Global Soil Moisture Estimation by Assimilating AMSR-E Brightness Temperatures in a Coupled CLM4–RTM–DART System

LONG ZHAO AND ZONG-LIANG YANG
Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas

TIMOTHY J. HOAR
National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 5 November 2015, in final form 3 July 2016)

ABSTRACT

Very few frameworks exist that estimate global-scale soil moisture through microwave land data assimilation (DA). Toward this goal, such a framework has been developed by linking the Community Land Model, version 4 (CLM4), and a microwave radiative transfer model (RTM) with the Data Assimilation Research Testbed (DART). The deterministic ensemble adjustment Kalman filter (EAKF) within DART is utilized to estimate global multilayer soil moisture by assimilating brightness temperature observations from the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E). A 40-member ensemble of Community Atmosphere Model, version 4.0 (CAM4.0), reanalysis is adopted to drive CLM4 simulations. Space-specific, time-invariant microwave parameters are precalibrated to minimize uncertainties in RTM. Besides, various methods are designed to upscale AMSR-E observations for computational efficiency and time shift CAM4.0 forcing to facilitate global daily assimilations. A series of experiments are conducted to quantify the DA sensitivity to microwave parameters, choice of assimilated observations, and different CLM4 updating schemes. Evaluation results indicate that the newly established CLM4–RTM–DART framework improves the open-loop CLM4-simulated soil moisture. Precalibrated microwave parameters, rather than their default values, can ensure a more robust global-scale performance. In addition, updating near-surface soil moisture is capable of improving soil moisture in deeper layers (0–30 cm), while simultaneously updating multilayer soil moisture fails to obtain intended improvements. Future work is needed to address the systematic bias in CLM4 that cannot be fully covered through the ensemble spread in CAM4.0 reanalysis.

1. Introduction

Reliable global soil moisture datasets are urgently needed for a wide range of applications including weather forecasting (de Goncalves et al. 2006; Drusch 2007), drought monitoring (Dai et al. 2004; Bolten et al. 2010), climate prediction (Dirmeyer et al. 2006; Bao et al. 2010), and water resource management (Blum 2005; Dobriyal et al. 2012). However, conventional soil-moisture-retrieving methods such as ground-based measurements (Robinson et al. 2008; Crow et al. 2012; Romano 2014), land surface modeling (Henderson-Sellers et al. 1993; Entin et al. 1999), and remote sensing (Kerr 2007; Loew et al. 2013) are subject to limitations in data quality, spatiotemporal resolution, or representativeness and coverage.

Recently developed land data assimilation (DA) techniques (Liang and Qin 2008; Reichle 2008; Tian et al. 2011) hold promise for estimating soil moisture with good accuracy by integrating soil wetness information from both model simulations and independent satellite observations (Vereecken et al. 2008). Considering the uncertainties in satellite soil moisture products (Jackson et al. 2010; Su et al. 2011; Chen et al. 2013; Al-Yaari et al. 2014), some researchers prefer to directly assimilate microwave brightness temperatures (TB) through radiative transfer models (RTMs; Margulis et al. 2002; Crow and Wood 2003; Yang et al. 2007; Loew et al. 2009; Shi et al. 2010). Model parameters are sometimes updated together with soil moisture to minimize systematic biases in land surface models (LSMs)
or RTMs (Moradkhani et al. 2005; Pan and Wood 2006; Reiche et al. 2008; Qin et al. 2009; Nie et al. 2011; Han et al. 2014), which greatly improve the DA performance. Based on these concepts, microwave land DA has been well tested and applied at catchment or regional scales (Crow and Van den Berg 2010; Tian et al. 2010; Montzka et al. 2011), and its global application is recently receiving an increasing amount of attention in the land DA community (De Lannoy et al. 2013; Harding et al. 2014), with major challenges in accounting for the heterogeneity in soil properties, land use, vegetation cover, and topography (Reiche et al. 2014).

Advanced LSMs play a fundamental role in the global soil moisture estimation framework. As a land component of the Community Earth System Model (CESM; Hurrell et al. 2013), the Community Land Model (CLM; Oleson et al. 2004; Lawrence and Chase 2007; Oleson et al. 2010) has a state-of-the-art representation of the hydrologic cycle, in which soil moisture is simulated using a refined layering scheme to resolve the strong near-surface soil moisture gradient (Z. Yang 1998, unpublished manuscript). Despite the enhancement in hydrological modeling, CLM is found to be biased in estimating soil moisture at global (Du et al. 2014) and regional (Long et al. 2013; Cai et al. 2014) scales. The past decade has also seen many CLM-based land DA studies aiming to improve soil moisture estimation by assimilating either in situ (De Lannoy et al. 2007; Tian et al. 2008; Zhang et al. 2012) or synthetic (Kumar et al. 2009; Han et al. 2014) observations, and ensemble-based DA has been proven to represent the model uncertainties well (De Lannoy et al. 2006). Although assimilating soil moisture in the vertical profile directly results in improved estimates of root-zone soil moisture (De Lannoy et al. 2007), there are also studies showing potential improvements of deep-soil states through the assimilation of surface soil moisture observations (Kumar et al. 2009). The latter finding is especially useful for global root-zone soil moisture estimation through the assimilation of satellite microwave TB in regions where no in situ soil moisture observations are available (Tian et al. 2010).

This study aims to improve CLM-estimated global soil moisture by assimilating Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) TB observations. To this end, we have developed a prototype of a microwave land DA system that couples CLM, version 4 (CLM4; Lawrence et al. 2011), and a microwave RTM into the Data Assimilation Research Testbed (DART; https://www.image.ucar.edu/DARes/DART/; Anderson et al. 2009). Methods in time shifting of forcing time series, upscaling of AMSR-E observations, and precalibration of RTM parameters are designed to speed up the computational procedure and to account for the uncertainties in the radiative transfer process. The resultant multilayer soil moisture is then evaluated against the available in situ measurements over the globe. Details on the DA system including CLM4, RTM, DART, datasets, and experimental design are provided in section 2. Section 3 presents evaluation results and discussion. Finally, a summary is given in section 4.

