg, MB, Canada, with a range of LCs including crops such as cereals, canola, corn, and soybean; some permanent grassland, wetlands, and mixed forest cover. The agricultural region is about 15 km $\times$ 70 km within the Red River Watershed, shown in Fig. 4. During the experiment, spaceborne microwave measurements from SMOS, AMSR-E, WindSAT, and RADARSAT-2 were collected along with airborne measurements from the NASA

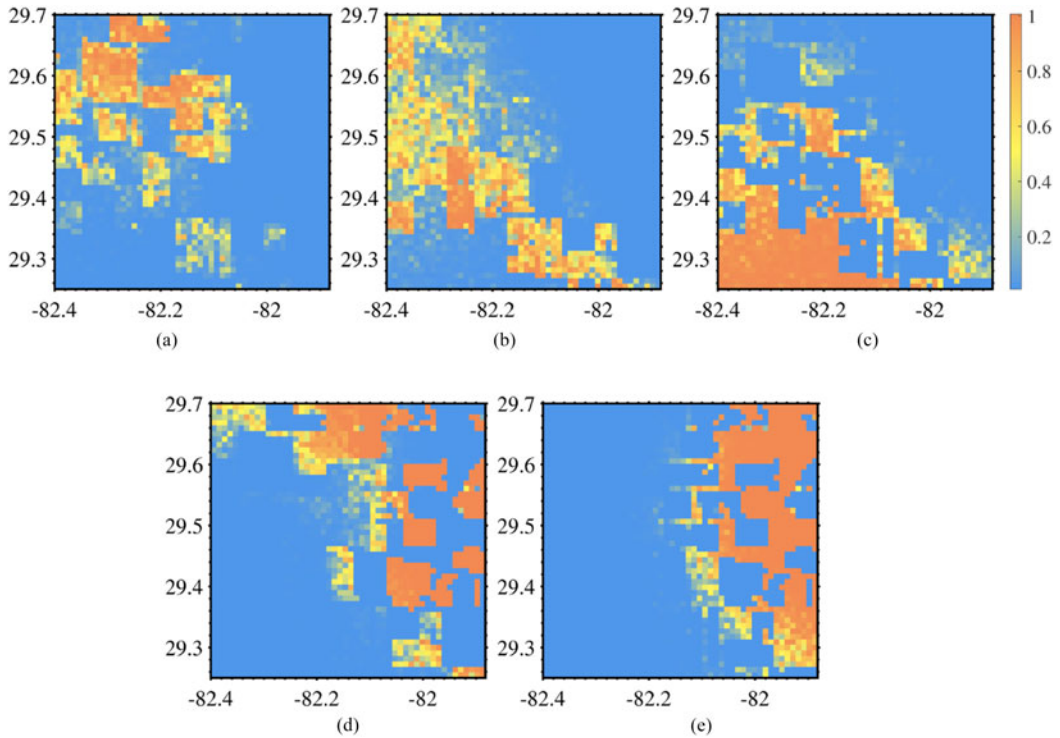

Fig. 6. (a)–(f) Membership fractions of five clusters on DoY 222.

TABLE II
DAYS SELECTED FOR EVALUATING SRRM ESTIMATES

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

These days capture variability in precipitation/irrigation (PPT) and land cover (LC).

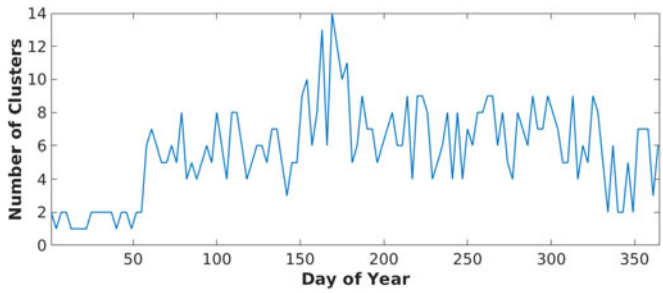

Fig. 7. Number of clusters for each day of the synthetic experiment.

L-band unmanned aerial vehicle synthetic aperture radar and the passive/active L-band sensor (PALS). In addition to the existing Agriculture and Agri-Food Canada (AAFC) permanent stations in the area, about 50 other temporary stations were used from the United States Department of Agriculture (USDA), AAFC, and Manitoba Agriculture, Food and Rural Initiatives (MAFRI). These stations provided micrometeorological conditions, SM, soil temperature, and land characteristics such as vegetation, roughness, and soil density at the times of satellite and air-plane overpasses.

For the implementation of the SRRM-based algorithm, the LAI and LST were obtained from the Moderate Resolution Imaging Spectroradiometer, aboard the NASA Terra and Aqua satellites, the coarse resolution H-Pol T_B was obtained from SMOS, and PPT was obtained from both the Tropical Rainfall Measurement Mission (TRMM) and the measurements at the USDA, AAFC, and MAFRI stations in the SMAPVEX-12 region.

