Retrieving Accurate Soil Moisture over the Tibetan Plateau Using Multisource Remote Sensing Data Assimilation with Simultaneous State and Parameter Estimations

WEIJING CHEN,a CHUNLIN HUANG,b ZONG-LIANG YANG,a AND YING ZHANGb

a Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas
b Key Laboratory of Remote Sensing of Gansu Province, Helixe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, Gansu, China

(Manuscript received 9 December 2020, in final form 3 August 2021)

ABSTRACT: Data assimilation provides a practical way to improve the accuracy of soil moisture simulation by integrating a land surface model and satellite data. This study establishes a multisource remote sensing data assimilation framework by incorporating a simultaneous state and parameter estimation method to acquire an accurate estimation of the soil moisture over the Tibetan Plateau. The brightness temperature of the Advanced Microwave Scanning Radiometer 2 (AMSR2) is directly assimilated into the coupled system of the Common Land Model (CoLM) and a microwave radiative transfer model (RTM) to improve the soil moisture simulation. The Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature product and the Beijing Normal University (BNU) leaf area index product are employed to not only improve the estimation of temperature and vegetation variables from the CoLM, but they also provide more accurate background information for the RTM during the brightness temperature assimilation. In situ measurements from the Naqu network are used to evaluate the results. The model simulation showed an obvious underestimation of soil moisture and overestimation of soil temperature, which was alleviated by the assimilation experiments, particularly in the shallow soil layers. The estimated parameters also showed advantages in the soil moisture simulation when compared with the default parameters. The assimilation experiment presents promising results in the combination of model and multisource remote sensing data for estimating soil moisture over the complex mountainous region in Tibet.

SIGNIFICANCE STATEMENT: We wanted to obtain accurate and spatiotemporal continuous soil moisture over the Tibetan Plateau. Therefore, we established a multisource remote sensing data assimilation framework in which AMSR2 brightness temperature and MODIS land surface temperature were assimilated into a land surface model by a simultaneous state and parameter estimation method. The results showed the soil moisture we obtained has better accuracy than both model simulation and satellite retrieval. Compared with in situ measurements, the model simulations underestimated soil moisture and overestimated soil temperature. With assimilation, the reduction of the mean bias error (MBE) (RMSE) for shallow-layer soil moisture at different sites ranges from 20% (19%) up to 98% (71%).

KEYWORDS: Soil moisture; Remote sensing; Data assimilation

1. Introduction

As an important component of the terrestrial hydrological cycle, soil moisture impacts not only the water cycle but also the energy and carbon cycles (Robinson et al. 2008; Seneviratne et al. 2010). According to the Fifth Assessment Report of the International Panel on Climate Change (IPCC), the variability in soil moisture has been shown to be closely related to climate change over the past few decades (Pachauri et al. 2014). Ground stations are effective way to obtain soil moisture measurements, but spaceborne microwave remote sensors have provided the most practical means of monitoring the spatial variability of soil moisture from regional to global scales. Particularly, microwave is useful for soil moisture detection because of the distinct difference between the dielectric constant of water and dry soil (Ulaby et al. 1986; Wigneron et al. 2003)

L-band, C-band, and X-band channels are widely used in microwave satellite sensors, such as the Scanning Multichannel Microwave Radiometer (SMMR), the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E; Njoku et al. 2003), the Soil Moisture and Ocean Salinity (SMOS; Kerr et al. 2001) and the Soil Moisture Active Passive (SMAP; Entekhabi et al. 2010). These sensors have been quite successful in providing global soil moisture products for a wide range of applications: hydrological/land surface modeling, land–atmosphere coupling, energy balance modeling, agricultural drought monitoring, and ecohydrology (Bisselink et al. 2011; Dorigno et al. 2017; Ford et al. 2014; Martínez-Fernández et al. 2016; Ray and Jacobs 2007; Wanders et al. 2014; Zhang and Jia 2013).

Moreover, integrating microwave observations with a land surface model by data assimilation has been widely used to acquire more accurate soil moisture. The performance of assimilating the retrieved soil moisture products from microwave sensors has varied in different regions according to previous studies (Crow and Ryu 2009; Chen et al. 2014; Draper et al. 2012; Kumar et al. 2015; Lievens et al. 2016; Tian et al. 2017). This may be partially ascribed to the accuracy of the retrieved soil moisture, which depends on the applicability of the inversion method and the uncertainty of the auxiliary information required in the inversion process. De Lannoy and Reichle (2016a) pointed out that the inversion of soil moisture may be
inconsistent with the model simulation because there is a certain difference in the land surface parameters and the auxiliary information, such as vegetation and temperature, used in the inversion process and the simulation/assimilation experiment (Wigneron et al. 2017). Furthermore, some studies have evaluated the retrieved soil moisture products from several satellites with in situ measurements from the world-distributed soil moisture networks and found that there were distinct divergences between the retrievals and in situ data (Al Bitar et al. 2012; Chen et al. 2013; Dorigo et al. 2015; Gherboudj et al. 2012; Su et al. 2013).

An alternative approach is to directly assimilate the brightness temperature or backscatter coefficient to avoid the errors from the inversion process. To this end, a forward model is necessary to establish the connection between the soil moisture simulation from a land surface model and the brightness temperature or backscattering coefficient observations. The less intractable forward model for passive microwave has broadened the application of the direct assimilation of brightness temperature. The impact of assimilating brightness temperature from single sensor on soil moisture estimation has been studied in different regions (De Lannoy and Reichle 2016a,b; Lievens et al. 2015; Yang et al. 2016; Crow et al. 2020). In addition, more studies have focus on assimilating multisource satellite observations to combine the potential advantages of different data (Lievens et al. 2017; Zhao and Yang 2018; Girotto et al. 2019), such as the assimilation of brightness temperature from different sensors, the assimilation of brightness temperature and data from optical sensors, etc. Soil temperature is an important factor affecting both simulated and observed brightness temperature (Raju et al. 1995; Zheng et al. 2012). Thus, in the construction of multisource remote sensing data assimilation framework, it is expected to improve the assimilation performance by considering the estimation of soil temperature.