### 2. Methodology and data

#### a. CLM4

CLM4 is an LSM that benefits from the community efforts in developing and refining model parameterizations and physical structures (Lawrence et al. 2011). Integrated in CESM (Kluzek 2013), CLM4 is well configured for parallel computation and has the “multi-instance” capability to simultaneously run multiple simulations within a single executable, which simplifies the setup of ensemble-based DA. We choose to run the offline CLM4 at a $0.9^\circ \times 1.25^\circ$ spatial resolution, with each grid cell consisting of one or more columns representing spatial heterogeneity induced by various land units (e.g., wetland, lake, urban, glacier, and vegetated area). In each column, the soil profile is divided into 15 layers and soil moisture is predicted from a multilayer model, where vertical soil water transport is governed by infiltration, runoff, gradient diffusion, gravity, canopy transpiration, and interactions with groundwater. The detailed layering information for the top 10 CLM4 soil layers is given in Table 1. A modified Richards equation (Zeng and Decker 2009) is applied to maintain the hydrostatic equilibrium of soil moisture distribution. Soil hydraulic and thermal parameters are derived from input soil ancillary data (e.g., sand, clay, or RTMs (Moradkhani et al. 2005; Pan and Wood 2006; Reiche et al. 2008; Qin et al. 2009; Nie et al. 2011; Han et al. 2014), which greatly improve the DA performance. Based on these concepts, microwave land DA has been well tested and applied at catchment or regional scales (Crow and Van den Berg 2010; Tian et al. 2010; Montzka et al. 2011), and its global application is recently receiving an increasing amount of attention in the land DA community (De Lannoy et al. 2013; Harding et al. 2014), with major challenges in accounting for the heterogeneity in soil properties, land use, vegetation cover, and topography (Reiche et al. 2014).

Advanced LSMs play a fundamental role in the global soil moisture estimation framework. As a land component of the Community Earth System Model (CESM; Hurrell et al. 2013), the Community Land Model (CLM; Oleson et al. 2004; Lawrence and Chase 2007; Oleson et al. 2010) has a state-of-the-art representation of the hydrologic cycle, in which soil moisture is simulated using a refined layering scheme to resolve the strong near-surface soil moisture gradient (Z. Yang 1998, unpublished manuscript). Despite the enhancement in hydrological modeling, CLM is found to be biased in estimating soil moisture at global (Du et al. 2014) and regional (Long et al. 2013; Cai et al. 2014) scales. The past decade has also seen many CLM-based land DA studies aiming to improve soil moisture estimation by assimilating either in situ (De Lannoy et al. 2007; Tian et al. 2008; Zhang et al. 2012) or synthetic (Kumar et al. 2009; Han et al. 2014) observations, and ensemble-based DA has been proven to represent the model uncertainties well (De Lannoy et al. 2006). Although assimilating soil moisture in the vertical profile directly results in improved estimates of root-zone soil moisture (De Lannoy et al. 2007), there are also studies showing potential improvements of deep-soil states through the assimilation of surface soil moisture observations (Kumar et al. 2009). The latter finding is especially useful for global root-zone soil moisture estimation through the assimilation of satellite microwave TB in regions where no in situ soil moisture observations are available (Tian et al. 2010).

This study aims to improve CLM-estimated global soil moisture by assimilating Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) TB observations. To this end, we have developed a prototype of a microwave land DA system that couples CLM, version 4 (CLM4; Lawrence et al. 2011), and a microwave RTM into the Data Assimilation Research Testbed (DART; https://www.image.ucar.edu/DARes/DART/; Anderson et al. 2009). Methods in time shifting of forcing time series, upscaling of AMSR-E observations, and precalibration of RTM parameters are designed to speed up the computational procedure and to account for the uncertainties in the radiative transfer process. The resultant multilayer soil moisture is then evaluated against the available in situ measurements over the globe. Details on the DA system including CLM4, RTM, DART, datasets, and experimental design are provided in section 2. Section 3 presents evaluation results and discussion. Finally, a summary is given in section 4.

### 2. Methodology and data

#### a. CLM4

CLM4 is an LSM that benefits from the community efforts in developing and refining model parameterizations and physical structures (Lawrence et al. 2011). Integrated in CESM (Kluzek 2013), CLM4 is well configured for parallel computation and has the “multi-instance” capability to simultaneously run multiple simulations within a single executable, which simplifies the setup of ensemble-based DA. We choose to run the offline CLM4 at a $0.9^\circ \times 1.25^\circ$ spatial resolution, with each grid cell consisting of one or more columns representing spatial heterogeneity induced by various land units (e.g., wetland, lake, urban, glacier, and vegetated area). In each column, the soil profile is divided into 15 layers and soil moisture is predicted from a multilayer model, where vertical soil water transport is governed by infiltration, runoff, gradient diffusion, gravity, canopy transpiration, and interactions with groundwater. The detailed layering information for the top 10 CLM4 soil layers is given in Table 1. A modified Richards equation (Zeng and Decker 2009) is applied to maintain the hydrostatic equilibrium of soil moisture distribution. Soil hydraulic and thermal parameters are derived from input soil ancillary data (e.g., sand, clay,
and organic contents) through pedotransfer functions (Clapp and Hornberger 1978; Cosby et al. 1984; Lawrence and Slater 2008). Note that only soil moisture within the top 10 hydrologically activated layers (0–3.8 m) is calculated, and the bottom five layers are retained to account for thermal interactions with the underlying deep ground.

Despite the advanced physical design of this model, CLM4 still poses challenges for data assimilation. The location information is stored for the grid cells alone, yet the model contains subgrid prognostic variables (e.g., columns). Upscaling from these subgrid variables to gridcell-based quantities is a necessity in the current system because it is impossible to discern what part(s) of the grid cell is the appropriate land unit, column, or plant functional type that would match the characteristics of the TB observation location. Therefore, in the DA system, a gridcell average of soil moisture and temperature is first aggregated from corresponding columns and then used to calculate an observational estimate of TB through RTM. CLM4-predicted soil moisture is then updated indirectly by comparing the modeled and AMSR-E observed TB.

b. RTM and parameter calibration

According to the microwave remote sensing theory, near-surface soil moisture is highly correlated with TB through soil dielectric constant that influences soil microwave emissivity and eventually TB. In addition, the presence of vegetation causes attenuation of the emitted microwave signals. A detailed radiative transfer process involving soil, vegetation, and topography is depicted in appendix A. According to Eqs. (A1)–(A11), the estimated TB can be generally expressed as a function of time-varying model states ($\theta, p, Freq, T_g, T_v, SM,$ and LAI) and time-invariant parameters (%sand, %clay, porosity, qmv, hmv, bmv, and xmv), where $\theta$ is the satellite incidence angle; $p$ denotes microwave polarization (vertical or horizontal); $Freq$ is the microwave frequency; $T_g, T_v,$ and SM are CLM4-predicted near-surface soil temperature, vegetation temperature, and near-surface soil moisture, respectively; LAI is leaf area index; %sand, %clay, and porosity are soil texture parameters; qmv and hmv are soil surface roughness–related parameters; and bmv and xmv are vegetation-related parameters.