IV. METHODOLOGY

A. Synthetic Study

The SRRM-based algorithm uses LST, 3-day PPT, LAI, LC at 1 km, and T_B at 10 km every three days as inputs. The Cauchy–Schwarz cost function described in Section II-A is used to cluster the region using the auxiliary inputs at 1 km, with the x and y coordinates of each pixel scaled from 0 to 1. The clustering algorithm uses two parameters—the number of clusters K and a regularization constant μ—that are determined by cross-validating against the absolute mean error in T_B following the regression for each day.

The optimal number of iterations that produced a suitable clustering result for downscaling was determined by the minimum RMSE for DoY 222, a day when both the LC and micrometeorological conditions were the most heterogeneous, providing the worst-case scenario for convergence of the clustering algorithm. Fig. 5 shows the spatially averaged RMSE between

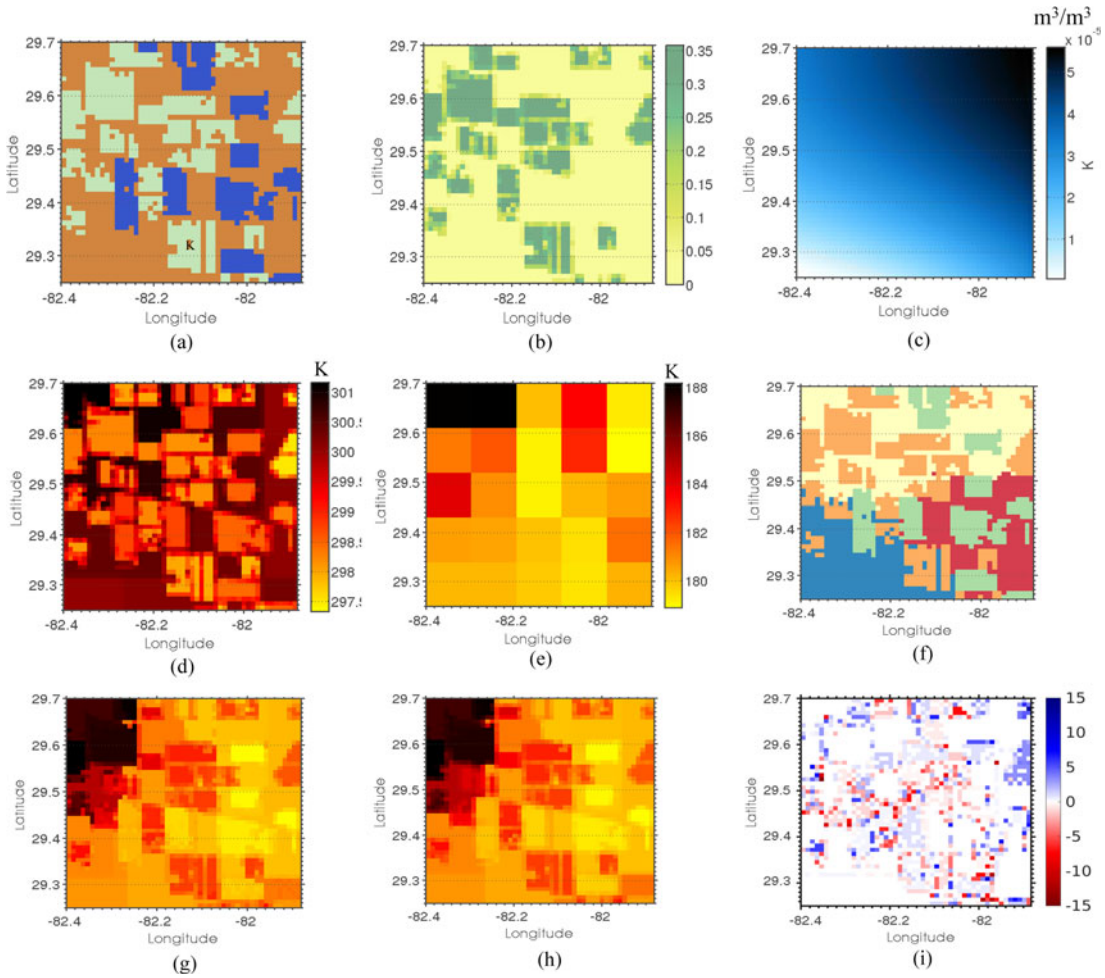

Fig. 8. DoY 222. (a) LC at 1 km (brown represents bare soil, green represents cotton, and blue represents sweet corn). (b) LAI at 1 km. (c) PPT at 1 km. (d) LST at 1 km. (e) T_B observations at 10 km. (f) Cluster memberships at 1 km. (g) True T_B at 1 km. (h) Disaggregated T_B using the SRRM method at 1 km. (i) Difference between the true T_B and downscaled T_B.

