Spatial Scaling Using Temporal Correlations and Ensemble Learning to Obtain High-Resolution Soil Moisture

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

***Abstract*— A novel algorithm is developed to downscale soil moisture (SM), obtained at satellite scales of 10–40 km to 1 km by utilizing its temporal correlations to historical auxiliary data at finer scales. Including such correlations drastically minimizes the size of the training set needed, accounts for time-lagged relationships, and enables downscaling even in the presence of short gaps in the auxiliary data. The algorithm is based upon bagged regression trees (BRT) and uses correlations between high-resolution remote sensing products and SM observations. The algorithm trains multiple RTs and automatically chooses the trees that generate the best downscaled estimates. The algorithm was evaluated using a multiscale synthetic data set in north central Florida for two years, including two growing seasons of corn and one growing season of cotton per year. The time-averaged error across the region was found to be 0.01 m^3/m^3, with a standard deviation of 0.012 m^3/m^3 when 0.02% of the data were used for training in addition to temporal correlations from the past seven days, and all available data from the past year. The maximum spatially averaged errors obtained using this algorithm in downscaled SM were 0.005 m^3/m^3, for pixels with cotton land cover. When land surface temperature (LST) on the day of downscaling was not included in the algorithm to simulate "data gaps," the spatially averaged error increased minimally by 0.015 m^3/m^3 when LST is unavailable on the day of downscaling. The results indicate that the BRT-based algorithm provides high accuracy for downscaling SM using complex nonlinear spatiotemporal correlations, under heterogeneous micrometeorological conditions.**

***Index Terms*— Machine learning, remote sensing, soil moisture.**

I. Introduction

SPATIO-TEMPORAL distribution of soil moisture (SM) is highly variable and significantly influences atmospheric and hydrological processes. Accurate SM information at spatial scales of <1 km is critical for applications such as agricultural drought monitoring [1], [2], risk management [3], and productivity predictions [4], with major implications for food security and sustainability. Microwave observations such as those from the European Space Agency's SM and Ocean Salinity mission [5] and the NASA-SM Active Passive (SMAP) mission [6], provide global observations of SM at spatial resolutions of 10–40 km every 2–3 days. These satellite-based SM obtained at coarse spatial resolutions need to be downscaled to about 1 km for applications in heterogeneous agricultural regions.

Most current downscaling algorithms [7] are based upon linear unmixing algorithms, potentially leading to significant loss of structural information in the data [8], especially under highly nonlinear heterogeneous and dynamic conditions such as those in agricultural lands. Hierarchical models [9]–[14] or empirical models based upon statistical [15]–[26] relationships among SM and other remotely sensed (RS) spatially cross-correlated data have been used for downscaling, assuming linear or quadratic relationships. The use of Bayesian descriptors of probability density functions allows downscaling even in the presence of complex nonlinear relationships between the coarse pixel and the multiple fine pixels. Recently, algorithms using higher order relationships have been developed to downscale SM using auxiliary high-resolution RS products—for example, methods based upon principle of relevant information (PRI) [2], [8], support vector regression [27], and copulas [28]. However, these methods require extensive training for obtaining low downscaling errors.

Current downscaling algorithms assume that all the augmented observations are available at the same time coarse SM is available and cannot be applied in the presence of temporal data gaps during cloud-cover, sensor recalibration exercises, solar flares, etc. In addition, the relationship of some high-resolution observations, such as precipitation (PPT), to SM is discernible only over extended periods of time. Historical SM and its temporal correlations to other variables can be utilized to downscale SM in the presence of data gaps, as well as to minimize errors in downscaled SM due to time-lagged relationships. These temporal correlations can also be leveraged for downscaling SM in regions where accurate characterization of the relationships between fine-resolution SM and auxiliary fine scale data inside a coarse pixel requires extensive training data of *in situ* SM. While implementation of downscaling algorithms using comprehensive training sets is realistic in data-rich agricultural regions in developed countries, the number of *in situ* stations sharply decreases in developing and underdeveloped countries. Augmenting the training

Manuscript received January 5, 2017; revised May 18, 2017; accepted June 7, 2017. Date of publication January 4, 2018; date of current version February 27, 2018. This work was supported by the NASA-Terrestrial Hydrology Program under Grant NNX13AD04G. *(Corresponding author: Subit Chakrabarti.)*

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

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

Digital Object Identifier 10.1109/TGRS.2017.2722236

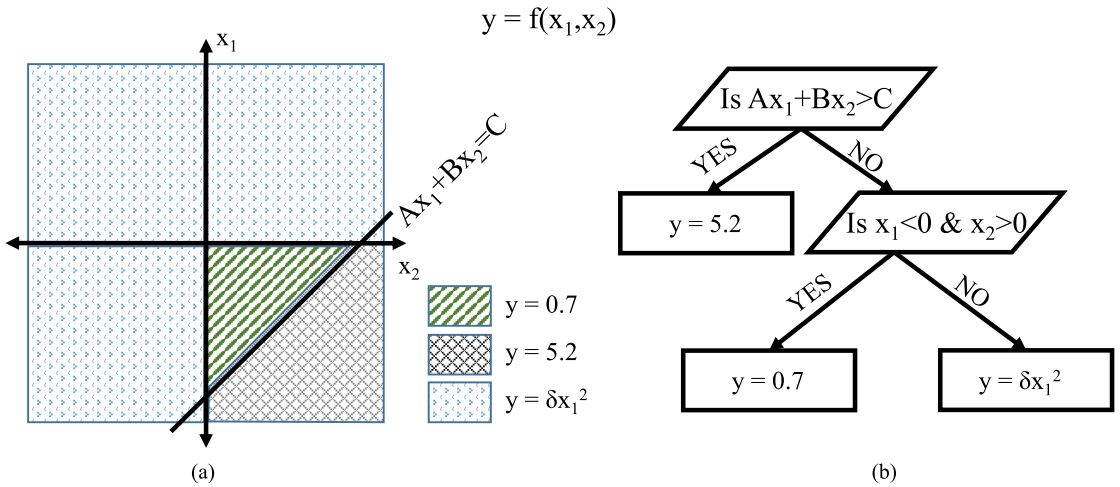

Fig. 1. (a) 2-D vector space with three generative models, shown by shaded spaces, corresponding to three regression models. (b) Associated RT that takes x_1 and x_2 as input and assigns the correct regression model corresponding to the vector space divisions. The parallelograms enclose partitioning rules and the rectangles enclose regression functions assigned to each partition.

set using historical observed data can increase the efficacy of downscaling algorithms. Significant gaps exist in utilizing historical RS and *in situ* data to implement downscaling algorithms involving implicit or explicit temporal prediction components.

Adding historical data to algorithms such as the PRI is impractical due to computational burden arising from nonstationarity, quick-switching physical regimes, and fitting large number of hyperparameters. Nonlinear regression techniques can be used as they have higher order prediction capabilities and can be parallelized for computational efficiency. However, these techniques tend to overfit [29] and need to be regularized to provide reliable results. Implicit regression methods are more suited to this problem as they are more efficient than explicit regression methods, such as PRI or nonlinear regression, and they can be easily regularized. In this paper, the most widely used implicit regression method, tree-based regression [30], is used to downscale SM using historical information of correlated auxiliary features. It is regularized using ensemble learning techniques [31] that combine results of multiple models to reduce overfitting.

The goal of this paper is to understand the impact of incorporating historical information of SM in improving downscaled SM estimates in the presence of temporal data gaps or in data sparse regions. The primary objectives of this paper are to: 1) develop an algorithm to downscale coarse scale SM using historical information of auxiliary variables and SM; 2) implement this algorithm to downscale SM to 1 km using land surface temperature (LST), land cover (LC), leaf area index (LAI), and PPT; and 3) quantify the algorithm performance for application in data-sparse regions and in the presence of temporal data gaps. In this paper, a synthetic data set [32] will be used for the above objectives. Operationally, satellite products such as LST, LAI, and LC from the moderate resolution imaging spectroradiometer (MODIS) instrument onboard National Aeronautics and Space Administration Aqua/Terra, and PPT from the National Aeronautics and Space Administration/Japan Aerospace Exploration Agency Global Precipitation Measurement mission could be used.

Section II describes the theoretical details of the downscaling framework. Section III discusses the details of the data set used in this paper. Section IV illustrates the steps for the implementation of the algorithm, presents the downscaling results for SM at 1 km, and discusses the impact of both temporal training and temporal data gaps on the downscaled SM. Section V summarizes the results and concludes this paper.

II. Theory

Downscaling, in general, is an ill-posed problem that results in a multiplicity of solutions after regression, requiring regularization. In this paper, the downscaled SM is obtained using regression trees (RTs) [30]. To regularize the fine-scale SM estimates using RTs, a powerful nonlinear and nonparametric ensemble learning technique called bootstrap aggregation (bagging) [31] is used.