Uncertainties from the above variables will potentially introduce extra biases into the RTM-calculated TB. Taking the vertically polarized, 6.9-GHz TB as an example, Fig. 1 shows the sensitivity of RTM-estimated TB (TB$_{est}$). For each parameter, the relationship is derived by first randomly sampling 2000 values within its physical range (‘‘min’’ and ‘‘max’’ in Table 2) while assigning all other parameters with a specified value (‘‘default’’ in Table 2), and then applying these 2000 combinations of parameters in RTM to obtain the same number of TB$_{est}$. Specifically, TB$_{est}$ shows high sensitivity to time-varying variables (e.g., $T_g, SM,$ and LAI) except for $T_v$. Since both $T_g$ and SM may have large

![Fig. 1. Relationship between RTM-estimated TB (vertically polarized, 6.9 GHz) and input land states and RTM parameters. The x axis indicates the randomly sampled parameter values within its physical range.](image-url)
diurnal variations, it is necessary to estimate TB at exactly the time when AMSR-E TB is observed, thereby requiring time matching between CLM4 states and AMSR-E observations if conducting assimilation at global scale with daily intervals. This is realized by rendering the model time from the consistent coordinate universal time (UTC) to the same local solar time (LST) through the time shifting of forcing data. By contrast, the accurate input of LAI is ensured by introducing satellite observational products. In terms of time-invariant parameters, $T_{\text{Best}}$ shows a relatively low sensitivity to soil texture parameters (e.g., $%\text{sand}$, $%\text{clay}$, and porosity), and thus they are directly inherited from CLM4 ancillary data. On the other hand, $T_{\text{Best}}$ shows a high sensitivity to vegetation and soil surface roughness parameters (e.g., $b_{\text{mv}}$, $h_{\text{mv}}$, and $q_{\text{mv}}$), except for $x_{\text{mv}}$, and therefore, they are calibrated prior to the data assimilation. However, since the highly packaged engineering design of CLM4 prevents the feasibility of using it for global-scale parameter calibration, we have followed Yang et al. (2007) to obtain the calibrated RTM parameters (see details in appendix B) when developing our prototype CLM4–RTM–DART system. Note that this calibration scheme is different from de Rosnay et al. (2009), who compare different combinations of land, vegetation, and soil dielectric parameterization schemes in the forward model. This study also differs from De Lannoy et al. (2013, 2014) in terms of calibrating key RTM parameters for the L-band TB estimation.

The calibration period is from 1 June to 1 October 2010. Considering the relative low sensitivity of TB to $x_{\text{mv}}$, only $b_{\text{mv}}$ was calibrated in Eq. (A5) to avoid overfitting and $x_{\text{mv}}$ was set to a constant value ($x_{\text{mv}} = -1.25$) following Jackson and Schmugge (1991). Precalibration was not implemented at regions within the Arctic and Antarctic Circles to avoid snowy or icy conditions. Finally, the precalibrated radiative transfer parameters (i.e., $b_{\text{mv}}$, $q_{\text{mv}}$, and $h_{\text{mv}}$, with their spatial distribution shown in Fig. 2) were used in the RTM module in our CLM4–RTM–DART system for AMSR-E TB assimilation.

c. DART and the CAM4.0 reanalysis

DART has incorporated a variety of mainstream DA algorithms, including the traditional ensemble Kalman filter (EnKF), the deterministic ensemble adjustment

![Fig. 2. Global distribution of precalibrated radiative transfer parameters ($q_{\text{mv}}$, $b_{\text{mv}}$, and $h_{\text{mv}}$) and “cost-min” denotes the minimum of the cost function (defined in appendix B) obtained in the calibration process.](image-url)
Kalman filter (EAKF) (Anderson 2001), and particle filter (PF). Zhang (2015) shows that EAKF outperforms EnKF in snow DA using observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) and Gravity Recovery and Climate Experiment (GRACE). Thus, EAKF is used in this study to update CLM4 states. By assimilating a specifically observed quantity, EAKF is capable of updating each state vector component based on their correlation (Anderson et al. 2009). On the one hand, this supports our current study that updates soil moisture and other CLM4 state variables through the assimilation of AMSR-E TB; on the other hand, this forms the theoretic basis of future studies on multisource land DA where there might be no apparent relationships between CLM4 state variables and observations.

The 6-hourly Community Atmosphere Model, version 4.0 (CAM4.0), reanalysis with spatial resolution of 1.9° × 2.5° is used as forcing data in this study. This reanalysis is generated by interfacing CAM4.0 to DART and assimilating a variety of observatories, including the radio occultation observations from the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC), temperature and wind components from radiosondes and aircraft, and satellite drift winds (Raeder et al. 2012). Compared with other ensemble-based DA that simply perturb one or several atmospheric forcing fields, the CAM4.0 ensemble reanalysis preserves the full covariance of each field while maintaining variability consistent with observational uncertainty. Based on the CAM4.0 reanalysis, the coupled CLM4 and DART has been used to estimate snow water equivalent in the Northern Hemisphere through the assimilating of MODIS snow cover fraction (Zhang et al. 2014), microwave radiance (Kwon et al. 2015), and joint MODIS snow cover fraction and GRACE terrestrial water storage (Zhang 2015). In this study, considering the computational cost and the EAKF performance (Reichle et al. 2002), we have randomly chosen 40 members of the CAM4.0 reanalysis to introduce uncertainties into CLM4 states. Note that in order to facilitate the assimilation of global AMSR-E observations at a daily interval, we have shifted the original UTC-formatted CAM4.0 reanalysis into LST format to indirectly render CLM4 to match with AMSR-E in time. A justification for the time shifting of CAM4.0 reanalysis as well as bias in the forcing is addressed in section 3c.

d. Satellite data

The “AMSR-E/Aqua daily global quarter-degree gridded brightness temperatures” (Knowles et al. 2011) available at the National Snow and Ice Data Center (NSIDC) are used. The AMSR-E sensor is on board the NASA EOS Aqua satellite, which has a sun-synchronous orbit and crosses the equator at about 1330 LST (0130 LST) in the ascending (descending) mode each day, thereby producing a daily global map with approximately the same observational time (LST). Multichannel passive microwave signals were measured at six frequencies (6.9, 10.7, 18.7, 23.8, 36.5, and 89.0 GHz) and at both the horizontal and vertical polarizations with an incidence angle of 55°. In this study, the descending (nighttime around 0130 LST), vertically polarized signals are assimilated for soil moisture estimation following Zhao et al. (2013).