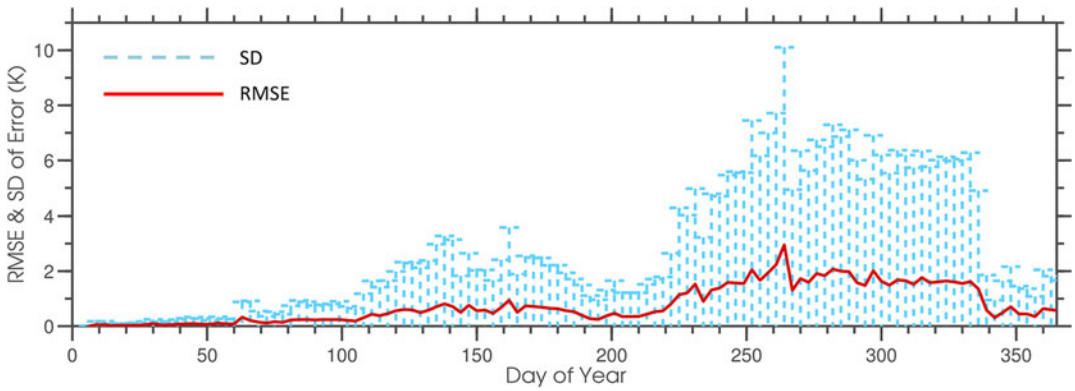

Fig. 9. RMSE and SD of error in downscaled T_B at 1 km.

downscaled T_B and the observations at 1 km on DoY 222 for different iterations of the clustering algorithms. All parameters were cross-validated for each individual iteration, except for the number of clusters, which were cross-validated once, using 50 iterations of the clustering algorithm. The error oscillated around a value of 5×10^{-4} K with a mean amplitude of 1.2×10^{-4} K after 30 iterations. In this study, 30 iterations of the clustering algorithm are used. After clustering, each pixel has a vector of K numbers $(m_1, m_2, \ldots, m_K)$ that sum to 1 describing its membership to each of the K clusters. The mem-

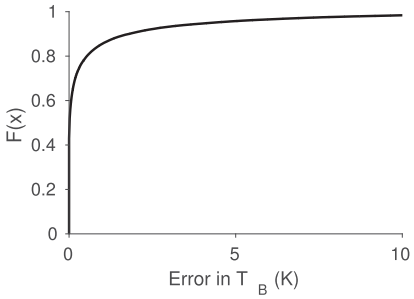

Fig. 10. Empirical CDF downscaled T_B at 1 km.

berships of the pixels in the study region to each of the five clusters are shown on DoY 222 in Fig. 6(a)–(f), respectively. The pixels at the center of the corn field at latitude 29.6 °N and longitude 82.3 °W belong to cluster 1, while the boundary pixels belong to clusters 1 and 3 with proportions of 0.6 and 0.4 proportionally. This proportional representation is needed because the boundary pixels exhibit characteristics of both bare soil and corn LCs.

During regression, K models, $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_K$, are developed, one for each cluster, using LST, three-day PPT, LAI, LC, T_B at 1 km, and T_B at 10 km as inputs to the regularized kernel regression algorithm using a randomly selected training set comprised of 33% of all pixels. The hard membership of each pixel, i, for model development purposes is determined by the maximum value in its membership vector, $\mathbf{m}^i = (m_1^i, m_2^i, \ldots, m_K^i)$. The downscaled value of T_B is computed for each point in the test, represented as a vector $\mathbf{x'}_i = (\mathrm{LST}_i^{1\ \mathrm{km}}, \mathrm{PPT}_i^{1\ \mathrm{km}}, \mathrm{LAI}_i^{1\ \mathrm{km}}, \mathrm{LC}_i^{1\ \mathrm{km}}, {T_B}_i^{10\ \mathrm{km}})$ by

$$ {T_B}_i^{1\mathrm{km}} = \mathbf{m}^{\mathrm{T}} \cdot \left(\hat{f}_1(\mathbf{x'}_i), \hat{f}_2(\mathbf{x'}_i), \ldots, \hat{f}_K(\mathbf{x'}_i) \right). \quad (6) $$

The SRRM-based algorithm is evaluated using the RMSE and SD of the errors over the entire season for each LC. Moreover, the downscaled T_B is compared to the true T_B. The Kullback–Liebler divergence (KLD) between the probability density function (PDF) of the estimated observations and the true T_B is calculated for different LCs over the season. The KLD is a member of the class of f-divergences that convey distances in the probability space. In this study, it is used to evaluate the proximity of the PDF of the downscaled estimates to the PDF of the true T_B. The effect of the heterogeneity in inputs on the error in downscaled T_B is investigated on DOY 222, when the heterogeneity in LC and meteorological conditions is maximum. In addition, four other days were also selected to understand the effect of the heterogeneity in inputs on the error in downscaled T_B. Variabilities in precipitation, 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. In this study, we use quantitative analyses of spatial variations in T_B, such as RMSE and KLD, and heterogeneous LC conditions to provide an index of expected errors in downscaled T_B. Finally, the utility and effectiveness of multiple regression models are evaluated by downscaling T_B without clustering, thus using a single regression model similar to UT-based methods [6], and comparing the result to the downscaled T_B with the clustering step.