In data assimilation, model parameters may be imperfectly defined and can affect the accuracy of the model simulations and the performance of assimilation. Calibration of parameters has been proved to produce a better model simulation and widely used in hydrological model (Wanders et al. 2014; Rajib et al. 2016; López López et al. 2017). But the simultaneous estimation of states and parameters by data assimilation can ease the demand for long-term historical data in the calibration process. The parameters can be updated as state variables along with the assimilation of observations using the augmented state vector or dual filter that has been proved to be effective in reducing the simulation deviation and improving the assimilation performance (Bateni and Entekhabi 2012; Yan and Moradkhani 2016; De Lannoy et al. 2013). This study aimed to establish a multisource remote sensing data assimilation framework by incorporating a simultaneous state and parameter estimation method to improve the performance of direct assimilation of brightness temperature on soil moisture estimation.

As a substitute sensor of AMSR-E, the Advanced Microwave Scanning Radiometer 2 (AMSR2) has continued to provide multichannel passive microwave-based data since May 2012 (Imaoka et al. 2010). It is noteworthy that a C-band frequency (7.3 GHz) channel has been added to the AMSR2 to mitigate radio-frequency interference (RFI) effects that the original C-band frequency (6.9 GHz) experienced in certain regions (Njoku et al. 2005). Chen et al. (2017) has evaluated the soil moisture retrievals from several sensors using measurements from two monitoring networks on the Tibetan Plateau. The results indicated that the AMSR2 products evidently overestimated or underestimated the temporal variation of soil moisture in different regions.

To further investigate the applicability of brightness temperature from the AMSR2 in soil moisture estimation and verify our proposed assimilation framework mentioned above, this study conducted an assimilation experiment in the Qinghai–Tibet Plateau region of China and was motivated to address the following questions: 1) Since the AMSR2 soil moisture retrievals have been demonstrated to have a challenging quality over the region, can it be possible to estimate the parameters and improve the soil moisture simulation by assimilating brightness temperature on the Tibetan Plateau? Moreover, whether there is a remarkable discrepancy in assimilating the 6.9-GHz brightness temperature and the newly added 7.3-GHz brightness temperature? 2) Can the assimilation of land surface temperature provide accurate soil temperature information at shallow layers for brightness temperature simulations? 3) Whether the estimated parameters derived from the AMSR2 brightness temperature assimilation, especially the soil parameters, are effective to improve the soil moisture simulation of the land surface model?

The paper is organized as follows: Section 2 provides the introduction and preprocessing of the data used in this study. Section 3 describes the models and methods, as well as the detailed experimental design. The results and discussions about the experiment results are presented in section 4. Finally, conclusions are reported in section 5.

2. Data

a. Study area and in situ measurements network

The Tibetan Plateau is the world’s highest plateau with a total area of approximately 2.5 million square kilometers, and an average altitude of over 4000 m. It plays an important role in the formation and development of the Asian monsoon climate and inland aridity (Zhang et al. 2015). The Tibetan Plateau splits the midlatitude westerlies and blocks the south Asian monsoon. In addition, the Tibetan Plateau’s elevated heating drives the Asian monsoon system. Several observation stations have been constructed to monitor soil moisture and temperature in the Tibetan Plateau, such as the Naqu network in the central Tibetan Plateau (Yang et al. 2013), the Maqu network in the eastern Tibetan Plateau (Su et al. 2011), and the Pali network in the southern Tibetan Plateau. Thus, the goal of this study is to acquire accurate soil moisture information over the Tibetan Plateau using the brightness temperature assimilation, with the study area within 26°–40°N and 74°–104°E (Fig. 1). The Naqu network was chosen for the validation of the land
surface model and data assimilation systems with the consideration of the time coverage of in situ data.

The Naqu network is located around the town of Naqu over an area of \(100 \text{ km} \times 100 \text{ km}\) (31°–32°N and 91°30′–92°30′E; Yang et al. 2013). A total of 57 stations were set up successively during 2010–12 to collect soil moisture and temperature data (Fig. 1). Soil moisture and temperature were measured at four depths (0–5, 10, 20, and 40 cm), and provided at 30 min and daily resolution. Direct measurements of soil moisture are calibrated with the measured dielectric by taking account of the influences of soil organic carbon. Moreover, some auxiliary parameters of this network, such as soil sand content, soil clay, and soil organic carbon content, are measured at each station. In situ hourly soil moisture and temperature are used to compare with the model simulation. To match the spatial resolution of the model (which is consistent with the spatial resolution of the input forcing data), the original in situ data were aggregated by arithmetically averaging at 0.1°. As a result, 29 different pixels with in situ measurements were obtained in the Naqu network.

b. Satellite data

Brightness temperature (TB) assimilated in this study are sourced from the level 3 product of AMSR2 (JAXA 2016). This dataset is distributed by the Japan Aerospace Exploration Agency (JAXA) with equidistant cylindrical projection, available at a 0.1° resolution and daily interval (JAXA 2013). The AMSR2 is on-board the first generation of the Global Change Observation Mission–Water (GCOM-W) satellite, which was launched in May 2012 and continues the Aqua/AMSR-E mission to understand the mechanisms of climate and water cycle variations. The AMSR2 has a sun-synchronous orbit crossing the equator at 0130/1300 local time with the incidence angle of 55°. The frequency channel set of the AMSR2 consists of seven bands (6.9, 7.3, 10.7, 18.7, 23.8, 36.5, and 89 GHz) with both horizontal and vertical polarizations. Given the high sensitivity of the low frequencies to soil moisture, the brightness temperature of 6.9 and 7.3 GHz are assimilated and compared in this study. Moreover, according to the study of Kim et al. (2015), the AMSR2 level 3 soil moisture product developed by the Vrije Universiteit (VU) Amsterdam and the National Aeronautics and Space Administration (NASA) is also employed to examine the performance of model simulations and assimilation experiments.