In DART, neighboring observations within a certain distance are allowed to impact the model states (Anderson et al. 2009). Thus, the computational load increases greatly with the increase in the observation number. Therefore, we have spatially upsampled the original 1/4° AMSR-E TB onto the CLM4 resolution (0.9° × 1.25°), which largely reduces the total number of AMSR-E observations and significantly saves the computing time in the practical implementation of a single DA cycle (from 1 h to less than 9 min). Justification for the AMSR-E TB upscaling is discussed in section 3d. Note that AMSR-E pixels along coastlines and over inland water bodies are eliminated during upscaling. Regarding the common radio-frequency interference (RFI) contamination at 6.9 GHz, monthly global RFI masks are first generated following the method of Njoku et al. (2005) and then applied prior to the assimilation. In terms of microwave frequency, only those less than 19 GHz are assimilated, as they are more sensitive to changes in near-surface soil moisture than in vegetation (Vinnikov et al. 1999; Prigent et al. 2005). Specifically, the relatively higher 18.7-GHz TB is assimilated globally in some data
assimilation cases (Table 3), while 6.9 GHz is the preferred lower frequency. However, at some locations where there is severe RFI contamination at 6.9 GHz (e.g., the contiguous United States and India), the 10.7-GHz TB is assimilated instead. Since there is no officially publicized observation error in the AMSR-E TB observations, we have set the observational error variance to 2 K, which approximately corresponds to 1% of the TB values observed at frequencies of 18.7 GHz or lower.

In addition to AMSR-E TB, the 0.05° Global Land Surface Satellites (GLASS) LAI product that developed and released by the Center for Global Change Data Processing and Analysis of Beijing Normal University (Liang and Xiao 2012) is adopted to provide auxiliary vegetation dynamic information in the radiative transfer processes (see appendix A). The GLASS LAI is produced based on LAI products of AVHRR (1982–99) and MODIS (2000–12) by using the so-called general regression neural networks (GRNNs) to refill spatio-temporal missing values (Xiao et al. 2014). Note that in order to alleviate the variable location issues and to reduce the total number of global observations, we have also upscaled the original 0.05° GLASS LAI into a coarser CLM4 resolution (0.9° × 1.25°).

e. The CLM4–RTM–DART system and DA experiments

As shown in Fig. 3, CLM4 fully utilizes its multi-instance capability to facilitate the ensemble-based DA. With the input of time-shifted CAM4.0 ensemble forcing and initial conditions, CLM4 is stopped every 24 h (nominally at 0000 UTC but actually corresponding to LST when AMSR-E TB is observed), and restart files are generated. DART then reads a subset of the CLM4 states (near-surface soil moisture, soil temperature, vegetation temperature, etc.) and employs the coupled RTM (the observation operator) to calculate an ensemble of prior TB with the input of GLASS LAI and precalibrated parameters, and the EAKF performs DA at all CLM4 grid cells. Note that no additional perturbations are applied to land states during the cycling of the assimilation experiment, and the spread among ensemble members is generated from ensemble forcing. A further discussion on ensemble spread is presented in section 3a(1). The updated CLM4 states are inserted back to the restart files and used as background initial conditions for the next forecast cycle.

Based on the newly established CLM4–RTM–DART system, several experimental cases were designed to investigate the system performance in global soil moisture estimation. Table 3 shows an open-loop (OL) run of CLM4 without DA and two DA runs using precalibrated (DA_1) and default (DA_0) microwave parameters in the RTM. Note that parameter values used in the latter case are listed in Table 2 (see “default” for bmv, xmv, qmv, and hmv). Both DA experiments assimilate the lower frequency of vertically polarized, nighttime AMSR-E TB to update the top two layers’ soil moisture in the CLM4 restart fields. Besides, three additional DA cases (DA_2, DA_3, and DA_4) were conducted by assimilating both the lower and higher frequencies of AMSR-E TB but with different updating schemes on CLM4 states. In contrast to DA_1 and DA_2
FIG. 4. (top) Global distribution of selected subregions for soil moisture evaluation and (bottom) the spatially averaged topsoil moisture time series. In each subplot of the bottom panel, the multiyear mean (black solid line) with standard deviations (shaded areas) and the 2010 soil moisture time series (blue solid line) are shown. Temporal coverage of multiyear time series is indicated in the subtitle.
that only update topsoil moisture, DA_3 simultaneously updates topsoil moisture, topsoil temperature, and vegetation temperature, while DA_4 updates multilayer soil moisture within the top eight soil layers.

In all the above cases, CLM4 was configured to run at the hourly time step at a 0.9° × 1.25° spatial resolution over the globe. A set of 40-member initial conditions were generated by using 40 members of CAM4.0 reanalysis to drive the same number of CLM4 simulations from 1 January to 1 June 2010. In DA cases (from DA_0 through DA_4), the CLM4-predicted near-surface soil moisture, soil temperature, and vegetation temperature are areal averaged within each grid cell to calculate TB. The overall DA period is from 1 June to 1 October 2010. In terms of localization, we have referred to Zhang et al. (2014) and conducted a series of localization-distance-based experiments and found the DA performance, judging from the prior and posterior root-mean-square error (RMSE), is insensitive to localization settings (not shown here for conciseness). This is reasonable as discontinuities are commonly found in land surface modeling due to high spatial heterogeneity. Finally, a constant localization distance of 1.7° (0.03 rad) is adopted in this study.

### f. In situ soil moisture observations and the evaluation scheme

Two in situ observational databases, namely, the International Soil Moisture Network (ISMN) and the North American Soil Moisture Database (NASMD), were used to evaluate the CLM4–RTM–DART-estimated soil moisture. ISMN was initiated by the International Soil Moisture Working Group (ISMWG) to facilitate the continuous global in situ soil moisture observations and the standardization of measuring technique and protocol (Dorigo et al. 2011). The ISMN database consists of worldwide voluntary contributions of scientists and networks and has hosted 35 networks with over 1400 stations by November 2012. In contrast to ISMN, the NASMD project is operated by scientists at...
Texas A&M University, aiming to understand how soil moisture influences climate on seasonal to interannual time scales in North America (see details on http://soilmoisture.tamu.edu/). The NASMD holds a more comprehensive collection of in situ soil moisture observations in all 50 United States as well as in Canada and Mexico. Originated from various observational networks and institutions, soil moisture observations from both ISMN and NASMD have been harmonized and quality controlled to ease their applications (see details at http://soilmoisture.tamu.edu/ and http://ismn.geo.tuwien.ac.at/). In general, ISMN and NASMD are complementary to each other and are both essential means for the geoscientific community to validate and improve global satellite observations and LSMs.