B. Implementation on the SMAPVEX-12 Dataset

Downscaling was performed for four days during the experiment when coarse SMOS T_B was available. The clustering algorithm groups the LAI, LST, LC, PPT, along with the x and y coordinates of the pixels in the region using the same methodology as that used with the synthetic dataset. Fine-scale *in-situ* T_B required for training is obtained from the 800-m PALS observations at the 42 sites where the AAFC SM sampling stations were located, as shown in Fig. 4(b). Additional 1008 pixels of PALS T_B were used to compare with the downscaled T_B. In an operational scenario, the fine-scale T_B estimates can be obtained from *in-situ* SM via models such as the MB model. The SMOS T_B is downscaled for all the five days using the SRRM-based algorithm to 1 km, and the root mean square difference (RMSD) is calculated between the downscaled T_B and the PALS T_B as a measure of the efficacy of the downscaling algorithm. The KLD of the differences was calculated to understand how the PDF of the downscaled T_B differs from the PDF of the PALS T_B.

V. Results

A. Synthetic Dataset

The Cauchy–Schwarz based clustering algorithm clusters the study region to identify subareas where a single downscaling regression model may be applied. The number of clusters is proportional to the complexity and heterogeneity in the auxiliary dataset. The optimal number of clusters for each day of the experiment is shown in Fig. 7. The number of clusters depends upon the spatial heterogeneity of all the features and not just the LC. For example, while the LC on both DoY 1 and DoY 364 is bare soil, the PPT on DoY 364 is nonzero unlike DoY 1 that has no PPT. In addition, the LAI on DoY 1 is zero, while on DoY 364, the LAI is nonzero due to crop remnants. Thus, the numbers of clusters on both the days is different. The number of clusters is consistently higher for the growing seasons of corn and cotton than during nonvegetated periods. The highest number of clusters, 14, is for the periods when both the LCs exist simultaneously. The inputs, clusters, true T_B, and the downscaled T_B at 1 km for DoY 222 are shown in Fig. 8. It should be noted that the coarse T_B is not the exact average of the true fine-scale T_B. The 10-km observations were obtained by averaging all the 1-km pixels inside a 10×10 km^2 pixel and adding Gaussian noise with 0 mean and SD of 5 K, as mentioned in Section IV. For example, in Fig. 8(e) and (g), values of coarse-scale T_B at latitude 29.4 °N and longitude 81.9 °W are higher than the average true T_B of the 1-km pixels in the region. The clustering in Fig. 8(e) largely follows the LC pattern, with all the cotton and the corn pixels assigned to different clusters. The algorithm is also affected by PPT and successfully demarcates areas with low and high PPT, as shown in Fig. 8(d) and (e). The differences between the true and downscaled T_B are low, with a median value of 2 K. The discontinuity in the true T_B in the North–West part of the region is also replicated