Land surface temperature (LST) retrieved from the Moderate Resolution Imaging Spectroradiometers (MODIS) on the Terra and Aqua platforms are chosen as the observations for soil temperature assimilation considering the reliable accuracy (Wan 2014). The MOD11C1 (Terra) and MYD11C1 (Aqua), version 6, products have provided daily LST values in a 0.05° latitude/longitude climate modeling grid (CMG) since July 2002 (Wan 2013). A pair of daytime and nighttime observations from both Terra and Aqua enrich the LST observations up to four times a day. Only those high quality LST retrievals that flagged with zero are used in this study. Data from every four CMGs are merged into a model grid cell using the arithmetical average before being assimilated into the land surface model.

The Beijing Normal University leaf area index (BNU LAI) is used to substitute for the default LAI values in the land surface model to include more precise vegetation information. This LAI dataset was generated by reprocessing the MODIS LAI products with a two-step integrated method, which was also validated by true LAI values collected over 26 in situ sites (Yuan et al. 2011). It contains the value of 8-day composites at global scale with two resolutions (1-km sinusoidal grids 30 arc s−1 geographic coordinate grids) from 2000 to 2016. Since the spatial resolution of the original LAI dataset is much higher than the land surface model, the LAI values used for each

---

**FIG. 1.** Land-cover-type map of the study area and the location of ground stations in the Naqu network. The land-cover-type dataset was obtained from the CoLM.
model grid cell were the area-weighted average of all the corresponding land-cover tiles within this model grid cell.

c. *Forcing data*

In this study, the China Meteorological Forcing Dataset (CMFD) is used as the input forcing data to drive the land surface model. This dataset includes seven near-surface meteorological forcing fields (air temperature, surface pressure, specific humidity, wind speed, downward shortwave radiation, downward longwave radiation, and precipitation rate) that are processed at a 3-h temporal resolution and a 0.1° spatial resolution (Yang and He 2018). It provides a high spatial–temporal resolution gridded data covering the geographic scope of China since January 1979. According to Chen et al. (2011), the CMFD is derived from the fusion of remote sensing products, reanalysis dataset and in situ observation data at weather stations. He et al. (2020) has comprehensively described and evaluated the CMFD dataset. In this study, the model time step was set to be hourly, thus the original dataset was disaggregated to hourly using conserving interpolation for precipitation and spline interpolation for other variables.

3. Models and methods

a. *Land surface model*

The Common Land Model (CoLM) is a state of the art land surface model developed by Dai et al. (2003), and its 2014 version is used in this study. In the CoLM, the water and energy balance calculations are performed independently based on the land-cover type. Every model grid cell can be subdivided into five tiles at most (each tile contains a single land-cover type): soil, urban and built up, wetland, land ice, and land water bodies. Each tile maintains its own prognostic variables at every time step, and then the mean condition of each grid cell responds to the area-weighted estimation over the different land-cover fractions (Dai et al. 2001).

Water and energy balance calculations involve 1 vegetation layer, 10 unevenly spaced vertical soil layers, and up to 5 snow layers (depending on the total snow depth). The thicknesses of 10 soil layers are 0.0175, 0.0276, 0.0455, 0.0750, 0.1235, 0.2039, 0.3359, 0.5539, 0.9133, and 1.1370m, respectively. The global surface datasets in the CoLM used to make surface characteristic data for the study area are provided by the model developer. For example, land-cover dataset is a reprocessing data based on the MODIS products, and the soil texture dataset is from a comprehensive, multilayer soil information dataset (Shangguan et al. 2014). More details can refer to the website of the CoLM.

b. *Radiative transfer model*

A radiative transfer model (RTM) is coupled with the CoLM in this study to conduct the forward brightness temperature simulations. The RTM is used as the observation operator to convert the soil moisture and soil temperature from the CoLM to the brightness temperature at the surface. Therefore, a zero-order tau–omega (τ–ω) model (neglecting multiple scattering effects) is selected to account for the vegetation contribution on microwave signal, and the simulated brightness temperature can be expressed as follows (Mo et al. 1982):

\[
TB_p = T_g(1 - \Gamma_p) \tau_p + T_c(1 - \omega)(1 - r_p)(1 + \Gamma_p \times r_p), \quad (1)
\]

\[
r_p = \exp(-\tau_p/\cos\theta), \quad (2)
\]

where \( p \) stands for the vertical/horizontal polarization. The terms \( T_g \) and \( T_c \) are the near surface soil temperature and canopy temperature, respectively, \( \Gamma_p \) is the rough surface reflectivity, which is relevant to the soil conditions, \( \omega \) is the single-scattering albedo of the vegetation, and \( r_p \) is the vegetation transmissivity, which is calculated by the incidence angle \( (\theta) \) and the vegetation optical thickness \( (\tau_p) \).

The Wigneron scheme coupled in the Community Microwave Emission Modeling Platform (CMEM) is applied to calculate the vegetation-related parameters: \( \tau_p \) and \( \omega \). The scheme is suitable for frequencies under 10 GHz according to CMEM (Wigneron et al. 2004):

\[
\tau_p(\theta) = \tau_{\text{NAD}}(\sin^2\theta \times \tan^2 \theta + \cos^2 \theta), \quad \text{and} \quad (3)
\]

\[
\tau_{\text{NAD}} = b_1 \times \text{LAI} + b_2, \quad (4)
\]

where \( \tau_{\text{NAD}} \) is the nadir value of the vegetation optical depth. Equation (4) shows a linear dependence of \( \tau_{\text{NAD}} \) on the LAI with two parameters \((b_1\) and \(b_2)\). The term \( \tan \theta \) is the polarization correction parameter. The terms \( b_1, b_2, \text{and} \tau_p \) are assumed to be empirical values in the CMEM, which are primarily dependent on the vegetation type (Wigneron et al. 2007).