A. Regression Trees

Linear or kernel regressions are global methods, with a single predictive model used over the full range of the input data. However, heterogeneous and dynamic input variables, such has PPT and LC, interact with each other and exhibit highly nonlinear correlations with SM making a single model suboptimal. RTs [30] subdivide or partition the space of the inputs, such as auxiliary variables and coarse scale SM, recursively and map each partition to the required output, such as downscaled SM. Fig. 1(a) shows the partitioned domain for a simple function $f(x_1, x_2)$, $\{x_1, x_2\} \in \mathbb{R}^2$. The function being modeled is defined piecewise using constants A, B, and δ as

$$f(x_1, x_2) = \begin{cases} 5.2, & \text{if } Ax_1 + Bx_2 > C \text{ for } \{x_1, x_2\} \in \mathbb{R}^2 \\ 0.7, & \text{if } Ax_1 + Bx_2 < C,\ x_1 < 0,\ x_2 > 0 \\ & \text{for } \{x_1, x_2\} \in \mathbb{R}^2 \\ \delta x_1^2, & \text{otherwise.} \end{cases} \tag{1}$$

The model is nonlinear and cannot be represented by any continuous parametric model. But a simple RT model can be

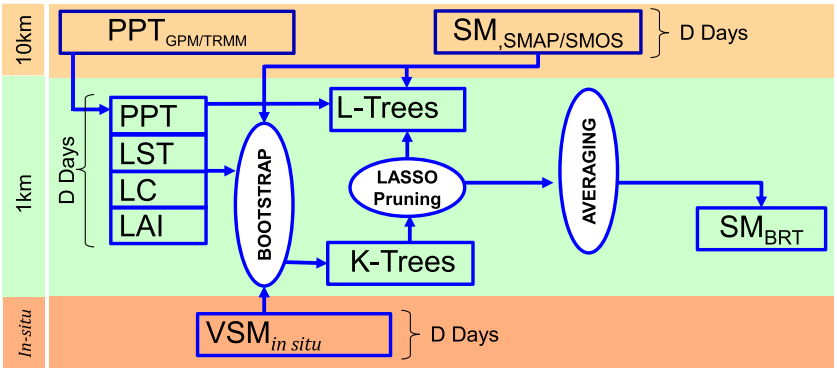

Fig. 2. Flowchart of showing the boosted regression trees (BRT) downscaling method. In the algorithm, PPT, LST, LC, LAI, and SM are used for the prior D days.

created for this function, as shown in Fig. 1(b). Because both the partitions within the input data space and the regression models can be nonlinear, the RT method can approximately learn any mapping function between the input and output data sets. If the number of partitions is too few, the regression models would be complex and fail to capture the structure of the data, while if the number of partitions is too many, the models would overfit and perform poorly in the presence of noise in the input data.

An algorithm to construct an RT consists of a heuristic, used to recursively select which input variable to use to partition the input space for each level of the tree, henceforth called the partitioned variable, and the stopping criterion that specifies when each partition reaches the optimal size. In this paper, the partitioned variable for each level of the tree is decided based upon the sum of squared errors on the validation set after regression. For each tree t, the prediction error $E(t)$ is defined as

$$\begin{aligned} E(t) &= \sum_{\forall l \in \text{Leaf}\{t\}} \sum_{\forall i \in l} \left(\frac{1}{n_l} \sum_{\forall i \in l} y_i - y_i \right) \\ &= \sum_{\forall l \in \text{Leaf}\{t\}} n_l \left(\sum_{\forall i \in l} \left(\frac{1}{n_l^2} \sum_{\forall i \in l} y_i - y_i \right) \right) \\ &= \sum_{\forall l \in \text{Leaf}\{t\}} n_l \cdot \text{Var}(l) \end{aligned} \tag{2}$$

where $y = \hat{f}(x)$ is the predicted variable and $\text{Var}(l)$ is the variance in the set of possible predictions for leaf l. The optimal partition results in minimization of the prediction error E on the validation set. In this paper, the stopping criterion is achieved when additional splits decrease $E(t)$ by <0.01 m^3/m^3 SM.

B. Bootstrap Aggregation and Ensemble Pruning

A single RT, as described above, is prone to overfitting the data. Bootstrap aggregation or bagging [31] is an ensemble learning technique that resamples the original training set uniformly to create multiple new data sets of the same size as the original data set. Each bootstrapped sample is fit using a single RT, resulting in the same number of RTs as the number of data sets and the mean of the outputs as the final output. This procedure has been shown to improve the stability of individual trees that might overfit the data due to their nonparametric nature [31].

For computational efficiency, ensemble pruning [33] is employed. This technique uses the ensemble of models generated by bagging and reduces the number of models by weeding models that perform poorly on a validation set, while maintaining the diversity of models. The L^1-constrained fitting for statistics and data mining least absolute shrinkage and selection operator (LASSO) [34] algorithm has been widely used for this purpose. For K RTs, the mean square loss [35] E_{overall} for a validation set $\{x_1 \ldots x_N\}$ is

$$E_{\text{overall}} = \sum_{n=1}^{N} w_n \left(f(x_n) - \sum_{k=1}^{K} \alpha_k \hat{f}_k(x_n) \right)^2 + \beta \sum_{k=1}^{K} \alpha_k \tag{3}$$

where $f(x_n)$ is the true value of the function, $\hat{f}_k(x_n)$ is the estimated value of the function by the kth RT, β is the regularization parameter, w_n are the weights to be learned, and α_k is the learner coefficients. Learning the weights is then equivalent to solving the following:

$$\min_{\alpha} \sum_{n=1}^{N} w_n \left(f(x_n) - \sum_{k=1}^{K} \alpha_k \hat{f}_k(x_n) \right)^2 \quad \text{s.t.} \sum_{k=1}^{K} \alpha_k \leq \frac{1}{\lambda} \tag{4}$$

where λ is a constant that regularizes the weights α_k. The higher λ is, the more trees are selected after ensemble pruning. The solution provides the optimum weights for each weak learner. The l learners with the highest weights are then selected. This technique naturally avoids overfitting and allows for extremely powerful generalization.

C. Algorithm Summary

The overall method of bagged regression trees (BRT) is shown in Fig. 2. For incorporating historical information into the algorithm, a spatiotemporal training data set is created by augmenting current *in situ* SM and auxillary data with data from the prior days. Our hypothesis is that using such an algorithm will significantly reduce the *in situ* data required

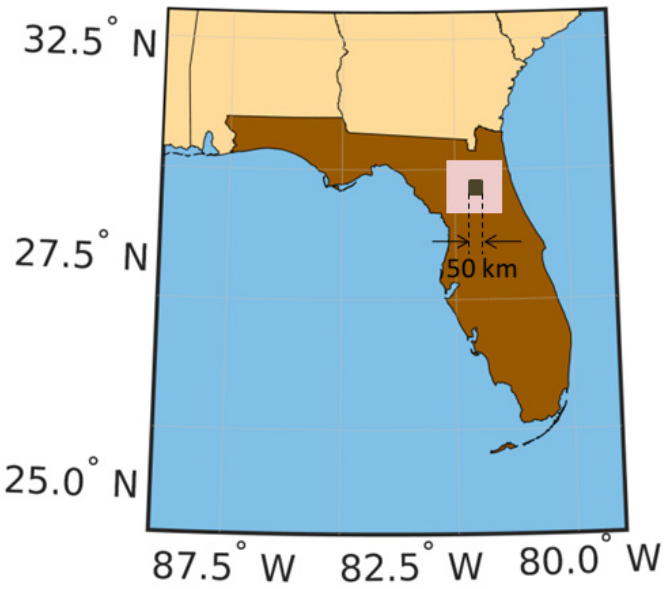

Fig. 3. Study region in North Central Florida. LSP-DSSAT-MB simulations were performed over the shaded 50×50 km^2 region.

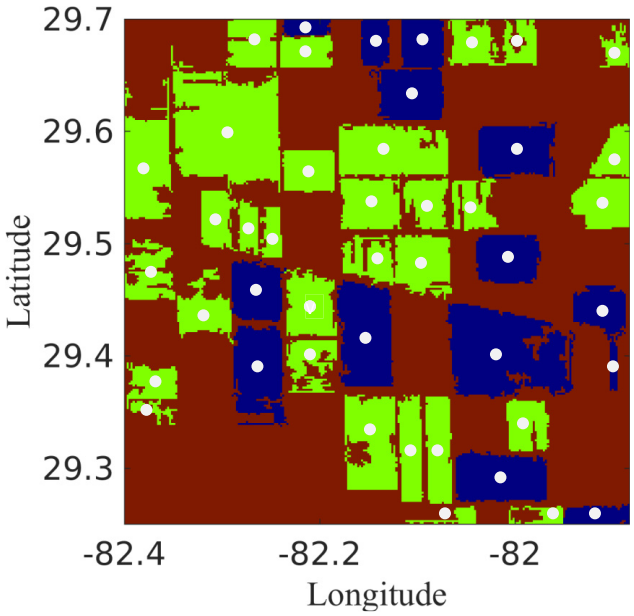

Fig. 4. LC at 200 m during cotton and corn seasons. Blue regions represent sweet-corn, green region represents cotton, and brown regions represent bare-soil LC. Homogeneous crop fields along with centers are shown for sweet-corn and cotton.

for training and can be implemented in data poor regions. For each LC, separate replicates of the spatiotemporal data set are created using the bootstrapping method. An equal number of RTs as the replicates are grown for each bootstrapped [31] version of the data, and LASSO-pruning is used to separate out trees that best describe the data, as described in Section II. Finally, the coarse SM is bootstrapped and regressed using the selected trees and the averaged output of these is the downscaled SM.

III. Experimental Data Set

The downscaling algorithm in this paper used a data set from the simulation framework consisting of a soil-vegetation-atmosphere transfer model, the land surface process (LSP) model, coupled with a crop growth model, the decision support system for agrotechnology transfer (DSSAT) model, detailed in [32]. A 50×50 km^2 region, equivalent to approximately 25 SMAP pixels with a spatial resolution of 9 km/pixel, was chosen in North Central Florida (see Fig. 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, have been conducted for various agricultural LCs over the last decade [36]–[38]. Simulated observations of SM, LST, and LAI were generated at 200 m for two years, from January 1, 2007 to December 31, 2008. Topographic features, such as slope, constant as the region is typically characterized by flat and smooth terrains with no runoff 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 located within the study region. The observations were spatially interpolated using splines to generate the meteorological forcings at 200 m. Long-wave radiation was estimated following [39]. The model simulations were performed over each contiguous homogeneous region of sweet-corn, bare soil, and cotton, as shown in Fig. 4, rather than all the pixels, to reduce computation time. A realization of the LSP-DSSAT model was used to simulate SM, 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. 4.