Because of high spatial heterogeneity, the representativeness of soil moisture observations at a single station is usually questioned and thus requires spatial upscaling to eliminate the scale mismatch between point-scale measurements and model (or satellite) gridded estimations (or retrievals; Crow et al. 2012). Given the coarse spatial resolution of CLM4-produced soil moisture, and further considering the diverse protocols in measuring depth, density, sensor, and sampling interval among different soil moisture networks, we have conducted a network-based evaluation within 12 specified subregions over the globe. The spatial distribution of these selected subregions as well as their multiyear mean and 2010 topsoil moisture time series are shown in Fig. 4. As detailed in Table 4, in addition to the aforementioned factors, these subregions are preselected with respect to vegetation cover, climatology, and data availability (June–October 2010), and each has at least eight soil moisture monitoring stations and four CLM4 grid cells. Similar to studies of Entin et al. (1999) and Xia et al. (2014), we have simply applied spatial averaging to both the station observations and CLM4 estimations within each subregion. Besides, the soil moisture measuring depths are different among subregions. Statistically, depths of 5, 10, and 20 cm possess the most effective observations from both NASMD and ISMN (see Table 5). Meanwhile, there are also stations with sensors (e.g., CS-616 in OZNET) measuring soil moisture within the top 0–30 cm, which can be aggregated from the 5-, 10-, and 20-cm observations. Therefore, soil moisture at three depths (5, 10, and 20 cm) plus one layer (0–30 cm) are chosen for evaluation (Table 6). Specifically, soil moisture centered at CLM4 layers of \( L_3 \), \( L_4 \), and \( L_5 \) are compared with observations at 5, 10, and 20 cm, respectively, whereas the aggregated CLM4 \( L_1-L_5 \) soil moisture is compared with the 0–30-cm observation. Note that near-surface soil moisture at OZNET was measured at 0–5 or 0–8 cm, which varies with stations, and here we treat these measurements as taken at 5 cm (see Table 4). In contrast to OZNET, near-surface soil moisture at CTP_SMTMN is measured at 0–5 cm; thus, we aggregated the top two CLM4 layers to get the corresponding estimates for evaluation. In the following descriptions, we consider topsoil moisture is nominally measured at 5 cm in subregions OZNET and CTP_SMTMN for simplicity.

Given the relatively short 4-month DA, this global dataset is expected to enhance the evaluation results over the spatial dimension. Meanwhile, different soil wetness conditions in these subregions, that is, the 2010 soil moisture compared with multiyear mean (Fig. 4), are used as a surrogate to indirectly represent the temporal variations and thus compensate the evaluation over the temporal dimension. In each subregion, the topsoil shows a drier (e.g., SCAN_R04), wetter (e.g., AARD), or similar level (e.g., SCAN_R02) of wetness condition as compared to its multiyear mean climatology. For each subregion, three statistic metrics, namely, the correlation coefficient \( R \), mean bias error (BIAS), and RMSE, were derived by comparing the

<table>
<thead>
<tr>
<th>Evaluating depth/layer (cm)</th>
<th>CLM4 layers (depth, cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>( L_3 ) (−6)</td>
</tr>
<tr>
<td>10</td>
<td>( L_4 ) (−11)</td>
</tr>
<tr>
<td>20</td>
<td>( L_5 ) (−21)</td>
</tr>
<tr>
<td>0–30</td>
<td>( L_1-L_5 ) (−0–28)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In situ (cm)</th>
<th>NASMD</th>
<th>ISMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>317</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>163</td>
<td>50</td>
</tr>
<tr>
<td>20</td>
<td>329</td>
<td>50</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>45</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>55</td>
<td>266</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>57</td>
<td>0</td>
</tr>
<tr>
<td>0–5</td>
<td>0</td>
<td>47</td>
</tr>
<tr>
<td>0–10</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>0–30</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>30–60</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>55–85</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>60–90</td>
<td>0</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluating depth/layer (cm)</th>
<th>CLM4 layers (depth, cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>( L_3 ) (−6)</td>
</tr>
<tr>
<td>10</td>
<td>( L_4 ) (−11)</td>
</tr>
<tr>
<td>20</td>
<td>( L_5 ) (−21)</td>
</tr>
<tr>
<td>0–30</td>
<td>( L_1-L_5 ) (−0–28)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluating depth/layer (cm)</th>
<th>CLM4 layers (depth, cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>( L_3 ) (−6)</td>
</tr>
<tr>
<td>10</td>
<td>( L_4 ) (−11)</td>
</tr>
<tr>
<td>20</td>
<td>( L_5 ) (−21)</td>
</tr>
<tr>
<td>0–30</td>
<td>( L_1-L_5 ) (−0–28)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluating depth/layer (cm)</th>
<th>CLM4 layers (depth, cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>( L_3 ) (−6)</td>
</tr>
<tr>
<td>10</td>
<td>( L_4 ) (−11)</td>
</tr>
<tr>
<td>20</td>
<td>( L_5 ) (−21)</td>
</tr>
<tr>
<td>0–30</td>
<td>( L_1-L_5 ) (−0–28)</td>
</tr>
</tbody>
</table>
experimental case–estimated and network-observed daily soil moisture time series.

3. Results and discussion

a. General results

1) MODEL BIAS

The overall evaluation results are illustrated in Figs. 5–7. Take SCAN_R01 as an example, all the DA cases that use precalibrated RTM parameters (from DA_1 through DA_4) slightly outperform OL with slightly higher correlation coefficients, less BIAS, and smaller RMSE values in estimating multilayer (except 20 cm) soil moisture. A similar performance is also found in the DA case with the default RTM parameters (DA_0) but with relatively less improvement. Besides, DA_1, DA_2, and DA_3 have nearly the same evaluation metrics, whereas significant degradation in DA_4, which simultaneously updates multilayer soil moisture, is observed at the 20-cm evaluations.

For all subregions, generally, OL tends to overestimate soil moisture with positive biases at all subregions and at almost all the evaluation depths (Fig. 6), which is consistent with previous findings of Du et al. (2014) and Cai et al. (2014). Nevertheless, the overestimation is ameliorated through DA (from DA_1 through DA_3) and reduced RMSE values are obtained (Fig. 7) at all subregions except for SNOTEL_R02. Despite this, the magnitudes of these improvements (relative to OL) are limited, except for CTP_SMTMN and SCAN_R04, where noticeable improvements are obtained. This phenomenon might be due to the existence of systematic biases in CLM4, which are spatially variant and cannot be fully covered by the limited spread in the 40 CLM4 members. As an example, Fig. 8 shows
the temporal evolution of near-surface soil moisture ensemble spread (represented by the standard deviation of all ensemble members) of case DA_1 for the globe and the other 12 subregions. On average, these spreads are only half of the RMSE values in Fig. 7, making it rather difficult to effectively assimilate satellite observations and eventually lead to marginal improvements in soil moisture estimation. A possible solution is to apply a spatiotemporally varying inflation to the state space in DART (Anderson et al. 2009) or to perturb the prognostic variables that represent errors in both model physics and parameters (Reichle et al. 2007, 2009; Draper et al. 2012; De Lannoy and Reichle 2016). An alternative way, as summarized in Kumar et al. (2012), is to eliminate the model bias prior to data assimilation by scaling observation to the model’s climatology (Reichle and Koster 2004; Draper et al. 2012) or by calibrating time-invariant model parameters prior to DA (Yang et al. 2007). The latter approach is adopted in this study.