The rough surface reflectivity \( \Gamma_p \) can be obtained through the semiempirical \( Q-H \) model, which is developed by Wang and Choudhury (1981) to account for the soil roughness effects.

\[
\Gamma_p = [(1 - Q)R_p + QR_p] \times H, \quad (5)
\]

where \( R_p(\omega) \) is the reflectivity of a smooth surface, which can be calculated using the Fresnel reflection equations; \( q \) stands for the horizontal polarization when \( p \) stands for vertical polarization, or vice versa. The terms \( Q \) and \( H \) are parameters related to the soil roughness and can be given by the following formulas (Bauer and Grody 1995):

\[
Q = 0.35 \times [1 - \exp(-0.6 \times \text{RMS}^2 \lambda)], \quad \text{and} \quad (6)
\]

\[
H = \exp[-(2 \times k \times \text{RMS} \cos\theta)^2], \quad (7)
\]

where \( k \) is the wavenumber \((=2\pi\lambda)\), which is determined by the wavelength \( \lambda \) (m). Here, RMS is the standard deviation of surface heights, a geophysical parameter to characterize the feature of soil roughness.

Another indispensable part of the RTM is soil dielectric modeling that implicitly reflects the impacts of soil water on the signal of microwave from ground. The scheme of Dobson et al. (1985) is used to calculate the soil dielectric. In the end, the Fresnel reflectivity in Eq. (5) varies with the varying soil dielectric.

Specifically, the dynamic variables, such as soil moisture, soil temperature and canopy temperature, used in the RTM are the
output of the CoLM. Considering the penetration depth of C-band microwave, the first layer of soil temperature and soil moisture were used (Zhao and Yang 2018). Moreover, the soil related parameters used in the RTM are consistent with the CoLM.

c. Ensemble Kalman filter

The assimilation experiments in this study are conducted within the framework of the ensemble Kalman filter (EnKF). The EnKF propagates an ensemble of state variables with each realization mutually independent. The prior/posterior distribution characteristics of state variables are represented by the ensembles, which also fulfill the capability of the EnKF to cope with the nonlinear model operator (Burgers et al. 1998). In the EnKF, the forward evolution of each ensemble member can be expressed as

\[
x_{i,t+1} = M(x_{i,t}, \alpha_{i,t}, \beta_{i,t}),
\]

where \(M(\cdot)\) is the model operator to forecast the state variables (CoLM in this study). The term \(x_{i,t+1}\) is the \(i\)th ensemble member of forecast state variables at time \(t+1\) (model simulations), and \(x_{i,t}\) is the \(i\)th ensemble member of analysis state variables at time \(t\) (after update/assimilation). The terms \(\alpha_{i,t}\) and \(\beta_{i,t}\) represent the \(i\)th ensemble member of meteorological forcings and parameters at time \(t\) required to run the CoLM.

When the observation \(Y_{t+1}\) is available at time \(t+1\), a stochastic perturbation \(v_{t}\) of the Gaussian distribution with a zero mean is added to the original observation to generate the perturbed observation ensemble and account for the error associated with the observation. Then, the ensemble member of state variables can be updated according to the following equation:

\[
Y_{i,t+1} = Y_{t+1} + v_{t}, \quad \text{and} \quad x_{i,t+1}^{f} = x_{i,t}^{f} + K_{i,t+1}(Y_{i,t+1} - y_{i,t+1}),
\]

where \(Y_{i,t+1}\) is the \(i\)th ensemble member of perturbed observations at time \(t+1\). The term \(Y_{i,t+1}\) is the \(i\)th ensemble member of the simulated observations, which can be written as: \(y_{i,t+1} = H(x_{i,t+1})\). The term \(H(\cdot)\) is the observation operator to establish a relationship between the model states and observations (the RTM in this study). The term \(K_{i,t+1}\) is the Kalman gain matrix at time \(t+1\), which is calculated as

\[
K_{i,t+1} = P_{i,t}^{f}H^{T}(HP_{i,t}^{f}H^{T} + R)^{-1},
\]

where \(R\) is the observation error covariance matrix. The term \(P_{i,t}^{f}H^{T}\) is the cross-covariance matrix between the model forecast states ensembles and the simulated observations ensembles; \(HP_{i,t}^{f}H^{T}\) is the ensemble covariance matrix of the simulated observations.

d. Experiment design

1) Model ensemble simulation

To validate the impact of AMSR2 brightness temperature assimilation on soil moisture estimation, the simulation and assimilation experiment will focus on the period from 1 May to 31 October 2014 with an hourly time step, mainly during the summer season/wet season (the microwave signals from snow and frozen soil differ from that of thawed soil, which is beyond the scope of this study). Five-year forcing data are looped twice to total 10 years for the model spin up before the experiment start date to obtain reasonable initial conditions. For the simulation and assimilation experiments, the model runs in ensemble mode with 30 realizations.

Random perturbation fields are required for the input meteorological variables to account for the uncertainties of the forcing dataset and to generate state ensembles. Given the implied physical relationships during the variation of meteorological variables (e.g., the increase of precipitation tends to associate with the decrease of air temperature), it is more reasonable to adopt multivariable random fields. In this study, normally distributed additive perturbations were applied to air temperature and longwave radiation and lognormally distributed multiplicative perturbations were added to precipitation and shortwave radiation (Table 1). The prescribed standard deviation error of variables and the cross correlations between variables used in this study are based on the study of Reichle et al. (2007) and Kwon et al. (2019), which are widely used for land data assimilation. Moreover, the autocorrelations of the meteorological variables in time and space are ignored for simplicity in this study.