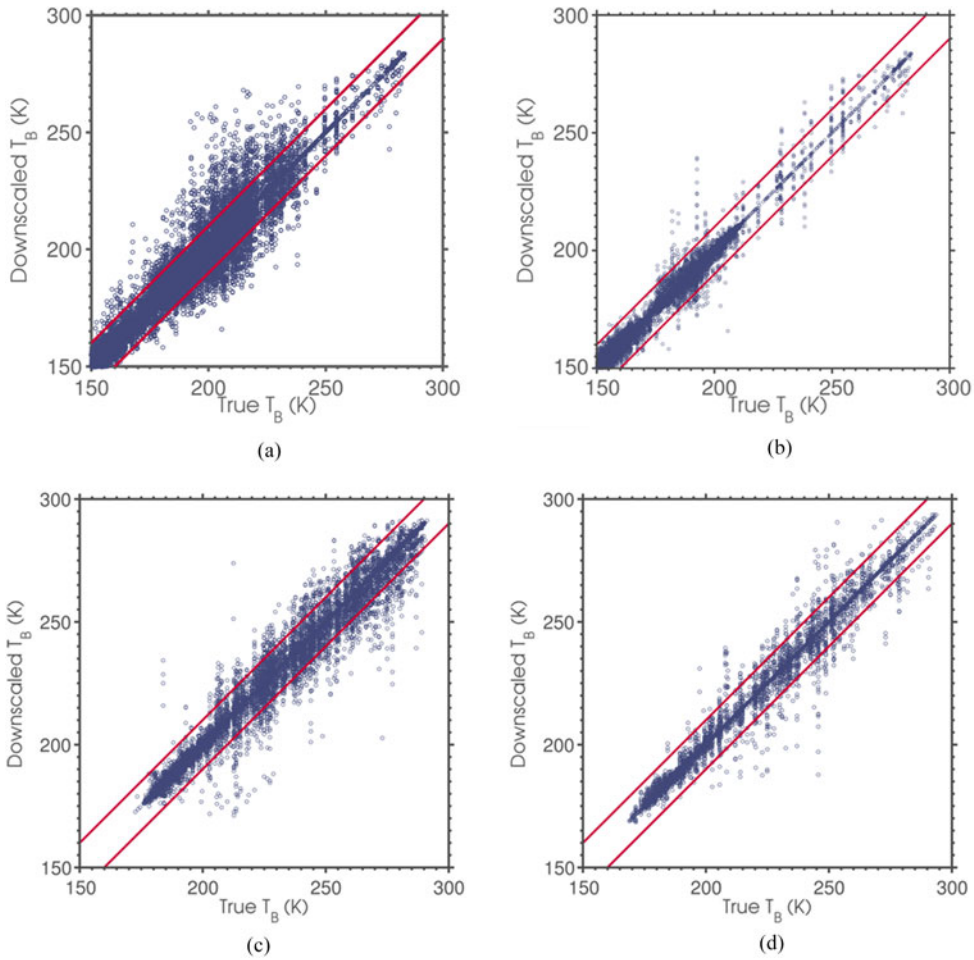

Fig. 11. Downscaled T_B as a function of true T_B at 1 km during the whole season for (a) bare-soil pixels with partially vegetated subpixels (vegetation fraction <0.5), (b) bare-soil pixels without vegetated subpixels, (c) corn pixels, and (d) cotton pixels. Lines corresponding to errors of 8 K are shown for each plot.

TABLE III
RMSE, SD, AND KLD ON SELECTED DAYS FOR THE SYNTHETIC DATASET

DoY	RMSE (in K)	SD (in K)	KLD
39	1.7	0.3	0.0245
135	3.1	3	0.0963
222	2.4	4.2	0.0963
156	4	2.2	0.2570
354	2	3.2	0.1458

in the downscaled T_B, as shown in Fig. 8(g) and (h). A daily time series of spatially averaged RMSE and SD in the region between the downscaled T_B and the true T_B in the simulation period is shown in Fig. 9. The RMSEs are, on average, 1.5 K higher for the vegetated period than during the bare-soil season. Fig. 10 shows the cumulative density function (CDF) of the errors in downscaled T_B. About 97% of the days have an RMSE of ≤ 3 K in the downscaled T_B. The SD of the error is higher during heterogeneous LC and/or PPT conditions than when they are homogeneous. The highest error SD of almost 10 K is on DoY 262, a day when the study region had both corn and cotton crops, and received heterogeneous PPT. The sensitivity of retrieved SM from T_B is 0.02–0.04 K $\cdot (\mathrm{m}^3/\mathrm{m}^3)^{-1}$ [30].

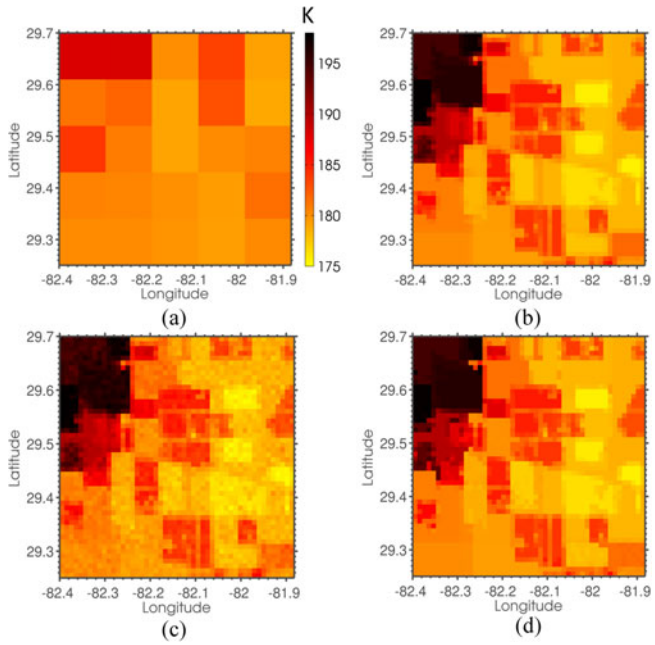

Fig. 12. Comparison of (a) coarse T_B, (b) true T_B, (c) downscaled T_B when the clustering step is not used, and (d) downscaled true T_B when the clustering step is used.

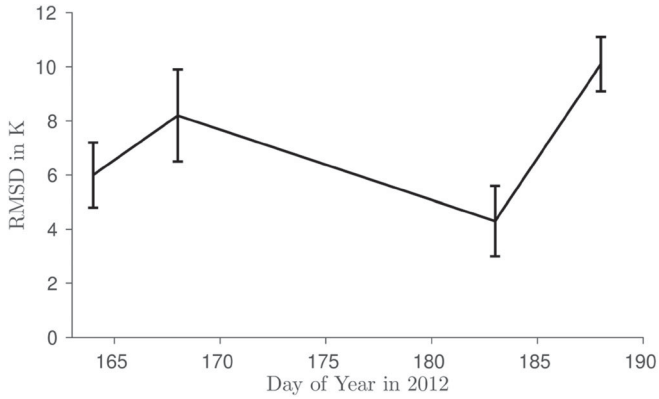

Fig. 13. RMSD between PALS T_B and downscaled T_B during SMAPVEX-12.