2) THE ROLE OF CALIBRATED MICROWAVE PARAMETERS

Similar to evaluation results in subregion SCAN_R01, the performance of DA_0 varies with location (see Figs. 5–7). This is reasonable because the precalibrated parameters are highly variable in space (Fig. 2), which causes DA_0 to perform either abnormally in regions where default parameters deviate far from the calibrated ones or well in regions where default parameters are close to the calibrated ones. Note that the precalibration scheme presented in appendix B is a highly nonlinear system. Therefore, the calibrated parameters may not be unique or optimal to Eq. (B2) within limited iterations, which further explains why sometimes DA_0 even performs slightly better than other DA cases (e.g., 5 cm at CTP_SMTMN, 20 cm at SCAN_R01, and 5 and 10 cm at SCAN_R04). Nevertheless, it is rather dangerous to simply use the default RTM parameters, as they may introduce great
uncertainties at locations that are impossible to be distinguished in advance. Specifically, Fig. 9 shows the comparison between CLM4–RTM-simulated (without data assimilation) and satellite-observed TBs. It is seen that the RMSE values (between simulation and observation) are reduced using precalibrated [optimized (opt)] than using the default (def) RTM parameters in most subregions except for CTP_SMTMN, ICN, SCAN_R04,

FIG. 8. Temporal evolution of near-surface soil moisture ensemble spread for case DA_1.

FIG. 7. As in Fig. 5, but for RMSE.
FIG. 9. Scatterplots between CLM4–RTM-simulated and satellite-observed TBs from June to September 2010. The "def" and "opt" denote simulation with the default and calibrated RTM parameters, respectively.
and SMOSMANIA. Nevertheless, further evaluations (also see section 3b) suggest that differences between simulated and observed TBs are further reduced through the assimilation of AMSR-E TB (figure not shown here for brevity). In general, the DA with pre-calibrated RTM parameters is recommended as it shows a more robust performance.

3) THE CREDIBILITY OF SIMULTANEOUSLY UPDATING MULTILAYER SOIL MOISTURE

The degradations in DA_4, which is typically observed at subregions CTP_SMTMN, ICN, SCAN_R04, SMOSMANIA, and SNOTEL_R02, might be due to two reasons. First, it is commonly recognized that there is a time delay in deeper soil moisture with responding to changes in near-surface conditions induced by evaporation and precipitation [see Fig. 2 in Changnon (1987)]. Since TB is sensitive to changes in topsoil moisture (0–5 cm), it is reasonable to update near-surface soil moisture by assimilating observed satellite TB. However, deeper soil moisture might be problematically correlated to system-estimated TB and result in unrealistic increments. This is different from the findings of Kumar et al. (2009), where soil moisture at deeper layers is also effectively updated based on the error correlation between surface and root-zone soil moisture values. Second, unlike topsoil whose ensemble spread is maintained by the input CAM4.0 ensemble forcing, ensemble spread in deeper soil moisture may collapse or damp whenever DA is performed. This will make the prior deep-soil moisture too confident to be affected by external observations and will therefore result in weak increments in deep-soil moisture as well. Again, perturbations to prognostic variables including soil moisture at deeper layers are expected to maintain the ensemble spread (Reichle et al. 2007; Kumar et al. 2012). Despite this, compared with the OL run, cases from DA_1 through DA_3, which only update the topsoil moisture, tend to estimate 10- and 0–30-cm soil moisture well, indicating that topsoil increments can propagate into soil moisture at deep layers.

4) COMPARISON OF FOUR DA SCHEMES

The performance of DA cases varies with locations and soil depths. To quantitatively evaluate their overall performance, we have used $R_{\text{diff}}$ value, which is defined as the absolute difference between OL- and DA-related correlation coefficients $[R(\text{DA}) - R(\text{OL})]$, as a surrogate to represent DA performance over OL. The higher the $R_{\text{diff}}$ value, the better the DA performs. We then averaged all the $R_{\text{diff}}$ values and calculated their statistics, as shown in Fig. 10. Statistically, DA_4 yields the lowest $R_{\text{diff}}$ and performs the worst among all five DA cases in estimating multilayer soil moisture. DA_0 is much better but still less sound than DA_1–DA_3, which show comparable performances. Given that DA_1 has the simplest CLM4 states updating scheme and only assimilating half of the observations, which significantly reduced the computational time as compared to DA_2 and DA_3.
(Table 3), it is adopted in the current DA system to produce global soil moisture products. Figure 11 further compares DA_1 and OL in terms of RMSE values. In general, improvements in DA_1 are obtained during the 2010 evaluation for most subregions that represent a wide range of soil moisture conditions relative to the multiyear climatology (Fig. 4).

b. DART prior and posterior

In DART, ensembles of TB calculated with input of CLM4 predicted and innovated land states are called “prior” and “posterior,” respectively. Figure 12 shows the evolution of RMSE between DA_1 derived and the observed TB over the globe and 12 subregions, where RMSE is calculated based on all CLM4 grid cells within this area. The credibility of this newly established DA system is reflected in the reduced deviations between the observed and model-simulated microwave TB. Clearly, RMSE values of posterior are constantly lower than that of prior, and the decrease in temporal mean of RMSE time series is obtained at all subregions, indicating the EAKF in DART behaves well in the system. In general, prior RMSE values are reduced by $1.4 \text{ K}$ on average for the 12 subregions ($-2.2 \text{ K}$ over the globe), with updates evident in the top two-layer soil moisture. Both this indirect comparison of prior and posterior TB and previous direct evaluation against in situ soil moisture observations suggest that this CLM4–RTM–DART
system holds promise for taking advantage of satellite microwave observations and the multilayer soil moisture estimation at the global scale. Despite this, the model bias in CLM4 is clearly observed in the system-estimated prior TB whose RMSE values are in the range of 5–10 K in most subregions. This again necessitates the calibration of CLM4 parameters, which is expected to further reduce the systematic bias and model uncertainties.
c. Time shifting of CAM4.0 reanalysis and forcing bias

As mentioned before, the CAM4.0 reanalysis is shifted from UTC to LST to facilitate the global assimilation of AMSR-E TB at a daily frequency. The time shifting of CAM4.0 reanalysis is based on integer hours, which may result in longitudinal artificial boundaries or stripes in the original forcing. To investigate this issue, we have randomly selected the seventeenth ensemble from the 40-member DA outputs and plotted the daily variation of reassembled ground temperature $T_g$ and the third layer soil moisture ($\text{sm}$; centered at around 6.2 cm) for 2 July 2010. A snapshot of the reassembled $T_g$ and sm estimates at 0000 UTC is shown in Fig. 13. After shifting back to the global UTC time, slight boundaries are observed in the $T_g$ map while rarely found in sm. In fact, since CLM4 was run in the offline mode, there were no horizontal communications among grid cells, and therefore, the time shifting of CAM4.0 reanalysis-induced time lags between longitudinally adjacent grid cells is expected to have negligible influences on the model outputs.