2) Parameter estimations via data assimilation

Soil moisture and brightness temperature simulations are sensitive to certain parameters. Thus, sand (used in both CoLM and RTM), clay (used in both CoLM and RTM), organic carbon (used only in CoLM) and RMS (used only in RTM) are chosen to be updated using the data assimilation method during the assimilation experiment. The parameter’s vector is composed as \(\beta = (\text{sand, clay, OC, RMS})\). The initial parameter ensembles are generated using the normally distributed additive perturbations of the model default parameter values. The upper and lower bounds of the parameters are used to confine the initial parameters ensembles and the updated parameter values, which are defined based on the model settings and literatures (Table 1). Moreover, to ensure a reasonable update of sand percentage and clay percentage, the sum of these two parameters should be further checked within 96% (defined by the CoLM). If not, the updated values should be refined as \(\%\text{sand} = \%\text{sand}[96/\text{sum}(\%\text{sand} + \%\text{clay})]\), \(\%\text{clay} = \%\text{clay}[96/\text{sum}(\%\text{sand} + \%\text{clay})]\).

3) Assimilation with satellite observations

Given the high sensitivity of low-frequency microwave signal to the soil moisture, 6.9 GHz V-polarization brightness temperatures are chosen as the observations to update 10 layers soil moisture, as well as the related parameters mentioned in section 3d(2). Two parallel filters are designed to recursively estimate soil moisture and the related parameters. In addition, more accurate estimation of soil temperature and vegetation information can contribute to reduce the simulation error of brightness temperature, as well as improve the accuracy of assimilated soil moisture. Thus, MODIS land surface
temperatures were assimilated to update 10 layers soil temperature, and the default leaf area index in the CoLM was substituted by the BNU LAI dataset. The assimilation process is simplified as follows: at each model step, all \( y_{i,t+1} \) are calculated and stored in memory first; then checking whether there are observations. If MODIS LST exits, the EnKF is used to update soil temperature. If AMSR2 TB exits, the EnKF is used to separately update soil moisture and parameters. Instant feedback is applied to state variables and parameters after assimilation, then the next step of the model begins.

A prescribed observation error is essential for the EnKF to generate the perturbed observation ensemble. However, it is very difficult to determine the observation error of satellite data at a global scale. Therefore, a number of studies have focused on testing the influence of observation error on the assimilation results (Walker and Houser 2004; Van Dijk and Renzullo 2011), but are still unable to reach a satisfactory unified conclusion (it is not the focus of this study). Therefore, by referring to previous relevant literature, a Gaussian observation error is introduced in this study for the brightness temperature and land surface temperature with standard deviation error is introduced in this study for the brightness temperature and land surface temperature with standard deviation, respectively. The mean bias error (MBE), the root-mean-square error (RMSE), the correlation coefficient (\( R \)), the normalized error reduction (NER) of MBE, RMSE, \( R \) are used as evaluation measures in this study:

\[
MBE = \frac{1}{T} \sum_{t=1}^{T} (X_t - \bar{X}_{true,t}),
\]

\[
RMSE = \left( \frac{1}{T} \sum_{t=1}^{T} (X_t - \bar{X}_{true,t})^2 \right)^{1/2},
\]

\[
R = \frac{\sum_{t=1}^{T} (X_t - \bar{X})(\bar{X}_{true,t} - \bar{X}_{true})}{\sqrt{\sum_{t=1}^{T} (X_t - \bar{X})^2} \sqrt{\sum_{t=1}^{T} (X_{true,t} - \bar{X}_{true})^2}}.
\]

\[
NER_{MBE} = 1 - \frac{|MBE_{DA}|}{|MBE_{OL}|},
\]

\[
NER_{RMSE} = 1 - \frac{RMSE_{DA}}{RMSE_{OL}},
\]

\[
NER_R = \frac{R_{DA} - R_{OL}}{R_{OL}}.
\]

where \( p_{relax} \) is an artificially set parameter controlling the ensemble inflation that is assumed to be 0.9 after several tests. The terms \( \sigma^2 \) and \( \sigma' \) are the prior and posterior ensemble standard deviation, respectively.

5) VALIDATION WITH IN SITU MEASUREMENTS

The assimilation results of soil moisture, soil temperature and parameters are represented by the ensemble mean. Because the model layers do not correspond to the in situ measurements depth directly, the average of first and second model layer is compared to in situ measurements at 0–5 cm, the fourth model layer is compared to in situ measurements at 10 cm, the fifth model layer is compared to in situ measurements at 20 cm, and the sixth model layer is compared to in situ measurements at 40 cm, respectively. The mean bias error (MBE), the root-mean-square error (RMSE), the correlation coefficient (\( R \)) and the normalized error reduction (NER) of MBE, RMSE, \( R \) are used as evaluation measures in this study.
layers, 917 and 872 measurements are missing for four soil layers in two grids, 904 measurements are missing for the first layer in one grid, and 2638 measurements for the fourth layer are missing in one grid. The terms $X_t$ and $X_{true}$ represent the ensemble mean of state variables (estimated by model simulation or assimilation experiment) and in situ measurements at time step $t$, respectively. MBE$_{OL}$ (RMSE$_{OL}$, $R_{OL}$) and MBE$_{DA}$ (RMSE$_{DA}$, $R_{DA}$) represent the statistical error index of model simulation and assimilation experiment. NER$_{MBE}$ (NER$_{RMSE}$) ranges between negative infinity and 1, and NER$_{R}$ ranges between $-1$ and positive infinity. A positive value of NER indicates an improvement of the assimilation experiment relative to the model simulation while a negative value indicates the degradation.