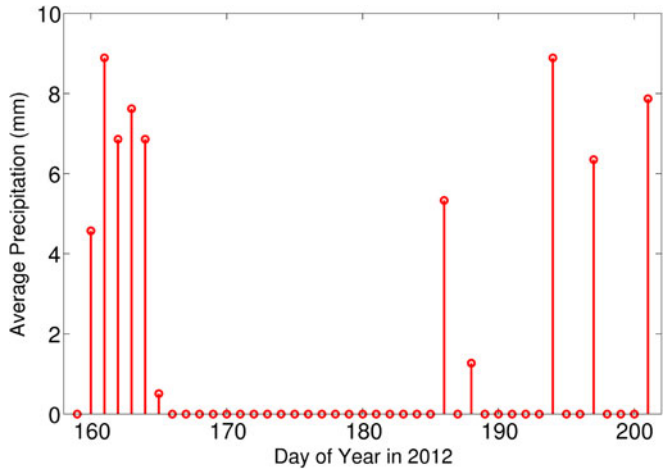

Fig. 14. Precipitation (in mm) received during the SMAPVEX-12 experiment.

Thus, the highest mean errors, on Day 262, correspond to ≤ 0.01 m^3/m^3 error in retrieved SM, making the downscaled product suitable for hydrological applications in agricultural regions.

Fig. 11 shows the downscaled T_B as a function of true T_B at 1 km for different LCs. The points are scattered around the $\hat{y} - y = 0$ line, with a positive variance, indicating that the SRRM did not introduce any bias. Interestingly, the bare-soil pixels during periods of vegetation have the highest RMSE, of about 7.8 K, resulting in a 0.035-m^3/m^3 error in retrieved SM. This is due to the effect of subpixel vegetation at 200 m in a pixel classified as a bare-soil pixel at 1 km, when the vegetation fraction is <0.5 at 1 km. If the pixels with fractional vegetation are omitted, the RMSE reduces drastically to 1.1 K. The RMSE of the errors for sweet corn and cotton were 3.35 K and 2.7 K, with SD of 1.65 K and 1.2 K, respectively. The RMSE, SD, and KLD for the five selected days are shown in Table III. Out of the five selected days, RMSE and SD are higher for DoY 135 and 156 when the region is vegetated with corn and cotton, respectively. In addition to DoY222, the RMSE is higher for DoY 354 than DoY 39 even when bare soil was the only LC during both days, due to the remnants of crops and slightly more heterogeneous precipitation in the region on DoY 354. The KLD for all the days is close to zero, showing that the PDF of the downscaled T_B is close to the PDF of the true T_B at 1 km. Fig. 12 shows a comparison of the downscaled T_B estimates at 1 km with and without the clustering step of the SRRM algorithm. If clustering is not performed, the SRRM algorithm reduces to a single-model nonlinear regression. While the RMSE for this single-model regression method is only 1.5 K higher than for the SRRM-based algorithm, the SD of the error is 8 K higher for the single-model regression method, which reduces its effectiveness in heterogeneous LCs and meteorological conditions and suggests that a single regression model cannot effectively downscale T_B over the entire region.

B. Implementation in the SMAPVEX-12 Region

The RMSD, SD, and KLD for downscaled T_B at 1 km compared to the airborne estimates for the four days during which downscaling is performed are shown in Fig. 13. For all days, the RMSDs are about 6.3 K, except on DoY 188. On DoY 188, the precipitation event occurred after the time of the SMOS overpass but before the PALS acquisitions, leading to a discrepancy between the coarse- and fine-scale T_B. The mean errors during the rest of the study differ by 4 K. The errors mainly arise from the uncertainty in precipitation, which was interpolated from weather station observations. The KLD for the SMAPVEX-12 dataset is about 0.47. This shows that the PDF of the downscaled T_B is considerably dissimilar to the PDF of the airborne T_B, but since the KLD is ≤ 1, the differences in nonzero moments between the distributions are low. This is expected for real datasets, since the coarse- and fine-scale T_B are from different instruments, having different sensor characteristics. Specifically, the cause of the errors in downscaled T_BM, in addition to sensor noise, is that the PPT received is high across all days, as shown in Fig. 14, except DoY 183, and the TRMM PPT has high regional bias during the experiment period. DoY 183 did not have any PPT and, therefore, has the lowest downscaling error of 4.3 K. Adjusting the TRMM Microwave Imager rain-rate according to [31] after comparing with ground PPT estimates is a possible solution, but presents additional complications. In this study, we also used PPT obtained from an inverse-distance-based interpolation of the recorded rain-gauge observations from MAFRI weather stations in the region. The errors were found to decrease by an average of 2.5 K across all days.

The LAI, LST, interpolated PPT, T_B at 10 km, airborne T_B, and downscaled T_B at 1 km are shown in Fig. 15(a)–(f) for DoY 183, respectively. The downscaled T_B is lower than the airborne T_B by about 5 K in general and especially the center of the study region, replicating the trend seen in the SMOS T_B at 10 km, as shown in Fig. 15(e) and (f). However, the finer trends of the high-resolution T_B that are absent in the SMOS T_BM are captured in the downscaled T_B. The spatially averaged RMSD is low at 4.3 K, which translates to a 0.023 m^3/m^3, that is meaningful in heterogeneous agricultural regions.