Although the CAM4.0 reanalysis has incorporated information from millions of atmospheric observations, systematic biases may exist, as commonly seen in most climate models, which make it necessary to investigate the CAM4.0 reanalysis bias in addition to its coarse spatio-temporal resolution. In the recent work of Zhang (2015), biases in CAM4.0 reanalysis are investigated by scaling precipitation and downward radiation to data from the Global Precipitation Climatology Project (GPCP) and the Clouds and the Earth’s Radiant Energy System (CERES), respectively. The “bias corrected” CAM4.0 reanalysis, while retaining the original ensemble spread, is then used to drive the offline CLM4 for snow estimation, and...
eventually shows mixed improvements and degradations in terms of snow depth and snow cover fraction in different regions. The findings indicate that the uncertainties in CAM4.0 are spatially variable and that bias correction in forcing is needed. Besides, DA improvements over OL found in this study suggest that land DA is able to compensate the impact of biased forcing in land model predictions when constrained by satellite observations.

d. The AMSR-E upscaling

This section investigates potential influences introduced through the AMSR-E TB upscaling that ignores the subgrid heterogeneity (Lakshmi et al. 1997). Yang et al. (2009) suggested that almost identical areal mean soil moisture values could be obtained by either assimilating subgrid-averaged TB or first assimilating TB in each subgrid and then averaging soil moisture. We have further investigated this issue from the perspective of the representativeness of upscaled TB. Assuming a CLM4-comparable grid cell of 1.25° × 1.25° that contains 25 AMSR-E subgrids, and each of these subgrids has specified states \( (T_g, SM, etc.) \) and different soil texture, roughness, and vegetation parameters. Basically, there are two schemes to estimate TB over the large grid cell (TB'). In scheme 1, TB in each of the 25 subgrids was first estimated individually through RTM and then spatially averaged to obtain TB'. In scheme 2, the states or parameters over all subgrids were first averaged and then TB' was obtained by running RTM once. The hypothesis is that if the upscaling is properly done, TB' estimated from the two schemes should be comparable.

A random sampling experiment was conducted toward the estimation of 6.9-GHz vertically polarized TB to verify the above hypothesis. As shown in Table 7, first, a total of 2000 subsets of gridcell states and parameters was generated through random sampling based on the “uniform” distribution with the input of land state and parameter’s physical range (min and max). Second, each subset was referred to as the spatial mean of this grid cell, and a standard deviation (std dev) was introduced to each land state variable or parameter to represent the subgrid heterogeneity. Third, 25 samples of each state variable or parameter, which were created through random sampling based on the normal distribution with the input of mean and standard deviation values, were assigned to 25 subgrids afterward. Finally, 2000 pairs of TBs were calculated using the aforementioned two schemes and compared in a scatterplot (Fig. 14). Statistically, the two schemes give approximately equivalent estimations of TB' with an overall coefficient of determination of 0.992 and root-mean-square difference (RMSD) of 1.81 K. This RMSD value of 6.9-GHz vertically polarized TB roughly corresponds to changes of 0.01 m³ m⁻² in topsoil moisture, which is fairly small compared to the standard deviation of soil moisture (0.10 m³ m⁻²) used during the random sampling experiment. This indicates the upscaled AMSR-E TB is capable of representing average condition of the corresponding grid cell, and thus is acceptable in microwave land data assimilation.

4. Summary and outlook

Global soil moisture data are urgently needed for a variety of research and application purposes while traditional soil moisture–retrieving methods are subject to various
limitations on spatiotemporal resolution and coverage or data quality. In this study, CLM4 was coupled into DART with EAKF to estimate soil moisture by assimilating AMSR-E microwave TB observations. Some methods were designed to balance the computational load and credibility of system output. Specifically, the original satellite data were upscaled to a coarser spatial resolution to reduce the total number of effective observations. An ensemble of CAM4.0 reanalysis was used to perturb CLM4 states and the original forcing time series were shifted to facilitate a global-scale daily DA. An independent precalibration of microwave parameters was conducted to minimize uncertainties in the RTM. Based on this newly developed CLM4–RTM–DART system, several experimental cases were designed to quantify the impacts of different assimilation and updating schemes. Although AMSR-E observations are only sensitive to changes in near-surface soil states, the capability of this CLM4–RTM–DART system in estimating multilayer soil moisture was also investigated.

In situ observations collected from soil moisture networks over the globe were used to evaluate DA outputs. Results suggest that updating the top 5-cm soil moisture only can potentially improve multilayer soil moisture estimation for all subregions, whereas simultaneously updating deep-soil moisture may introduce extra biases into the system and result in unrealistic estimations. In addition, no big differences are seen when soil temperature and vegetation temperature are updated together with soil moisture. In contrast to default microwave parameters, DA cases using precalibrated radiative transfer parameters show robust performances. Furthermore, differences between the model-estimated and satellite-observed TB values are reduced with the updated CLM4 states (soil moisture, etc.). At the present stage, the DA scheme utilizing precalibrated microwave parameters, assimilating a lower frequency of spatially upscaled AMSR-E TB, and updating the top two layers’ soil moisture is demonstrated to be the most efficient among all DA cases and is thus adopted in the current CLM4–RTM–DART system.

Note that these findings are drawn based on evaluation from limited locations, and a more comprehensive evaluation is expected when more in situ data are available. Evaluations suggest that potential systematic biases exist in CLM4 that cannot be fully represented by CAM4.0 reanalysis. Possible solutions include introducing an inflation to the CLM4 state space, calibrating CLM4 model parameters, or rescaling satellite observations. Note that soil temperature bias, which is a first-order error source in the radiative transfer model, is not addressed in this paper, although efforts have been made in compensating the time-mismatching issue and in calculating TB with the closest land states in time. Nevertheless, current findings indicate the newly established CLM4–RTM–DART system holds promise for producing improved global soil moisture data. Constant refinements will be implemented to further improve the system performance and applicability. Future studies will focus on calibrating the land model prior to data assimilation and systematically quantifying biases in soil temperature and other state variables.