4. Results and discussions

4.1. Soil moisture evaluation over the study area

Figure 2 shows the comparison of the monthly averaged soil moisture at 5 cm over the study area estimated by the model simulation and the assimilation experiment. Each row corresponds to the results of a specific month. The first column is the soil moisture from the model simulation (OL). The second column is the soil moisture from assimilation experiment using the brightness temperature of 6.9 GHz (DA$_{6.9}$). The spatial mean and standard deviation value of the corresponding variables are marked above each subplot. The soil moisture simulation increases from May and then decreases after reaching its peak in August, along with a similar trend for precipitation rate in the study area. Figure 3 displays the total precipitation and the correlation between the daily precipitation and the daily soil moisture from OL during the experiment period. As indicated by the correlation, the simulated soil moisture shows a high positive correlation with precipitation, which is greater than 0.6 in most areas. Areas with a smaller correlation are accompanied by less total precipitation. Additionally, there is an evident southeast–northwest gradient of both soil moisture simulation and precipitation, which is largely attributed to the influence of the southwest monsoon from the Indian Ocean in summer. As a result of topography blocking, most of the water vapor transported by the monsoonal flow forms precipitation in the southeast and can hardly reach the northwest (Maussion et al. 2014). In addition to the high association with precipitation, it is worthy to note that the soil moisture simulation also shows some relationship with the land-cover types (cf. Figs. 1 and 2). This is primarily due to the fact that land-cover types affect the infiltration of rainfall and the movement of water flux, as well as the soil hydrothermal properties in the CoLM.

The characteristics of soil moisture patterns have been obviously changed by the assimilation experiment. First, the soil moisture shows obvious increase over the study area after assimilation, indicating by the increase of the spatial mean value ranging between 0.031 and 0.053. Second, the change of soil moisture is not consistent in different regions after assimilation. For example, from June to September, soil moisture increased evidently in the northwest region, but decreased slightly in the southeast region. This opposite trend relieves the considerable discrepancy in soil moisture between the southeast and northwest. The spatial standard deviation value of the soil moisture also confirms that the differences in soil moisture over the study area has been reduced after assimilation. Finally, the northern region is mostly drier than the southern region over the study area as shown by the soil moisture from both the OL and DA$_{6.9}$, especially during the wet season (from June to September).

b. Validation of soil moisture and temperature

1) Soil moisture evaluation over the 0.2°

The observation grid in the Naqu network within 31.6°-31.8°N and 91.7°-91.9°E (labeled as blue rectangle in Fig. 1) contains approximately 20 ground stations. Given that there is a sufficient number of ground stations within this grid, it can enhance the reliability of results validation and deal with the mismatch of spatial representativeness between in situ measurements and model/satellite grid data. The soil moisture measurements from the 20 ground stations are arithmetically averaged prior to the comparison. Figure 4 displays the time series of hourly soil moisture at four layers from model simulation (OL), assimilation experiment (DA$_{6.9}$), and the AMSR2 soil moisture retrieval (AMSR2) of this 0.2° grid during the experiment period. The statistical metrics are given in Table 2. The discrete points of the AMSR2 represent the soil moisture at the transit time. The curves of OL deviate from that of in situ measurements with a large negative bias at all four layers. The soil moisture from OL responds to the temporal variation of precipitation, but fails to capture the temporal variation of in situ measurements. Moreover, the lower temporal dynamic variability and extremely underestimated soil moisture in May and October suggested that a more reasonable parameterization for soil freeze–thaw process is expected to be integrated into the CoLM. The AMSR2 retrieval shows an inconsistent deviation from the in situ measurements with satisfactory accuracy during the dry season but excessive soil moisture during the wet season, which leads to an obvious overestimation during the experiment period. Compared with OL, the AMSR2 retrieval can correctly reflect the values of soil moisture during the dry season but cannot accurately explain the variation characteristics of soil moisture during the wet season.

DA$_{6.9}$ exhibits a good agreement in the temporal variation of soil moisture with the in situ measurements, and captures the variation in both shape and magnitude. DA$_{6.9}$ attains a satisfactory accuracy of soil moisture at 5 cm, showing more advantages than both OL and the AMSR2 retrieval. DA$_{6.9}$ reduces the MBE and RMSE from $-0.113$ and $0.119$ cm$^3$ cm$^{-3}$ (OL) to $-0.018$ and $0.039$ cm$^3$ cm$^{-3}$, but slight degradation is shown by $R$ decreasing from 0.804 (OL) to 0.796. Improvements of soil moisture at subsurface layers via assimilation can also be found in Fig. 4 and Table 2. Similar to the surface layer, the correlation shows a slight decrease at deeper layers after assimilation. An accurate estimation of soil moisture at deep layers depends on more relevant satellite observations containing the soil moisture profile.
FIG. 2. The monthly averaged soil moisture \( (\text{cm}^3 \text{ cm}^{-3}) \) at 5 cm estimated by (left) model simulation (OL); assimilation experiment using the brightness temperature of (center) 6.9 GHz (DA6.9) and (right) 7.3 GHz (DA7.3). Each row corresponds to the results of a specific month.
information. Considering the relatively larger errors of soil moisture at 40 cm for both OL and DA6.9, a more appropriate scheme for the calculation of water fluxes between soil columns in the CoLM is expected to improve soil moisture estimation at deeper layers.

2) SOIL MOISTURE AND TEMPERATURE EVALUATION OVER ALL 29 0.1° GRIDS

As mentioned in section 2.1, there are 29 model grids with in situ measurements in the Naqu network. To evaluate the assimilation performance, the MBE, RMSE, and $R$ between in situ measurements and AMSR2 soil moisture retrieval, soil moisture estimated by the model simulation (OL) and assimilation experiment (DA6.9) are calculated. The left panel of Fig. 5 illustrates the boxplots showing the MBE (top), RMSE (middle), and $R$ (bottom) of soil moisture at four layers, while the spatial distribution of the NER_MBE (top), NER_RMSE (middle), and NER_R (bottom) of soil moisture at 5 cm are plotted in the right panel of Fig. 5.