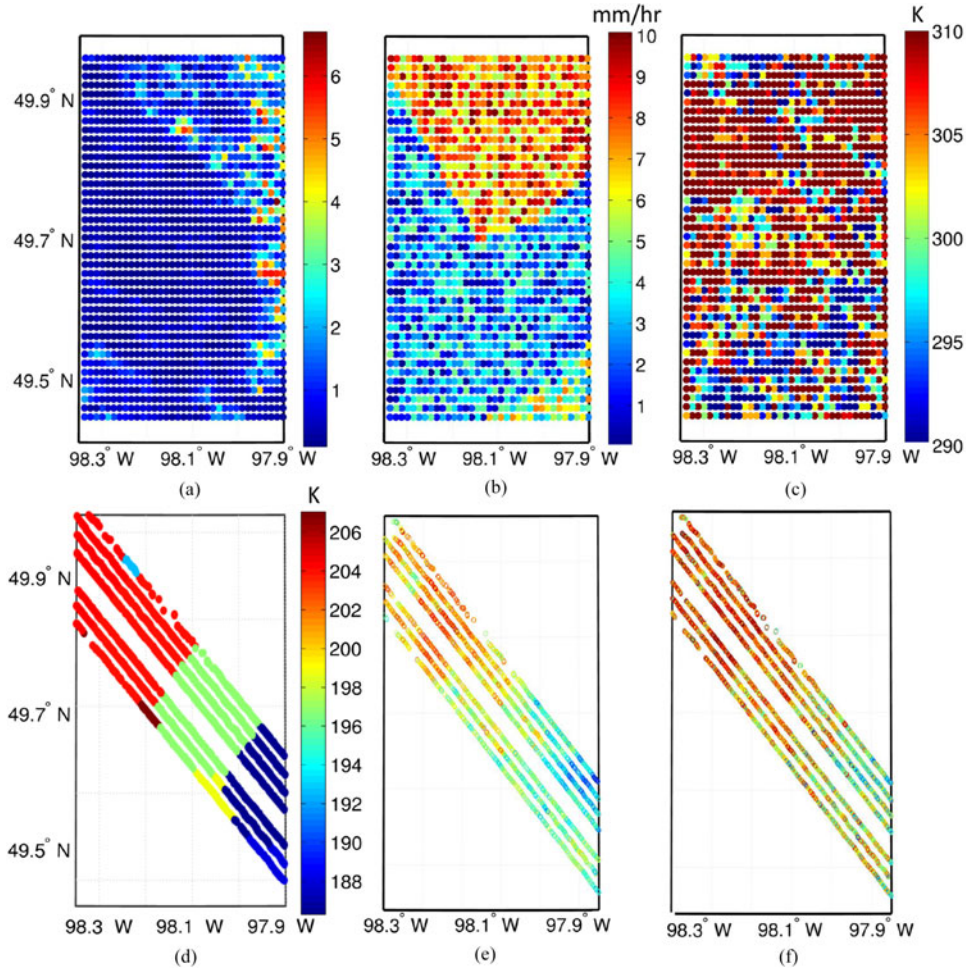

Fig. 15. DoY 183, 2012. (a) LAI. (b) PPT (in mm). (c) LST (in K). (d) SMOS T_B at 10 km. (e) Downscaled T_B at 1 km. (f) Airborne T_B at 1 km.

VI. Summary and Conclusion

In this study, a novel methodology based upon SRRM models was used to downscale satellite-based TB to 1 km for agricultural applications. It was evaluated using a multiscale synthetic dataset and implemented in the SMAPVEX-12 region. 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 high-resolution T_B. For a synthetic dataset in North Central Florida, the algorithm had RMSE of 5.76 K with SD of 2.8 K, during the vegetated season, and an RMSE of 1.2 K with an SD of 0.9 K during nonvegetated periods. The methodology was further applied to downscale SMOS T_B observations, and the KLD between the airborne observations and downscaled T_B was found to be much higher than the KLD between true and downscaled T_B for the synthetic dataset. However, the RMSD between the downscaled T_B and airborne observations was still limited to about 6 K, resulting in an difference of about 0.03 m^3/m^3 in retrieved SM. Thus, this algorithm is particularly useful during heterogeneous LC and micrometeorological conditions.

References

[1] 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. 2015.

[2] 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, Sep. 2014.

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

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

[5] O. Merlin, M. Escorihuela, M. Mayoral, O. Hagolle, A. A. Bitar, and Y. Kerr, "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, 2013.

[6] 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.

[7] J. Kim and T. Hogue, "Improving spatial soil moisture representation through integration of AMSR-E and MODIS products," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 446–460, Feb. 2012.

[8] N. Djamai, R. Magagi, K. Goita, O. Merlin, Y. Kerr, and A. Walker, "Disaggregation of SMOS soil moisture over the Canadian Prairies," *Remote Sens. Environ.*, vol. 170, pp. 255–268, 2015.