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

The model simulations at 200 m were spatially averaged to obtain PPT, LST, LAI, and SM at 1 and 10 km. Linear averaging is typically sufficient to illustrate the effects of resolution degradation [40]. To simulate rain-fed systems, all the water inputs from both PPT and irrigation were combined together, and the "PPT" in this paper represents these combined values.

The simulation period in both the years 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 SM 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 [9], [41], [42].

The errors added had zero mean and standard deviations of 5 K, 1 mm/h, 0.1 and 0.02 m^3/m^3 for LST, PPT, LAI at 1 km, and SM at 10 km, respectively, similar to [43] and [44] algorithms.

IV. Methodology and Results

The inputs to the BRT method are LST, PPT, LAI, and LC at 1 km. The downscaled SM estimates in this paper are instantaneous and indicate the SM values at the time of acquisition of the coarse scale SM. The first part of this paper, described in Section IV-A, used only spatial correlations to downscale SM, similar to the current downscaling studies [25], [43]. Two scenarios are implemented and compared. The first in which randomly selected 33% of the data set, or 750 out of 2500 pixels, are used for training (BRT_{750}), similar to other training-set-based downscaling algorithms [43]–[45]. And second, in which 0.02% of the data set, or 30 pixels, are used for training (BRT_{30}). The second scenario represents realistic situations for data poor regions such as developing countries and the minimum training data needed to meet the mission requirement of 0.04 m^3/m^3 error in SM. An equal number of points are chosen from each LC for training. Within each LC points are chosen via random sampling. However, it is not expected that every geographic region will need exactly 30 points to ensure a maximum root mean square error (RMSE) of 0.04 m^3/m^3 in downscaled SM. In general, more training data set samples are needed to downscale with increasing heterogeneity in the region in terms of the region's LC, soil texture, micrometeorological conditions, and slope/elevation in the coarse pixel. To accurately characterize the performance of the algorithms over changing LC and micrometeorology, both the BRT_{750} and BRT_{30} algorithms are implemented using the SM and auxiliary variables from January 1 to December 31, 2008.

In the second component of this paper, described in Section IV-B, the BRT-based downscaling is modified and extended to utilize temporal correlations, in addition to the spatial correlations (BRT_{st}). This method is nonparametric and thus, can be easily extended to include temporal training The BRT_{st} algorithm is trained using 0.02% of the data set in space, and all the prior data available for those training pixels, to provide temporal correlations. In this paper, the prior data consist of all the data available from one year prior to the day of downscaling in 2008.

In addition to handling spatial data sparsity, the BRT_{st} algorithm can be used to downscale SM in the presence of data gaps. In the third component of this paper, the effects of data gaps in LST are investigated. Remotely sensed LST, which is usually retrieved from the MODIS instrument onboard NASA Aqua and Terra satellites, is the most likely to be unavailable due to presence of clouds [46]. The near real-time LST products, which we envision will be used for this downscaling algorithm, are additionally affected by gaps caused due to solar gaps and imperfect satellite transmissions. In this paper, the effect of data gaps in LST is simulated by withholding LST data for the entire region on the day of downscaling, and on prior days, from the input data set.

A. BRT-Based Downscaling With Spatial Correlations

For downscaling based on only spatial correlations, the RTs are trained using the auxiliary data set on each day that the coarse SM is available. In (2)–(4), from Section II

$$\begin{aligned} f^t(x_n^t) &= \mathrm{SM}_{n,\mathrm{insitu}}^t \\ x_n^t &= \left(\mathrm{LST}_n^{1\ \mathrm{km}}, \mathrm{PPT}_n^{1\ \mathrm{km}}, \mathrm{LAI}_n^{1\ \mathrm{km}}, \mathrm{LC}_n^{1\ \mathrm{km}}, \mathrm{SM}_n^{10\ \mathrm{km}}, \mathrm{X}_n, \mathrm{Y}_n\right)^t \\ \hat{f}^t(x_n^t) &= \mathrm{SM}_{n,\mathrm{downscaled}}^t \end{aligned} \tag{5}$$

where X_n and Y_n are the x and y coordinates of the nth pixel and superscript t denotes the time index (in DoY) and $n \in [1, 2500]$. The average RMSEs for the BRT_{750} method for three LCs are shown in Table III. The errors for both vegetated and bare-soil pixels are very low, with bare-soil pixel having a higher mean error due to effects of fractional LC in the boundary pixels and crop remnants after harvest. The PRI [43] method also results in errors close to zero under this conditions, with a similar training set size. The BRT_{30} algorithm produces errors with a mean of 0.067 m^3/m^3 and a standard deviation of 0.051 m^3/m^3, similar to the other algorithms that use only spatial correlations. This is expected since the spatial correlations in such low-volume training sets are not strong enough to yield definitive relationships that can be utilized to downscale SM.

B. BRT-Based Downscaling With Spatio-Temporal Training

For the BRT_{st} algorithm, an RT is trained when coarse-scale SM is available, using historical information of LST, LAI, and PPT from the previous year for 0.02% of the pixels. Thus

$$\begin{aligned} f^t(x_{nt}) &= \mathrm{SM}_{\mathrm{nt,insitu}} \\ x_{nt} &= \Big(\left[\mathrm{LST}_{(t-D_1)n}^{1\ \mathrm{km}}, \mathrm{LST}_{(t-D_1+1)n}^{1\ \mathrm{km}}, \ldots, \mathrm{LST1}_{tn}^{1\ \mathrm{km}}\right], \\ &\quad \left[\mathrm{LAI}_{(t-D_2)n}^{1\ \mathrm{km}}, \mathrm{LAI}_{(t-D_2+1)n}^{1\ \mathrm{km}}, \ldots, \mathrm{LAI}_{tn}^{1\ \mathrm{km}}\right] \\ &\quad \left[\mathrm{PPT}_{(t-D_3)n}^{10\ \mathrm{km}}, \mathrm{PPT}_{(t-D_3+1)n}^{10\ \mathrm{km}}, \ldots, \mathrm{PPT}_{tn}^{1\ \mathrm{km}}\right], \\ &\quad \mathrm{SM}_{tn}^{10\ \mathrm{km}}, \mathrm{X}_{tn}^{10\ \mathrm{km}}, \mathrm{Y}_{tn}^{10\ \mathrm{km}}\right]\Big) \\ \hat{f}^t(x_{nt}) &= \mathrm{SM}_{\mathrm{nt,downscaled}} \end{aligned} \tag{6}$$

where $n \in [1, 2500]$ is the space-index, $t \in [1, 365]$ is the time index, and D_1, D_2, and D_3 are the number of values of LST, LAI, and PPT to include in the feature vector from previous days, respectively. For training purposes, LAI, available every seven days, is assumed to be constant between consecutive observations, while LC, PPT, and LST are available every day. The feature vector for downscaling SM is described in more detail in Fig. 5(a). The 30 pixels, n in Fig. 5, are randomly selected to serve as the training set. For each time-index t, the D_1, D_2, and D_3 observations of LST, LAI, and PPT, respectively, are included as the columns of the feature matrix that enable the algorithm to use temporal correlations and capture spatial SM variability. The range of the time index t in the rows of the feature matrix, ranging from $t-1$ to $t-365$, indicates that a moving window of historical auxiliary and *in situ* data from the past one year is used to augment the training data set. For notational simplicity, let us assume that LST was

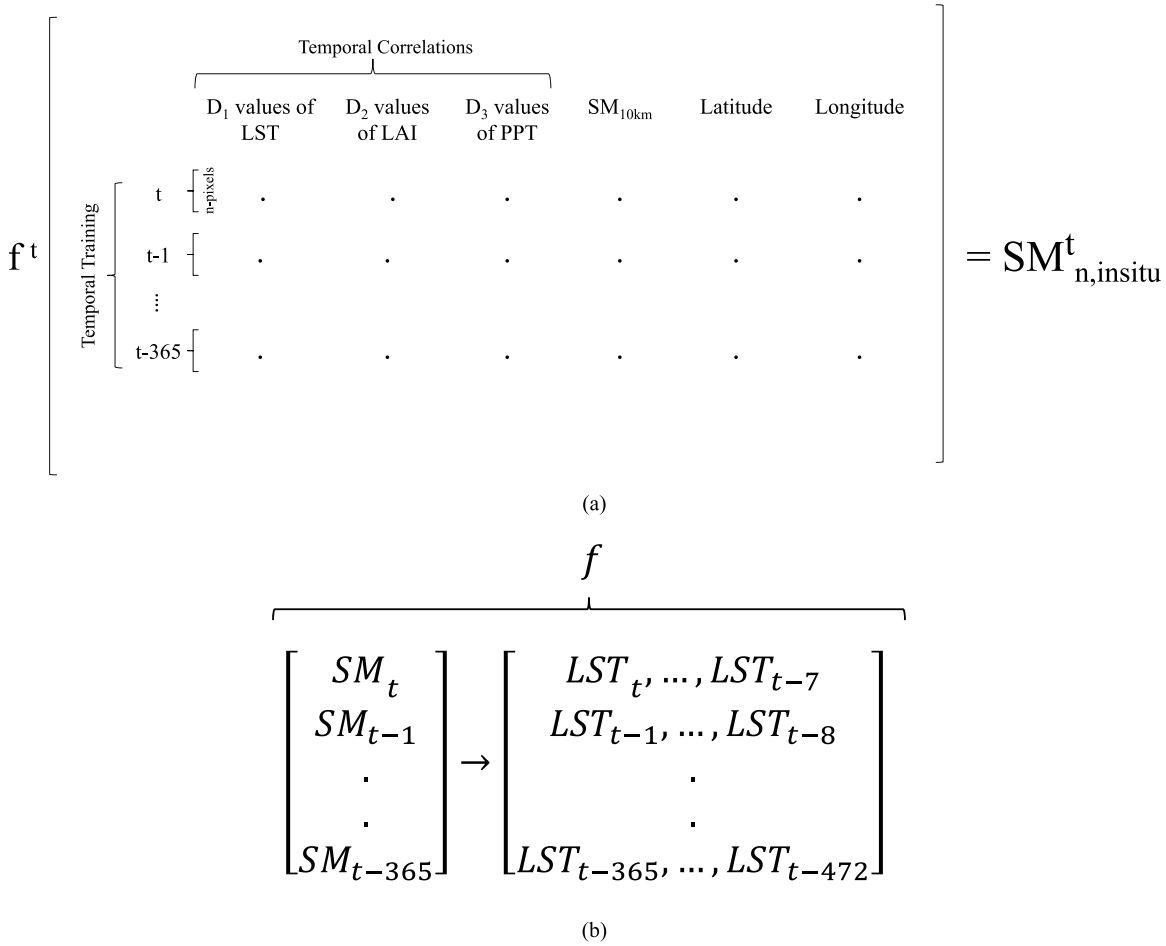

Fig. 5. (a) Complete feature vector for the bagged regression trees algorithm with spatiotemporal training. In the algorithm, PPT, LST, LC, LAI, and SM are used for the prior D_1, D_2, and D_3 days. (b) Downscaling function if SM is just dependant on LST.