It can be seen from the boxplot of the MBE, OL underestimates the soil moisture in either shallow or deep layers at most grids, as pointed out by Yang et al. (2009). The median values of the MBE, RMSE, and $R$ of soil moisture from OL at 5 cm are −0.108, 0.118, and 0.757 cm$^3$ cm$^{-3}$, respectively. The positive MBE of the AMSR2 retrieval indicates the overestimation of soil moisture. Compared with OL, the AMSR2
retrieval has a comparative MBE and RMSE but a lower $R$. The AMSR2 retrieval presents an unsatisfactory accuracy in the Naqu network, which cannot meet the mission requirement as reported by Chen et al. (2017). As for the performance of assimilation, the 5-cm soil moisture exhibits an acceptable accuracy with a much lower median value of the MBE ($-0.063 \text{ cm}^3 \text{ cm}^{-3}$) and RMSE ($-0.068 \text{ cm}^3 \text{ cm}^{-3}$) and a slight higher median value of $R$ ($0.771$). The 10- and 20-cm soil moisture values show evident improvements after assimilation according to the MBE and RMSE and a slight improvement according to the $R$. A relatively small improvement is found for soil moisture at 40 cm. Considering the limited penetration of microwave, there is a weak correlation between brightness temperature and soil moisture at deep layers. Therefore, in order to

<table>
<thead>
<tr>
<th></th>
<th>5 cm</th>
<th></th>
<th>10 cm</th>
<th></th>
<th>20 cm</th>
<th></th>
<th>40 cm</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OL</td>
<td>DA</td>
<td>AMSR2</td>
<td>OL</td>
<td>DA</td>
<td>OL</td>
<td>DA</td>
<td>OL</td>
</tr>
<tr>
<td>MBE</td>
<td>$-0.113$</td>
<td>$-0.018$</td>
<td>$0.068$</td>
<td>$-0.080$</td>
<td>$-0.028$</td>
<td>$-0.073$</td>
<td>$-0.035$</td>
<td>$-0.123$</td>
</tr>
<tr>
<td>RMSE</td>
<td>$0.119$</td>
<td>$0.039$</td>
<td>$0.096$</td>
<td>$0.085$</td>
<td>$0.046$</td>
<td>$0.076$</td>
<td>$0.046$</td>
<td>$0.124$</td>
</tr>
<tr>
<td>$R$</td>
<td>$0.804$</td>
<td>$0.796$</td>
<td>$0.781$</td>
<td>$0.865$</td>
<td>$0.851$</td>
<td>$0.868$</td>
<td>$0.804$</td>
<td>$0.832$</td>
</tr>
</tbody>
</table>

**TABLE 2.** Statistic metrics of soil moisture from model simulation (OL), assimilation experiment (DA), and AMSR2 soil moisture retrieval (AMSR2).

[Figure 5. (left) Boxplots showing the statistical metrics of soil moisture at four layers. (right) The spatial distribution of (top) NER_MBE, (middle) NER_RMSE, and (bottom) NER_R of soil moisture at 5 cm.]

Unauthenticated | Downloaded 06/29/24 11:13 PM UTC
effectively and correctly improve the soil moisture estimation at deep layers, more informative satellite data are expected to be introduced into the data assimilation framework, such as GRACE (reflecting the change of total water storage). Generally, fusion of multisource satellite data can be helpful to acquire an accurate soil moisture profile.

The effect of assimilation on soil moisture at 5 cm is clearly visible according to the NER_MBE, NER_RMSE, and NER_R values at each grid. The positive NER_MBE and NER_RMSE values at most grids indicate that the assimilation experiment successfully improves the soil moisture estimation, except for four grids (improve, blue; deteriorate, red). Furthermore, the reduction of the MBE is more apparent, with a value of greater than 20% and up to 98% than that of the RMSE, with a value of more than 19% and up to 71%. The deterioration of assimilation at the four grids is partly due to the accurate soil moisture simulation (MBE of soil moisture from OL are 0.019, 0.039, 0.006, and 0.041 cm$^3$ cm$^{-3}$). The poor correlation between the differences in soil moisture simulation and in situ measurements and the differences in brightness temperature simulation and the AMSR2 observation (according to the equation and theory of the EnKF) can also cause the deterioration. However, no more valid conclusions can be drawn from the deterioration. In terms of the NER_R, the improvement can be found in about half of the grids with a value of no more than 40%, while other grids show slight degradation in R.

Figure 6 is the same as Fig. 5, but it is for the result of soil temperature. Corresponding to the underestimation of soil moisture, OL tends to overestimate soil temperature, showing similar simulation accuracy at four layers. DA$_{6.9}$ greatly reduces the extremely overestimated surface soil temperature, with the median value of MBE decreasing from 5.996 to 0.186 K. However, DA$_{6.9}$ slightly underestimates the soil temperature at deep layers. According to the RMSE and R,
DA6.9 produces more accurate soil temperature estimation at four layers than OL. DA6.9 helps improve the soil temperature accuracy in two different ways: 1) the assimilation of the MODIS land surface temperature product; 2) the change of soil moisture after the assimilation of the AMSR2 brightness temperature. The increase of soil moisture from DA6.9 can alleviate the overestimation of soil temperature due to the larger heat capacity, thermal conductivity, and latent heat of evaporation of water. As indicated by the NER_MBE, NER_RMSE, and NER_R of soil temperature at 5 cm, DA6.9 shows a positive effect on surface soil temperature estimation at all grids, and the improvement in the MBE is more prominent.

c. Validation of estimated parameters

The sand, clay and organic carbon content of soil are measured at every ground station in the Naqu network, thus the “true” parameters for a model grid are represented by the station-average values within this grid. Figure 7 shows the comparison of the default parameters used in the CoLM, the estimated parameters from assimilation experiment and in situ measurements of 29 model grids. The default parameters show little variation at most grids, which is different from the characteristic of measurements. DA6.9 shows degradation of the estimation of clay content in almost all grids, but improvement of estimation of sand content and organic matter content after assimilation can be found in 45% and 20% of the grids, respectively. It is worth noting that the clay content measurements are quite small, and an obvious overestimation of clay content can be seen in the default values from the CoLM. With assimilation, the overestimation of clay content is aggravated at most grids. As indicated by Fig. 5, soil moisture is underestimated by the CoLM before assimilation. Therefore, the overestimation of clay content tends to be compatible with the increase of soil moisture after assimilation. A high clay content means a high water-holding capacity and low water conductivity (Saxton and Rawls 2006). It is obvious that the estimated RMS tends to be similar at all grids and shows limited difference with the default values, which is different from other parameters. It should be pointed out that soil texture affects the simulation of both soil moisture and brightness temperature, which may make assimilation have a more obvious effect on soil texture estimation.