Acknowledgments. We thank the Jackson School of Geosciences, The University of Texas at Austin; the NSFC (Grant 91337217); and the Texas Advanced Computing Center. Kun Yang is thanked for discussing parameter calibration related to the soil microwave radiative transfer model. The CAM4.0 reanalysis data are prepared by Kevin Raeder (raeder@ucar.edu).

APPENDIX A

The Radiative Transfer Model

By omitting atmospheric impacts, microwave TB observed at the top of the atmosphere can be calculated as

\[ \text{TB}_p = T_g(1 - \Gamma_p) e^{-\tau_c} + T_v(1 - \omega)(1 - e^{-\tau_c}) \times (1 + \Gamma_p e^{-\tau_c}) \quad \text{and} \]

\[ \Gamma_p = (1 - Q)R_p + QR_q e^{-H}, \]

where \( p \) and \( q \) denote microwave polarization (horizontal or vertical); \( \Gamma_p \) is soil microwave reflectivity that calculated through a semiempirical “\( Q/h \)” model developed by Wang and Choudhury (1981); \( T_g \) and \( T_v \) are near-surface soil temperature and vegetation temperature, respectively; \( \tau_c \) is vegetation optical thickness; and \( \omega \) refers to the single-scattering albedo of vegetation. Some microwave frequency-dependent parameters are given by

\[ H = \text{hm}v(x) \text{cos} \theta, \]

\[ Q = 0.35[1 - e^{-(\text{hm}v(x))^{0.391}}], \]

\[ \tau_c = \frac{\text{hm}v(100\lambda) \text{ym}v \omega_c}{\text{cos} \theta}, \quad \text{and} \]

\[ \omega = \frac{0.00083}{\lambda}, \]

where \( \theta \) is the incidence angle of microwave sensor (55° for AMSR-E), \( \lambda \) and \( k \) are the wavelength and
wavenumber \( (k = 2\pi/\lambda) \) for a specific microwave frequency (Freq). Equations (A4) and (A6) are fitted from limited field measurements provided by Fujii (2005). Equation (A3) follows Wegmuller and Matzler (1999), and Eq. (A5) follows Jackson and Schmugge (1991). Parameters \( q_{mv} \) (0 \( \leq \) 1) and \( h_{mv} \) (0 \( \leq \) 1) are surface roughness related, while parameters \( b_{mv} \) (0 \( \leq \) 5) and \( x_{mv} \) (2.1 \( \leq \) 2.0) are vegetation related. The vegetation water content \( \omega_{v} \) can be approximately related to LAI (Paloscia and Pampaloni 1988) and estimated by

\[
\omega_{v} = e^{LAI^{0.3}} - 1. \tag{A7}
\]

In Eq. (A2), parameter \( R \) that represents the Fresnel power reflectivity of a smooth soil surface is given by

\[
R_{h} = \frac{\cos \theta - \sqrt{\varepsilon_{r} - \sin^{2} \theta}}{\cos \theta + \sqrt{\varepsilon_{r} - \sin^{2} \theta}}, \tag{A8}
\]

\[
R_{v} = \frac{\varepsilon_{r} \cos \theta - \sqrt{\varepsilon_{r} - \sin^{2} \theta}}{\varepsilon_{r} \cos \theta + \sqrt{\varepsilon_{r} - \sin^{2} \theta}}, \tag{A9}
\]

\[
\varepsilon_{r} = [1 + (1 - w_{s}) (\varepsilon_{s}^{a} - 1)] + SM^{b} \varepsilon_{lw}^{a} - SM^{1/a}, \tag{A10}
\]

and

\[
\beta = 1.09 - 0.0011 \times \%\text{sand} + 0.0018 \times \%\text{clay}, \tag{A11}
\]

where \( h \) and \( v \) denote the horizontal and vertical polarization, respectively; \( \varepsilon_{r} \) is the soil dielectric constant; \( w_{s} \) is soil porosity; \( SM \) is soil moisture; \( \varepsilon_{s} = 4.7 + 0.0i \) is
the dielectric constant of a dry soil; $\varepsilon_{\text{fw}}$ is the dielectric constant of free water; $\alpha = 0.65$; and $\beta$ is a soil-texture-dependent coefficient. Equation (A10) follows Dobson et al. (1985), and Eq. (A11) follows Ulaby et al. (1986).

The above radiative transfer schemes have been used by Yang et al. (2007) and Qin et al. (2009) to estimate microwave TB in a land DA system for regional soil moisture estimation.

**APPENDIX B**

**Calibration of Radiative Transfer Parameters**

Figure B1 shows the procedure for precalibration of radiative transfer parameters for the CLM4–RTM–DART system. Note that performing precalibration should be viewed as an integral part of the DA system that calibrates model parameters by assimilating AMSR-E TB observations as demonstrated by Yang et al. (2007) and Zhao et al. (2013). These studies have shown the scheme is capable of reducing uncertainties in land surface and radiative transfer models.

As a proof of concept in this study, we use the Simple Biosphere Model, version 2 (SiB2; Sellers et al. 1996), to estimate microwave TB in a land DA system for regional soil moisture estimation.

A cost function is defined as the sum of the difference between estimated ($TB_{\text{Best}}$) and observed ($TB_{\text{obs}}$) TBs, which is also called the observation error term. To account for the background error term due to the biased initial soil moisture, a soil wetness index (SWI) is introduced and defined as the ratio of a higher (Freq1) and lower (Freq2) microwave frequencies of TB (Qin et al. 2009):

$$SWI = \frac{TB_{\text{Freq1}}}{TB_{\text{Freq2}}},$$  

(B1)

where $v$ denotes the vertical polarization. Theoretically, a higher SWI value corresponds to, approximately, a wetter near-surface soil condition. Accordingly, the cost function is defined as

$$F = \frac{1}{2} \sum_{t_0}^t [(TB_{\text{est},v} - TB_{\text{obs},v})^2 + (TB_{\text{est},v} - TB_{\text{obs},v})^2],$$  

(B2)

where est and obs denote the estimated and observed value, respectively. The shuffled complex evolution (SCE; Duan et al. 1993) algorithm is used to minimize the cost function by adjusting both SiB2 and RTM parameters. Note that the observation error term is calculated for the entire calibration period, while the background error term is considered at each observing time by adjusting the near-surface soil moisture such that the recalculated SWI is close to $(SWI_{\text{est}} + SWI_{\text{obs}})/2$.

**REFERENCES**