the only feature. Then, at time t, SM is estimated using LST from the previous seven days. Then, the SM is

$$\text{SM} \rightarrow f(\text{LST}_t, \ldots, \text{LST}_{t-7}). \tag{7}$$

Then, the function f that relates *in situ* SM to the LST for previous seven days is found using all the available *in situ* SM points from the past (one year) and their corresponding LSTs from previous seven days. This is shown in Fig. 5(b). So, the LSTs (or features) are always only seven days from the *in situ* SM. The 30 locations in the study area that is used for training spatiotemporally are shown in Fig. 6.

The spatially averaged errors in downscaling were compared for different values of D_1, D_2, and D_3, to understand the impact of utilizing temporal correlations. These errors were also averaged for 122 days in the year 2008. These spatiotemporally averaged errors, while underestimating the possible errors due to downscaling in areas with highly varying LC and/or meteorological conditions, provide a reasonable estimate of the upper bound of how much historical data to include, beyond which the variables become almost uncorrelated to the current values of SM. Fig. 5(a)–(c) shows these spatiotemporal errors for different values of D_1, D_2, and D_3. The errors when previous values of LST, PPT, and LAI are used are shown in Fig. 7. For LST and LAI, the spatially averaged error increases significantly after seven days

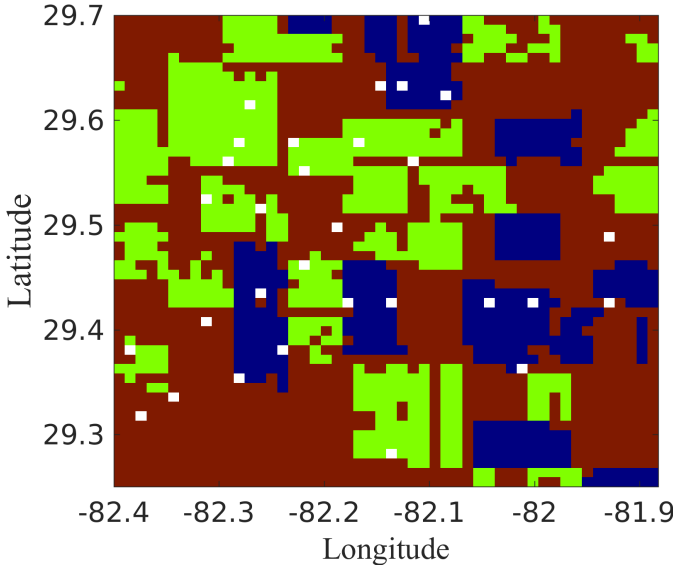

Fig. 6. Thirty locations within the study area used to train the bagged regression trees algorithm with spatiotemporal training are shown in white, along with the LCs. Blue regions represent sweet-corn, green regions represent cotton, and brown regions represent bare-soil LC.

and seven weeks, respectively. This indicates that including additional LST information from earlier than seven days or including additional LAI information from earlier than six

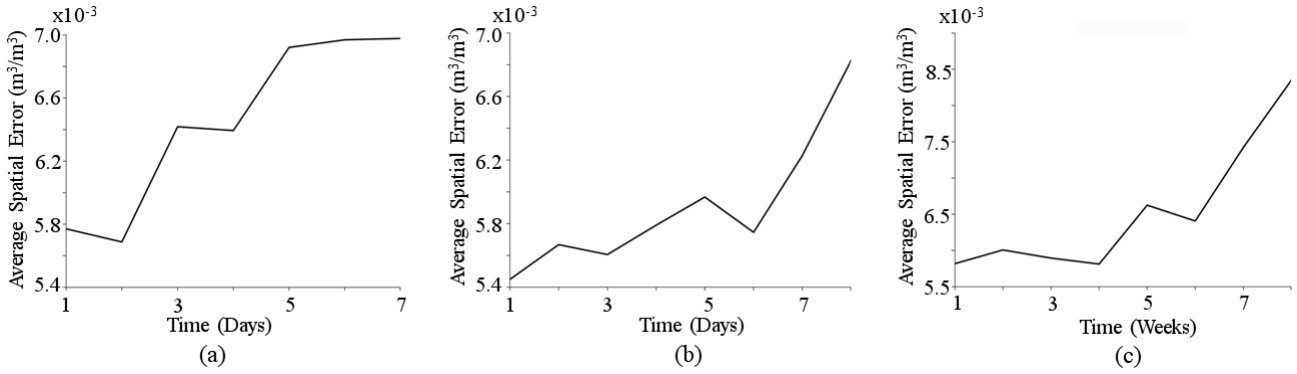

Fig. 7. Spatially averaged error in downscaled SM versus time lag of observations for (a) LST, (b) PPT, and (c) LAI.

weeks does not provide any added value to the downscaled estimates. The spatially averaged errors, for PPT, stabilize after five days. However, the effect of PPT on SM is highly nonlinear and is a function of the magnitude of and time-delay from the last rainfall event. In general, if we do not consider spurious correlations, such as those on the third day in Fig. 5(b), PPT does not have a noticeable effect on SM beyond a week's lag, even for the rain-fall events, as high as 3.6 mm/hr. For the ease of implementation, seven prior values[1] of PPT, LST, and LAI data, an overestimate for D_1, D_2, and D_3 equal to seven, were added to the feature vector [x_{nt} in (5)].

The sensitivity of the BRT_{st} algorithm to the number of trees [K in (2)] is investigated by dividing the training set into ten equal parts randomly and using nine parts for training and one part for evaluating the algorithm. This methodology, known as tenfold cross-validation, is repeated with different randomly selected partitions to approximate the errors that the BRT_{st} algorithm would incur on an average. Fig. 8 shows the tenfold cross-validation error as a function of the number of RTs trained.

Fig. 8 shows the generalization error as a function of the number of trees. Theoretically, the curve for the generalization error would require an infinite number of data points. It is methodologically simpler to conservatively select the number of trees based on an approximate minimum in the generalization error curve. This choice is primarily constrained by the computational time that can be afforded. Fifty trees is a conservative estimate of the number of trees needed for minimum generalization error. For greater than 50 trees, the estimated error in downscaling decreases by only 0.05 m^3/m^3, which is an inconsequential amount compared to the extra computational time, of about 10 min for each day, incurred for each additional tree. Thus, in this paper, 50 trees were used to estimate the downscaling function. To decrease overfitting, after the number of trees in the model is decreased even further after training through regularization, as shown in Fig. 9(a). The lambda value of 2×10^{-2} is used, which reduces the number of active trees needed to 20 for an average increase in error of 0.003 m^3/m^3 as shown in Fig. 9(b). While applying

[1]Note that the number of days/value is different for each feature. For example, LST is available every day so seven values represent seven days, while LAI is available weekly, so seven values represent seven weeks.

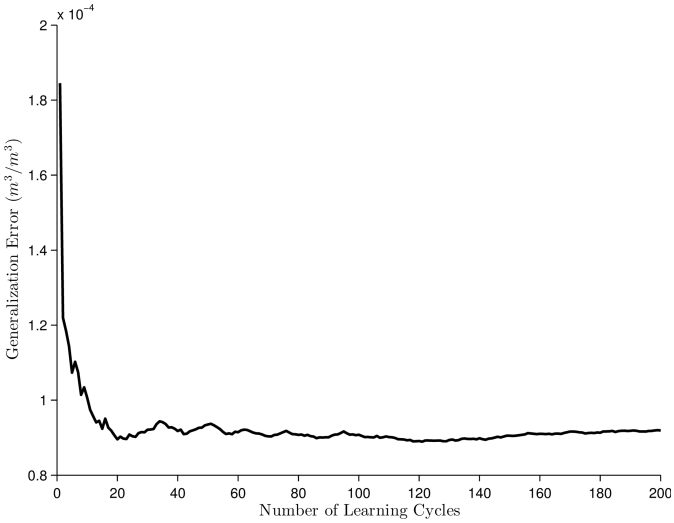

Fig. 8. Generalization error versus number of learning cycles for the bagged regression trees algorithm with spatiotemporal training.

the trained RT ensemble for downscaling, decreasing the number of trees from 50 to 20 reduces the computational time marginally by about 5 min per day. This reduction is expected to be much more significant when downscaling for multiple years for a higher number of pixels.

The spatially averaged errors in SM during the year 2008 are $\leq$ 0.01 m^3/m^3, with the highest errors, of about 0.023 m^3/m^3 during simultaneous corn and cotton LCs as well as late season bare soil due to remnant crops. Table III shows the RMSEs for the BRT_{st} algorithm, for various LCs, as shown in Fig. 10. The RMSE for the BRT_{st} method is higher than for the BRT_{750} algorithm by about 0.018 m^3/m^3. This is expected as the training set used by the BRT_{750} algorithm is 24 times larger in volume. The expected RMSE increase in the SM downscaled using the BRT_{st} algorithm is marginal considering the significant advantages from a reduced training data set.