[9] F. Lenti, F. Nunziata, C. Estatico, and M. Migliaccio, "On the spatial resolution enhancement of microwave radiometer data in Banach spaces," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1834–1842, Mar. 2014.
[10] X. Wu, J. Walker, C. Rudiger, and R. Panciera, "Effect of land-cover type on the SMAP active/passive soil moisture downscaling algorithm performance," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 846–850, Apr. 2015.
[11] K. Nagarajan, J. Judge, A. Monsivais-Huertero, and W. Graham, "Impact of assimilating passive microwave observations on root-zone soil moisture under dynamic vegetation conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4279–4291, Nov. 2012.
[12] M. Piles, M. Vall-llossera, L. Laguna, and A. Camps, "A downscaling approach to combine SMOS multi-angular and full-polarimetric observations with MDIS VIS/IR data into high resolution soil moisture maps," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, vol. 1, pp. 1247–1250.
[13] N. Das, D. Entekhabi, and E. Njoku, "Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2018–2028, Apr. 2014.
[14] Y. Wang, J. Shi, L. Jiang, J. Du, and B. Tian, "The development of an algorithm to enhance and match the resolution of satellite measurements from AMSR-E," *Sci. China: Earth Sci.*, vol. 54, no. 3, pp. 410–419, 2011.
[15] A. Gambardella and M. Migliaccio, "On the superresolution of microwave scanning radiometer measurements," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 796–800, Oct. 2008.
[16] 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.
[17] R. Jennsen, D. Erdogmus, K. Hild, J. Principe, and T. Eltoft, "Optimizing the Cauchy-Schwarz PDF distance for information theoretic, nonparametric clustering," in *Proc. 5th Int. Conf. Energy Minimization Methods Comput. Vision Pattern Recognit.*, , 2005, vol. 1, pp. 34–45.
[18] S. Chakrabarti, J. Judge, A. Rangarajan, and S. Ranka, "Disaggregation of remotely sensed soil moisture in heterogeneous landscapes using holistic structure based models," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4629–4641, Aug. 2016.
[19] B. Kibria, "Performance of some new ridge regression estimators," *Commun. Statist—Simul. Comput.*, vol. 32, no. 1, pp. 419–435, 2003.
[20] 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
[21] J. Casanova *et al.*, "Field observations during the fifth microwave, water, and energy balance experiment (MicroWEX-5): From March 9 through May, 2006. Circular no. 1514," Center Remote Sens., Univ. Florida, Gainesville, FL, USA, Tech. Rep., 2006. [Online]. Available: http://edis.ifas.ufl.edu/AE407
[22] T. Lin *et al.*, "Field observations during the third microwave, water, and energy balance experiment (MicroWEX-3): From June 16 through December 21, 2004. Circular no. 1481," Center Remote Sens., Univ. Florida, Gainesville, FL, USA, Tech. Rep., 2004. [Online]. Available: http://edis.ifas.ufl.edu/ae361
[23] Florida Automated Weather Network, 2011. [Online]. Available: http://fawn.ifas.ufl.edu/
[24] W. H. Brutsaert, "On a derivable formula for long-wave radiation from clear skies," *Water Resources Res.*, vol. 11, no. 5, pp. 742–744, 2010.
[25] P. Liu, R. DeRoo, A. England, and J. Judge, "Impact of moisture distribution within the sensing depth on L- and C-band emission in sandy soils," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 6, no. 2, pp. 887–899, Apr. 2013.
[26] 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 Resources*, vol. 26, no. 2, pp. 137–149, 2003.
[27] 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, 2008.
[28] J. Privette *et al.*, "Early spatial and temporal validation of MODIS LAI product in the Southern African Kalahari," *Remote Sens. Environ.*, vol. 83, pp. 232–243, 2002.
[29] 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, 2002.
[30] Y. Kerr, P. Waldteufel, J. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, Aug. 2001.
[31] W. Berg, T. L'Ecuyer, and C. Kummerow, "Rainfall climate regimes: The relationship of regional TRMM rainfall biases to the environment," *J. Appl. Meteorol. Climatol.*, vol. 45, pp. 434–454, 2006.

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, and is currently working toward the Ph.D. degree in the Department of Electrical Engineering, University of Florida, Gainesville, FL, USA, since 2012.

He is a Research Assistant jointly at the Center of Remote Sensing and the Computational Neuro-Engineering Laboratory, University of Florida. His research focuses on 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 in the Agricultural and Biological Engineering Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA. Her research interests include microwave remote sensing applications to terrestrial hydrology for dynamic vegetation; modeling of energy and moisture interactions at the land surface and in the vadose zone; spatial and temporal scaling of remotely sensed observations in 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 in the IEEE-GRSS. She also serves the American Geophysical Union as the past Chair of the Remote Sensing Technical Committee in the Hydrology Section.

Anand Rangarajan received the Ph.D. degree from the University of Southern California, Los Angeles, CA, USA.

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

Sanjay Ranka received the B.Tech. degree in computer science from IIT, Kanpur, India, and the Ph.D. in computer science from the University of Minnesota, Minneapolis, MN, USA.

He is currently a Professor in the Department of Computer Information Science and Engineering, University of Florida, Gainesville, FL, USA. His current research interests are focused on a variety of issues related to data science: energy-efficient computing, high-performance computing, data mining, and informatics. Most recently, he was the Chief Technology Officer at Paramark, where he developed real-time optimization software for optimizing marketing campaigns.