From the above results, assimilation does not lead to a satisfactory improvement in parameter estimation. The parameters obtained by assimilation can be called effective parameters, which can hardly be perfectly consistent with the in situ measurements. This is due to that the discrepancy between model simulation and observation is ascribe to not only the uncertainties in parameters but also the model misconceptualizations, input information errors, and spatial scale effects. To further validate the influence of the estimated parameters on soil moisture simulation, the CoLM runs with default parameters (sim1) and estimated parameters (sim2), respectively, and the MBE of soil moisture is calculated. Figure 8 plots the spatial distribution of the NER_MBE \((1 - \text{MBE}_{\text{sim2}}/\text{MBE}_{\text{sim1}})\) of soil moisture at 5 cm. It is obvious that the estimated parameters can reduce the bias and improve the accuracy of surface soil moisture simulation. NER_MBE varies from \(-3.315\) to \(0.969\), and the two grids with a negative value is indeed corresponding to

![Figure 7](image_url)

**Fig. 7.** The comparison of the parameters estimated by assimilation experiment and in situ measurements.
the ones with soil moisture estimation deterioration in assimilation.

d. Comparison of 7.3- and 6.9-GHz assimilation

In this study, the effect of assimilating the 7.3-GHz brightness temperature on soil moisture estimation over the Tibetan Plateau is also investigated. All the experimental settings for data assimilation are maintained, except for substituting the 6.9-GHz V-polarization brightness temperatures (DA6.9) by the 7.3-GHz V-polarization brightness temperatures (DA7.3). The monthly average soil moisture at 5 cm over the study area obtained from DA7.3 are displayed in the third column of Fig. 2. It is worth noting that the soil moisture estimation from DA7.3 shows a subtle difference with that from DA6.9 in most of the Tibetan Plateau, varying between 0.02 and 0.02 cm$^3$ cm$^{-3}$. The difference of the spatial mean of soil moisture over the study area between DA6.9 and DA7.3 is no greater than 0.007 for every month.

Moreover, the performances of DA6.9 and DA7.3 are evaluated using the in situ measurements at 29 model grids. Figure 9 compares the NER_RMSE values of soil moisture at 5 cm from DA7.3 and DA6.9. As can be seen, the improvement of soil moisture estimation is similar between DA7.3 and DA6.9 in the Naqu network. Although the lower frequency is more sensitive to soil moisture, the assimilation of 6.9-GHz brightness temperature does not seem to be dominant in the soil moisture estimation. In addition, the degradation in assimilation performance at the three grids is not greatly improved by DA7.3. According to the results, neither the 6.9-GHz channel nor the 7.3-GHz channel showed obvious advantages on the soil moisture estimation. This is probably because the 6.9-GHz channel is not much affected by the RFI in most regions of the Tibetan Plateau as detected by previous studies (Parinussa et al. 2015; de Nijs et al. 2015). More in situ measurements or related studies are required to distinguish the superiority between 6.9- and 7.3-GHz over this area in the future.

5. Conclusions

The low frequency of microwaves has great advantages in monitoring the variation of soil moisture at the global scale. As the microwave detector with multiple channels, the AMSR2 shows a great potential in soil moisture studies. However, the limited accuracy of the soil moisture retrieval from the AMSR2 has been reported for the Tibetan Plateau. Thus, this study conducted a brightness temperature assimilation experiment to acquire accurate soil moisture with spatial–temporal continuity on the Tibetan Plateau. The assimilation experiment was performed as the joint assimilation of the V-polarization brightness temperature of 6.9 GHz from the AMSR2 and land surface temperature from the MODIS, which was intended to accurately estimate the soil moisture and soil temperature. Moreover, four parameters (sand, clay, organic carbon, and surface roughness) related to soil moisture are estimated by data assimilation.

Assimilation changes the spatial distribution of soil moisture over the Tibetan Plateau with a decrease in the southeast region and an increase in the northwest region. However, soil moisture still exhibits a slightly gradual decline from the southeast to the northwest. When evaluating the performance of assimilation experiment using in situ measurements, evident improvements were achieved for both soil moisture and soil temperature over most of the regions. The assimilation results outperform the model simulation by evidently reducing the underestimation of soil moisture and the overestimation of soil temperature. Moreover, the comparison of the accuracy of the AMSR2 soil moisture retrieval and that of the assimilated soil moisture further confirms the necessity of directly assimilating brightness temperature in soil moisture estimation. Although there are some discrepancies between the estimated parameters and measurement, the parameters derived from assimilation are proved to obtain more accurate soil moisture estimation than the default ones, which demonstrates the feasibility of parameters estimation with state data assimilation.

It is worth noting that the experimental period of this study is limited to the summer period. Therefore, there are still a number of issues that need to be addressed when using the brightness temperature assimilation to obtain the long-term
soil moisture estimation over the Tibetan Plateau. For example, the freeze–thaw state of the soil affects the infiltration of water flux, and it is important to accurately identify the freeze–thaw transition of the soil to estimate the soil moisture in winter. Overall, the promising results of assimilation demonstrate the capability of improving soil moisture simulation by assimilating AMSR2 brightness temperature over the Tibetan Plateau. With the availability of abundant satellite data, data assimilation strategy in this study is expected to be applicable for the soil moisture estimation with microwave data at the continental scale.

Acknowledgments. This work is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA19040504 and the National Natural Science Foundation of China under Grants 41801271 and 41671375.


REFERENCES