Five days were selected from the season to understand the effect of the heterogeneity in inputs on the error in disaggregated SM. Variabilities in PPT, ranging from uniformly wet to uniformly dry, and in LC, ranging from bare soil to vegetated with both cotton and sweetcorn, were used as criteria for selecting the days, as shown in Table II and Fig. 11.

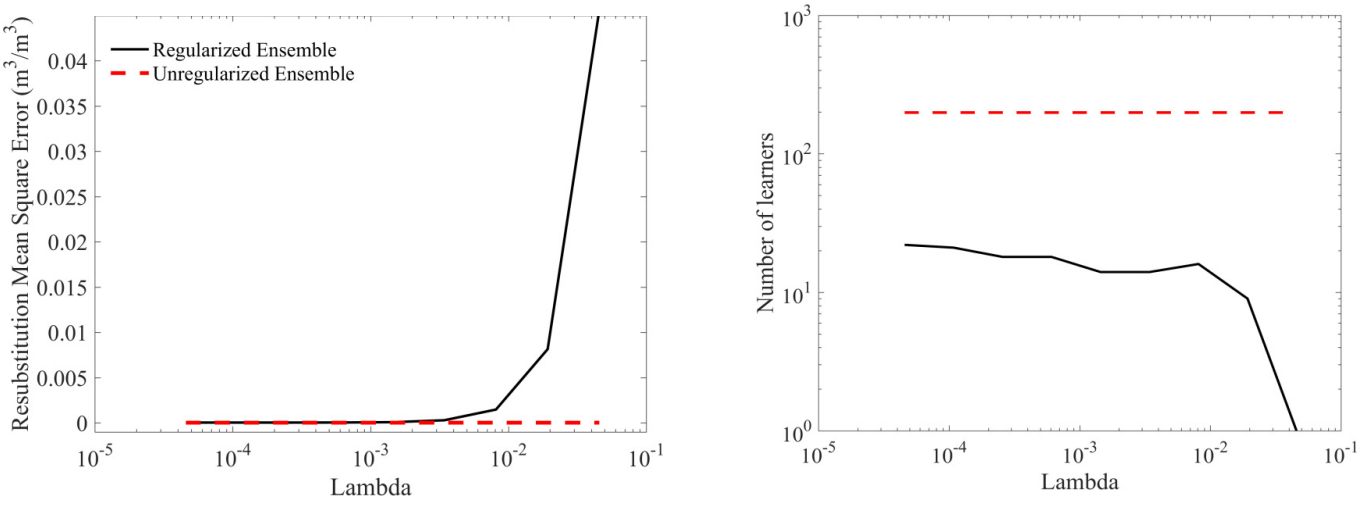

Fig. 9. (a) Resubstitution Error versus Lambda for the bagged regression trees algorithm with spatiotemporal training. (b) Number of learners versus lambda for the bagged regression trees algorithm with spatiotemporal training.

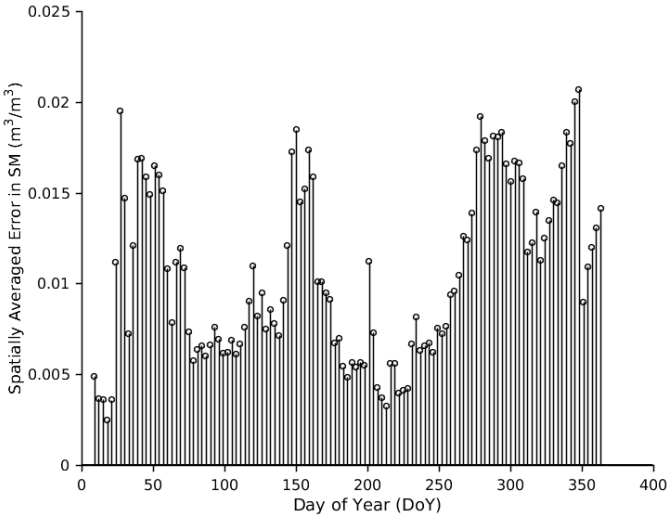

Fig. 10. Spatially averaged RMSE in disaggregated SM at 1 km for each day of the year in the validation year when coarse SM is available.

TABLE II

DAYS SELECTED FOR EVALUATING BRT ESTIMATES. THESE DAYS CAPTURE VARIABILITY IN PPT/IRRIGATION AND LC

DoY	Precipitation & Land-Cover
39	Dry, Baresoil
135	Dry, Sweet Corn
156	Wet, Cotton
222	Wet, Sweet Corn and Cotton
354	Wet, Bare

For DoY 354, the LST, as shown in Fig. 11(f), shows field-structures even though the LAI for the region, shown in Fig. 11(e), is zero because of crop remnants after harvest. For the five selected days as shown in Table II, the true SM and

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 BRT_{750}, BRT_{30}, AND $\mathrm{BRT}_{\mathrm{ST}}$ METHODS. A–BARESOIL PIXELS WITH VEGETATED SUBPIXELS AT 250 m TILL DoY 332, B—BARESOIL PIXELS AFTER DoY 332 C–BARESOIL PIXELS WITHOUT ANY VEGETATED SUBPIXELS AT 250 m TILL DoY 332

Land Cover	$\mathrm{RMSE}(\mathrm{BRT}_{750})$	$\mathrm{RMSE}(\mathrm{BRT}_{30})$	$\mathrm{RMSE}(\mathrm{BRT}_{\mathrm{st}})$
Corn	1.8615×10^{-17}	0.0735	0.0022
Cotton	2.4828×10^{-04}	0.0910	0.0053
Baresoil$^{\mathrm{A}}$	5.6222×10^{-5}	0.0845	0.0543
Baresoil$^{\mathrm{B}}$	5.628×10^{-6}	0.1041	0.0517
Baresoil$^{\mathrm{C}}$	2.5948×10^{-6}	0.0452	0.0012

TABLE IV

PEARSON'S CORRELATION COEFFICIENT BETWEEN RESIDUAL ERROR AND FEATURES ON DAY OF YEAR (DoY) 222

Features	Correlation
Land Surface Temperature	-0.1715
Leaf Area Index	0.0793
Precipitation	0.0010
Land Cover	0.1235

the downscaled SM are shown in Figs. 12–16. Figs. 12 and 13 show the downscaled SM for early and late-season bare soil conditions. For the bare soil pixels, the maximum error is 0.03 m^3/m^3, during the early season, and the maximum errors are in the southwest corner of the field where there are no *in situ* training sites. This is also observed during the late season bare soil conditions in Fig. 13. For the sweet-corn LC, shown in Fig. 14, and the cotton LC, shown in Fig. 15, the error in SM is the highest at the boundaries of the field and the

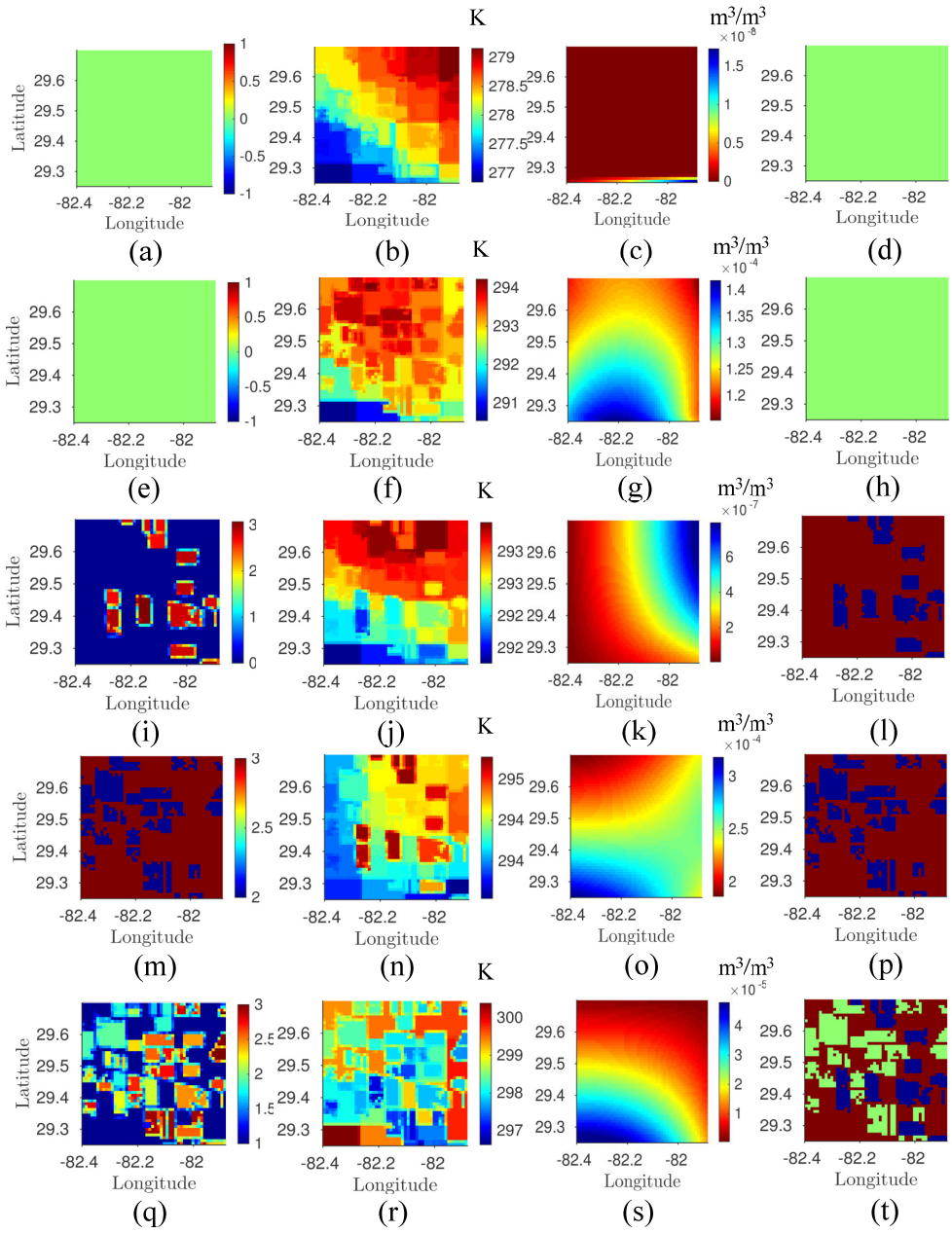

Fig. 11. LAI, LST, PPT, and LC for (a)–(d) DoY 39, (e)–(h) DoY354, (i)–(l) DoY 135, (m)–(p) DoY 156, and (q)–(t) DoY 222, respectively.

south west corner of the field. On DoY 222, shown in Fig. 16, even when there was a maximum heterogeneity in LC with corn, cotton, and bare soil, the errors are minimal, although the standard deviation of the error is high at 0.027 m^3/m^3. The time-averaged error across the region was 0.027 m^3/m^3, with a standard deviation of 0.012 m^3/m^3. The maximum error of 0.0543 m^3/m^3 was observed during the vegetated period for the few bare-soil pixels that had fractional vegetation cover. Excluding these pixels and the bare-soil pixels at the end of the season when the LAI and LC contradict each other due to crop remnants, the error is 0.012 m^3/m^3. The maximum time-averaged error is moderately higher at $\leq$ 0.01 m^3/m^3 for the cotton pixels and 0.018 m^3/m^3 for corn pixels. The algorithm does tend to undervalue very high SMs and overvalue very low SMs due to the smoothing nature of regularized regression. This is needed to ensure that any noise in the correlated features does not lead to noise in downscaled SM.

The sensitivity of the downscaling algorithm to the auxiliary features used was investigated through a residual analysis of the error between downscaled SM and true SM for DoY 222. For this, the Pearson correlation coefficient was calculated between the error and each of the features, as shown in Table IV. A high correlation indicates that the algorithm is not very sensitive to the feature, while a low correlation indicates that the effect of the feature is fully accounted for by the downscaling algorithm. To understand how well the algorithm extends to obtain SM at even finer resolutions, the LST, LAI, PPT, and LC at 200 m are used along with SM at 10 km to

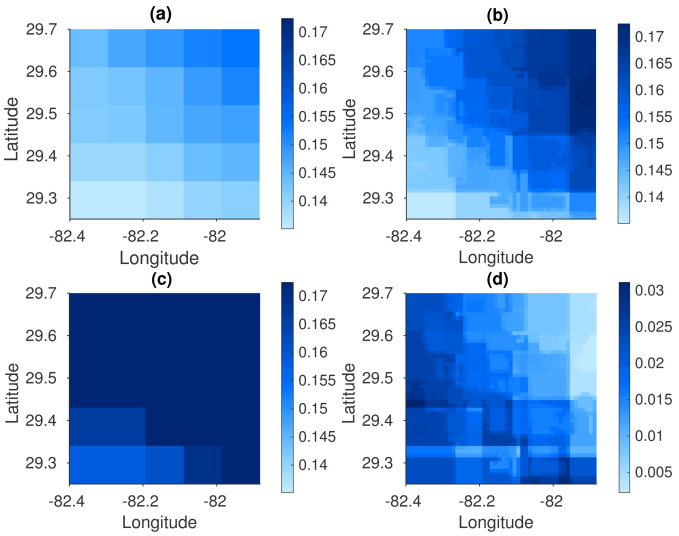

Fig. 12. DoY 39. (a) Coarse SM at 10 km. (b) True SM at 1 km. (c) Downscaled SM using the bagged regression trees algorithm with spatiotemporal training. (d) Absolute difference between true and downscaled SM.

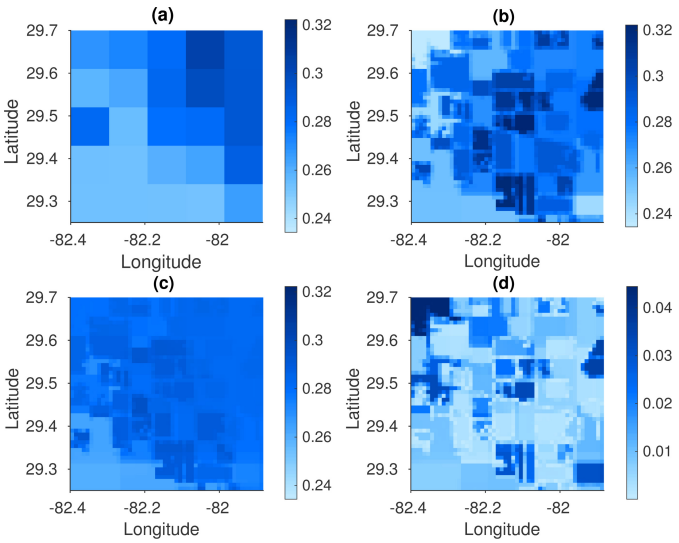

Fig. 13. DoY 354. (a) Coarse SM at 10 km. (b) True SM at 1 km. (c) Downscaled SM using the bagged regression trees algorithm with spatiotemporal training. (d) Absolute difference between true and downscaled SM.

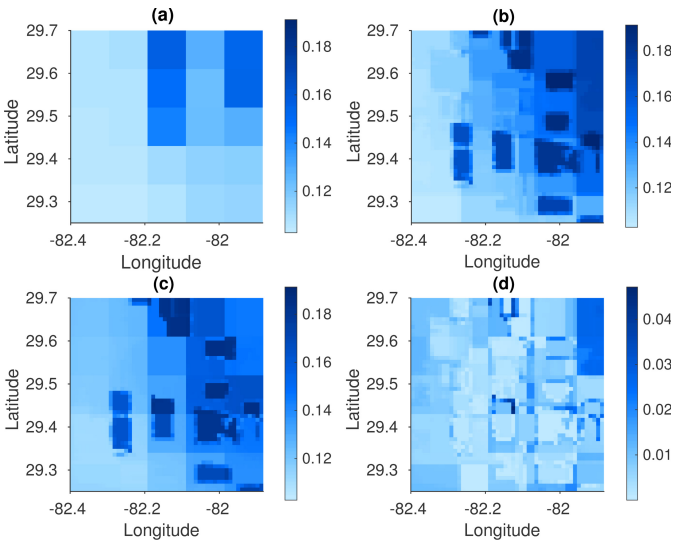

Fig. 14. DoY 135. (a) Coarse SM at 10 km. (b) True SM at 1 km. (c) Downscaled SM using the bagged regression trees algorithm with spatiotemporal training. (d) Absolute difference between true and downscaled SM.

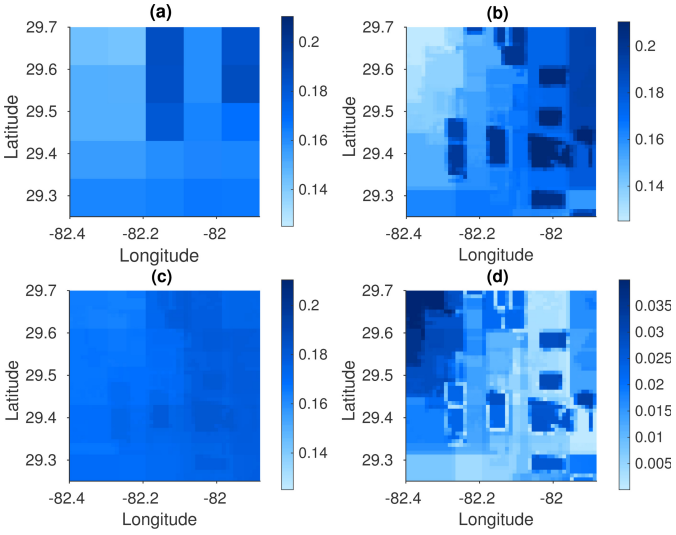

Fig. 15. DoY 156. (a) Coarse SM at 10 km. (b) True SM at 1 km. (c) Downscaled SM using the bagged regression trees algorithm with spatiotemporal training. (d) Absolute difference between true and downscaled SM.

obtain downscaled SM at 200 m. Theoretically, the limit of downscaling is the resolution at which the fine-scale estimates are available. However, downscaling algorithms attempt to estimate an inversion of the point-spread function that relates the coarse and fine-scale SM. This inverse function becomes increasingly ill-conditioned as the coarse scale and fine-scale resolutions get farther apart. Since the model simulations for this synthetic data set were performed at 200 m, the algorithm was implemented to obtain SM at that resolution for one day with the highest heterogeneity in features DoY 222. This is performed using the BRT_{st} algorithm on DoY 222. As shown in Fig. 17, the corn fields are visually apparent in the downscaled SM while the cotton fields are not. This is due to the higher differences between the corn and baresoil SM compared to the differences between the cotton and baresoil SM. The RMSE between the SM downscaled to 200 m and true SM at 200 m is 0.035 m^3/m^3, compared to 0.027 m^3/m^3 when the SM is downscaled to 1 km. Thus, the errors do increase when SM is disaggregated to finer resolutions, even when auxiliary data are available at those resolutions.

The computational cost for the BRT_{st} algorithm is $\mathcal{O}(mn)$, for m number for features and n number of pixels. Thus, in addition to low training data requirements, it can also be easily extended for regions with different geophysical attributes by including additional relevant features such as active observations at L and C band, slope, and soil type.

C. BRT_{st} Downscaling in the Presence of Data Gaps in LST

Typically, LST would be required at the time of downscaling. Because the BRT_{st} uses historical data, downscaling can be performed in spite of short data gaps in LST. To investigate the robustness of the algorithm during short data gaps in LST, an RT is trained without including LST information in the

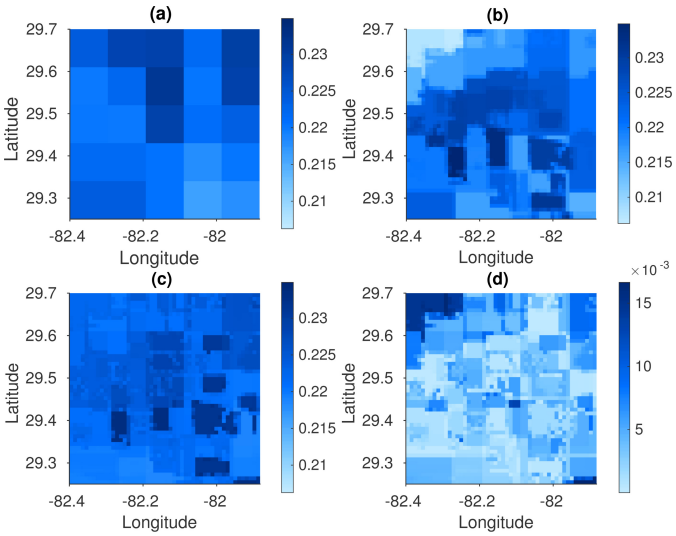

Fig. 16. DoY 222. (a) Coarse SM at 10 km. (b) True SM at 1 km. (c) Downscaled SM using the bagged regression trees algorithm with spatiotemporal training. (d) Absolute difference between true and downscaled SM.

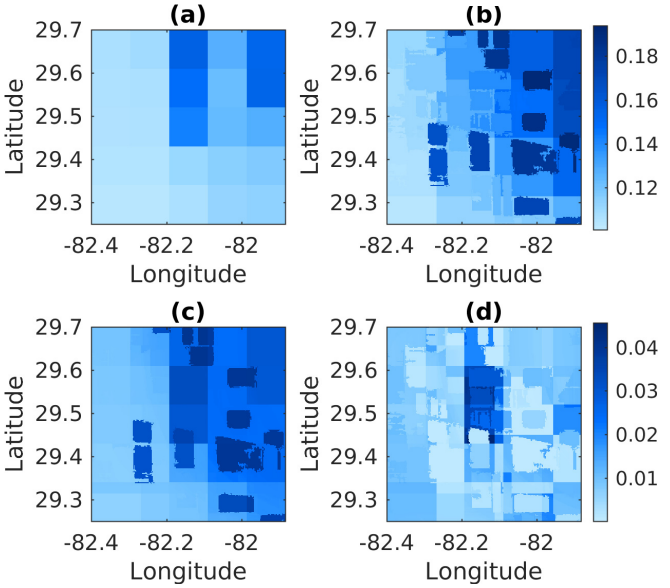

Fig. 17. (a) Coarse SM at 10 km, (b) true SM at 200 m, (c) downscaled SM at 200 m, and (d) difference between downscaled SM and true SM at 1 km for DoY 222, 2007.

feature vector. For this, (6) is modified as follows:

$$
\begin{aligned}
f^t(x_{nt}) &= \mathrm{SM}_{\mathrm{nt,insitu}} \\
x_{nt} &= \Big(\big[\mathrm{LST}^{1\ \mathrm{km}}_{(t-D_1)n}, \mathrm{LST}^{1\ \mathrm{km}}_{(t-D_1+1)n}, \ldots, \mathrm{LST1}^{1\ \mathrm{km}}_{tn}\big] \times U, \\
&\quad \big[\mathrm{LAI}^{1\ \mathrm{km}}_{(t-D_2)n}, \mathrm{LAI}^{1\ \mathrm{km}}_{(t-D_2+1)n}, \ldots, \mathrm{LAI}^{1\ \mathrm{km}}_{tn}\big] \\
&\quad \big[\mathrm{PPT}^{10\ \mathrm{km}}_{(t-D_3)n}, \mathrm{PPT}^{10\ \mathrm{km}}_{(t-D_3+1)n}, \ldots, \mathrm{PPT}^{1\ \mathrm{km}}_{tn}\big], \\
&\quad \mathrm{SM}^{10\ \mathrm{km}}_{tn}, \mathrm{X}^{10\ \mathrm{km}}_{tn}, \mathrm{Y}^{10\ \mathrm{km}}_{tn}\big]\Big) \\
U &= \begin{pmatrix} u_{(t-D_1)n} & 0 & \cdots & 0 \\ 0 & u_{(t-D_1+1)n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{tn} \end{pmatrix} \\
\hat{f}^t(x_{nt}) &= \mathrm{SM}_{\mathrm{nt,downscaled}} \qquad (8)
\end{aligned}
$$

where U is the unavailability matrix whose elements indicate the days on which LST is withheld from the feature-vector x_{nt}. For example, if u_{tn} is set to 0, LST on the day of downscaling is not available. The distinct scenarios investigated here are when $u_{tn} = 0$, $u_{tn} = u_{(t-1)n} = 0, \ldots, u_{tn} = u_{(t-1)n} = \cdots = u_{(t-D_1)n} = 0$.

Fig. 18 shows an increase in the average error in downscaled SM from the $\mathrm{BRT_{st}}$ algorithm when LST is "unavailable" for on, or upto three consecutive days prior to the day of downscaling, for DoY 222 in the year 2008. It also shows an increase in the number of pixels with error >0.04 $\mathrm{m^3/m^3}$. The error increases marginally by ≤ 0.015 $\mathrm{m^3/m^3}$ when LST is unavailable on the day of downscaling and one day prior to the day of downscaling. The error increases substantially to 0.05 $\mathrm{m^3/m^3}$ when LST is unavailable for three days and continues increasing with a slope of almost 0.015 $\mathrm{m^3/m^3}$ when the unavailability of LST is extended to further days. A similar trend is shown by the number of pixels with error in downscaled SM>0.04 $\mathrm{m^3/m^3}$, with the highest increase occurring at three days of unavailability. The $\mathrm{BRT_{st}}$ algorithm is robust to LST being unavailable on the day of downscaling and two days prior with upto 95% of the pixels having an error in downscaled SM of <0.04 $\mathrm{m^3/m^3}$.

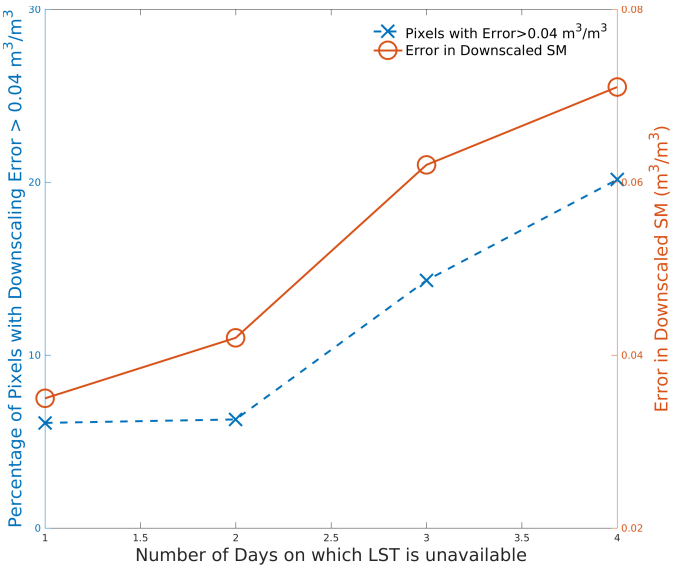

Fig. 18. Error in downscaled SM as a function of the number of gaps in LST data for the day of year 222.

V. Conclusion

In this paper, we implemented and evaluated a downscaling methodology based upon the BRT algorithm that utilizes spatiotemporal correlations in SM and other auxiliary variables including *in situ* SM to estimate fine-scale SM at 1 km. Because the availability of reliable *in situ* SM is usually sparse, the algorithm presented in this paper uses temporal interrelationships to reduce the quantity of high-resolution data points needed from 33% of the overall data set to just 0.02%, or 30 points in a 50×50 $\mathrm{km^2}$ area. However, in an operational scenario, measurement noise in *in situ* SM observations may necessitate additional training points or cause increase in average errors. For regions where no *in situ* is available, the training step can be performed in a different geographical area

with similar micrometeorology and soil characteristics. Other downscaling algorithms that do not use *in situ* SM are limited to using low-magnitude cross correlations between SM and other high-resolution RS data. This results in high errors in downscaled SM, especially in heterogeneous LCs and dynamic meteorological conditions.

Multiple RTs were trained using synthetically generated data that mimic *in situ* SM and RS products, namely, PPT, LST, LAI, and LC on the day of downscaling and the prior year. The best performing RTs were automatically chosen using ensemble pruning and their outputs were combined using bootstrap aggregation to generate the downscaled estimates. The time averaged error across the region was found to be 0.01 m^3/m^3, with a standard deviation of 0.012 m^3/m^3. The robustness of the BRT_{st} to simulated "data gaps" in LST was investigated and data gaps of upto three days, 95% of the pixels were found to have an error in downscaled SM of <0.04 m^3/m^3. Although the errors in downscaled SM are higher for this case than with all data available, this algorithm provides a reasonably accurate estimate which is within the error bounds dictated by SMAP mission requirements.

It is envisioned that the BRT algorithm evaluated in this paper may be applied using satellite-based higher resolution remote sensing data. For example, PPT data may be obtained from the Global PPT Measurement missions and the LAI, LST, and LC products are available from the MODIS sensor aboard Aqua and Terra satellites. Historical data for PPT are available from the NASA Tropical Rainfall Measurement Mission satellite from 1997 to 2015 and data for LAI, LST, and LC are available since 2002. The BRT_{st} algorithm can utilize these vast data sets to provide accurate high resolution SM. The algorithm can be extended to including additional geophysical features such as slope, soil properties, and active observations. The computational cost for the BRT_{st} algorithm is $\mathcal{O}(mn)$, for m number for features and n number of pixels.

ACKNOWLEDGMENT

The authors would like to thank the University of Florida High-Performance Computing Center for providing computational resources and support for all the model simulations conducted in this paper.

REFERENCES

[1] B. Narasimhan and R. Srinivasan, "Development and evaluation of soil moisture deficit index (SMDI) and evapotranspiration deficit index (ETDI) for agricultural drought monitoring," *Agricult. Forest Meteorol.*, vol. 133, nos. 1–4, pp. 69–88, 2005.

[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. Remote Sens.*, vol. 7, no. 9, pp. 3867–3879, Sep. 2014.

[3] M. Borga, E. N. Anagnostou, G. Blöschl, and J.-D. Creutin, "Flash flood forecasting, warning and risk management: The HYDRATE project," *Environ. Sci. Policy*, vol. 14, no. 7, pp. 834–844, 2016.

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

[5] Y. H. Kerr, P. Waldteufel, J. P. 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.

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

[7] J. Peng, A. Loew, O. Merlin, and N. Verhoest, "A review of spatial downscaling of satellite remotely-sensed soil moisture," *Rev. Geophys.*, vol. 55, pp. 1–26, Sep. 2017.

[8] 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.

[9] 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, Apr. 2002.

[10] A. Loew and W. Mauser, "On the disaggregation of passive microwave soil moisture data using *a priori* knowledge of temporally persistent soil moisture fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 3, pp. 819–834, Mar. 2008.

[11] G. Kim and A. P. Barros, "Space–time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data," *Remote Sens. Environ.*, vol. 81, pp. 393–403, Sep. 2002.

[12] R. Bindlish and A. P. Barros, "Subpixel variability of remotely sensed soil moisture: An inter-comparison study of SAR and ESTAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 2, pp. 326–337, Feb. 2002.

[13] M. Gebremichael, R. Rigon, G. Bertoldi, and T. M. Over, "On the scaling characteristics of observed and simulated spatial soil moisture fields," *Nonlinear Process. Geophys.*, vol. 16, pp. 141–150, Apr. 2009.

[14] S. Manfreda, M. F. McCabe, M. Fiorentino, I. Rodriguez-Iturbe, and E. F. Wood, "Scaling characteristics of spatial patterns of soil moisture from distributed modelling," *Adv. Water Res.*, vol. 30, pp. 2145–2150, Oct. 2007.

[15] M. Piles *et al.*, "A downscaling approach for SMOS land observations: Evaluation of high-resolution soil moisture maps over the Iberian Peninsula," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3845–3857, Sep. 2014.

[16] K. Ranney, J. Niemann, B. Lehman, T. Green, and A. Jones, "A method to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data," *Adv. Water Res.*, vol. 76, pp. 81–90, Feb. 2015.

[17] C. Hsu, L. Johnson, R. Zamora, T. Schneider, and R. Cifelli, "Downscaling advanced microwave scanning radiometer (AMSR-E) soil moisture retrievals using a multiple time-scale exponential rainfall adjustment technique," *J. Geophys. Remote Sens.*, vol. 4, no. 1, pp. 1–15, 2015.

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

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

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

[21] 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, Jan. 2013.

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

[23] O. Merlin, C. Rudiger, A. A. Bitar, P. Richaume, J. Walker, and Y. Kerr, "Disaggregation of SMOS soil moisture in Southeastern Australia," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1557–1571, May 2012.

[24] 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.

[25] 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.

[26] M. Piles, G. Petropoulos, N. Gonzalez-Zamora, and G. Ireland, "Towards improved spatio-temporal resolution soil moisture retrievals from the synergy of SMOS and MSG SEVIRI spaceborne observations," *Remote Sens. Environ.*, vol. 180, pp. 403–417, Jul. 2016.

[27] P. Srivastava, D. Han, M. Ramirez, and T. Islam, "Machine learning techniques for downscaling SMOS satellite soil moisture using modis land surface temperature for hydrological application," *Water Res. Manage.*, vol. 27, no. 8, pp. 3127–3144, 2013.

[28] N. E. C. Verhoest *et al.*, "Copula-based downscaling of coarse-scale soil moisture observations with implicit bias correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3507–3521, Jun. 2015.
[29] M. A. Babyak, "What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression type models," *Psychosomatic Med.*, vol. 66, no. 3, pp. 411–421, 2004.
[30] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification Regression Trees*. Wadsworth International Group, Belmont, CA, USA, 1984.
[31] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[32] 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, Apr. 2013.
[33] N. Li, Y. Yu, and Z.-H. Zhou, *Diversity Regularized Ensemble Pruning*. Berlin, Germany: Springer, 2012, pp. 330–345. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33460-3_27
[34] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc. B, (Methodol.)*, vol. 58, no. 1, pp. 226–288, 1996.
[35] I. Narsky and F. Porter, *Statistical Analysis Techniques in Particle Physics: Fits, Density Estimation and Supervised Learning*. Hoboken, NJ, USA: Wiley, 2013.
[36] T. Bongiovanni *et al.* (2009). "Field observations during the eighth microwave water and energy balance experiment (MicroWEX-8): From June 16 through August 24, 2009 center for remote sensing," Univ. Florida, Gainesville, FL, USA, Tech. Rep. AE476. [Online]. Available: http://edis.ifas.ufl.edu/ae476
[37] J. Casanova *et al.* (2006). "Page 1 circular 1514 field observations during the fifth microwave water and energy balance experiment (MicroWEX-5): From March 9 through May 26, 2006 center for remote sensing," Univ. Florida, Gainesville, FL, USA, Tech. Rep. AE407. [Online]. Available: http://edis.ifas.ufl.edu/AE407
[38] T. Lin *et al.* (2004). "Field observations during the third microwave, water, and energy balance experiment (MicroWEX-3): From June 16 through December 21, 2004 center for remote sensing," Univ. Florida, Gainesville, FL, USA, Tech. Rep. 1481. [Online]. Available: http://edis.ifas.ufl.edu/ae361
[39] W. H. Brutsaert, "On a derivable formula for long-wave radiation from clear skies," *Water Resour. Res.*, vol. 11, no. 5, pp. 742–744, 2010.
[40] W. Crow and E. Wood, "The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97," *Adv. Water Res.*, vol. 26, no. 2, pp. 137–149, 2003.
[41] C. Huang, X. Li, and L. Lu, "Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter," *Remote Sens. Environ.*, vol. 112, pp. 1320–1336, Apr. 2008.
[42] 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, Sep. 2002.
[43] 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, Jan. 2015.
[44] S. Chakrabarti, J. Judge, A. Rangarajan, and S. Ranka, "Downscaling microwave brightness temperatures using self regularized regressive models," in *Proc. IGARSS*, vol. 9. Jan. 2015, pp. 2580–2583.
[45] S. Ghosh and P. Mujumdar, "Statistical downscaling of GCM simulations to streamflow using relevance vector machine," *Adv. Water Res.*, vol. 31, pp. 132–146, Sep. 2007.
[46] C. Frey and C. Küenzer, "Two algorithms to fill cloud gaps in LST time series," in *Proc. EGU Gen. Assembly Conf. Abstracts*, vol. 15. Apr. 2013, p. 8475.

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 the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 2017.

He is currently a Post-Doctoral Research Associate with the Center for Remote Sensing, Department of Agricultural and Biological Engineering, University of Florida. His research interests include the application of novel machine learning techniques to remotely sensed observations.

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

She is currently the Director of the Center for Remote Sensing and an Associate Professor with the Agricultural and Biological Engineering Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA. Her research interests include microwave remote sensing applications to terrestrial hydrology for dynamic vegetation, modeling of energy and moisture interactions at the land surface and in the vadose zone, spatial and temporal scaling of remotely sensed observations in heterogenous landscapes, and data assimilation.

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

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

She is currently with the Center of Remote Sensing, University of Florida. She also supervises crop vegetation sampling, data collection, sensor monitoring, and data archiving for the center. Her master's research focused on modeling crop growth and development, and on improving the modeled estimates of biomass and yield through assimilating remote sensing observations. Her research interests include GIS work for decision support models, which includes processing field, and remotely sensed data.

Anand Rangarajan (M'90) received the Ph.D. degree from the University of Southern California, Los Angeles, CA, USA.

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

Sanjay Ranka (S'87–M'88–SM'01–F'02) received the B.Tech. degree in computer science from IIT Kanpur, Kanpur, India, and the Ph.D. degree in computer science from the University of Minnesota, Minneapolis, MN, USA.

He was the Chief Technology Officer with Paramark, Baltimore, MD, USA, where he developed real-time optimization software for optimizing marketing campaigns. He is currently a Professor with the Department of Computer Information Science and Engineering, University of Florida. He has co-authored four books, 225 journals, and refereed conference articles. His current research interests include variety of issues related to data science: energy efficient computing, high-performance computing, data mining, and informatics.

Dr. Ranka is a fellow of AAAS. His research has received the Best Student Paper Runner Up Award at IGARRS 2015 and BICOB 2014, the Best Student Paper Award at ACM-BCB 2010, and the Best Paper Runner Up Award at KDD-2009. He is an Associate Editor-in-Chief of the *Journal of Parallel and Distributed Computing* and an Associate Editor for the ACM/IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, SUSTAINABLE COMPUTING: SYSTEMS AND INFORMATICS, KNOWLEDGE AND INFORMATION SYSTEMS, and the *International Journal of Computing.* He was an Editor of the IEEE TRANSACTIONS ON COMPUTERS and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS.