Development of the Coupled Atmosphere and Land Data Assimilation System (CALDAS) and Its Application Over the Tibetan Plateau

Mohamed Rasmy, Toshio Koike, David Kuria, Cyrus Raza Mirza, Xin Li, and Kun Yang

***Abstract*—Land surface heterogeneities are important for accurate estimation of land–atmosphere interactions and their feedbacks on water and energy budgets. To physically introduce existing land surface heterogeneities into a mesoscale model, a land data assimilation system was coupled with a mesoscale model (LDAS-A) to assimilate low-frequency satellite microwave observations for soil moisture and the combined system was applied in the Tibetan Plateau. Though the assimilated soil moisture distribution showed high correlation with Advanced Microwave Scanning Radiometer on the Earth Observing System soil moisture retrievals, the assimilated land surface conditions suffered substantial errors and drifts owing to predicted model forcings (i.e., solar radiation and rainfall). To overcome this operational pitfall, the Coupled Land and Atmosphere Data Assimilation System (CALDAS) was developed by coupling the LDAS-A with a cloud microphysics data assimilation. CALDAS assimilated lower frequency microwave data to improve representation of land surface conditions, and merged them with higher frequency microwave data to improve the representation of atmospheric conditions over land surfaces. The simulation results showed that CALDAS effectively assimilated atmospheric information contained in higher frequency microwave data and significantly improved correlation of cloud distribution compared with satellite observation. CALDAS also improved biases in cloud conditions and associated rainfall events, which contaminated land surface conditions in LDAS-A. Improvements in predicted clouds resulted in better land surface model forcings (i.e., solar radiation and rainfall), which maintained assimilated surface conditions in accordance with observed conditions during the model forecast. Improvements in both atmospheric forcings and land surface conditions enhanced land–atmosphere interactions in the CALDAS model, as confirmed by radiosonde observations.**

***Index Terms*—Cloud data assimilation, land–atmosphere interactions, land data assimilation (LDAS), microwave remote sensing, numerical weather prediction (NWP).**

Manuscript received June 9, 2011; revised January 3, 2012; accepted February 22, 2012. Date of publication April 24, 2012; date of current version October 24, 2012.

M. Rasmy and T. Koike are with the River and Environmental Engineering Laboratory, Department of Civil Engineering, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: rasmy@hydra.t.u-tokyo.ac.jp; tkoike@hydra.t.u-tokyo.ac.jp).

D. Kuria is with Kimathi University College of Technology, 10100 Nyeri, Kenya (e-mail: dn.kuria@gmail.com).

C. R. Mirza was with the River and Environmental Engineering Laboratory, Department of Civil Engineering, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: craza@hydra.t.u-tokyo.ac.jp).

X. Li is with the Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China (e-mail: lixin@lzb.ac.cn).

K. Yang is with Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100085, China (e-mail: yangk@itpcas.ac.cn).

Digital Object Identifier 10.1109/TGRS.2012.2190517

I. Introduction

LAND surface heterogeneities, which control land surface processes and therefore land–atmosphere interactions, have an essential role in determining both water and energy budgets. Soil moisture is a crucial element in accurate land surface modeling owing to its control on the partitioning of water and energy fluxes, which in turn regulates land–atmosphere interactions. Several studies showed that soil moisture affects both large- and small-scale circulations [1], [4], [30], modulates meteorological droughts, floods, and precipitation [30], [43], [45]. As *in situ* soil moisture information cannot be used in numerical weather prediction (NWP) applications, satellite remote sensing (particularly L- or C-band passive microwave observations) is the most promising technique permitting quantitative estimation of soil moisture on daily basis and at regional and global scales [11], [13], [28]. Although many satellite-derived soil moisture outputs are available, knowledge to integrate these outputs into NWP models is very limited. Recent studies on assimilating soil moisture products within a NWP model had several limitations and thus recommended for the development of more advanced assimilation schemes [9], [10], [48]. As a result, [46] developed a satellite land data assimilation system coupled with a mesoscale model (LDAS-A) to directly assimilate lower frequency microwave radiances. The results obtained for Western Tibet showed that the LDAS-A system is capable of improving land surface variables (i.e., soil moisture and surface temperature) and land–atmosphere interactions in a mesoscale model. They further showed that improvements in solar radiation, in addition to soil moisture, are necessary to introduce realistic land–atmosphere interactions into a mesoscale model.

Furthermore, a land data assimilation can only ameliorate the uncertainties and improve representation of model surface conditions when satellite observations are available. Because of limited satellite overpasses (e.g., Advanced Microwave Scanning Radiometer on the Earth Observing System (AMSR-E), at a bidaily maximum), the improved land surface conditions often suffer from substantial errors and drift from the biases in the predicted forcings of LSMs during the model forecast. This is particularly the case for precipitation and solar radiation, which are the most critical inputs for accurate land surface modeling. For example, the exact timing and amount of rainfalls are often incorrect in model predictions and introduce strong biases in the assimilated land surface conditions at very short time scales. In addition, when a model predicts rainfall events,

solar radiation at surface is also reduced owing to model-simulated cloud conditions. As a result, the partitioning of turbulent fluxes, land–atmosphere interactions, and associated atmospheric processes will be misguided during the subsequent forecast [42]. To reduce biases in atmospheric forcings, a perturbation can be added to the predicted forcings [31]. However, the perturbation methods requires expenses of additional model integrations [15] and pitfalls that arise from modeling accurate forcing for land surface models are still unsolved, and are the most challenging issues faced by the land and atmosphere modeling communities. Solving these issues is essential for both land surface and NWP model applications, and will greatly help to understand and predict regional/global water and energy budgets accurately.

Clouds directly influence the surface meteorological forcings (i.e., downward radiation and rainfall) and thus strongly affect the estimation of Earth's surface water and energy budgets. As a result, this study focused on improving cloud representations in NWP models. Previous studies showed that the uncertainties in cloud microphysical processes and initialization of moisture fields [e.g., water vapor (WV) and cloud liquid water content (CLWC)] are major stumbling blocks in precipitation forecasts [25], [32], [37]. The assimilations of observed WV into NWP models showed significant impact on short-range rainfall prediction and development of convective systems [21], [22], [27]. Since ground-based measurements of CLWC and WV are very limited, the assimilation of satellite observations at various frequencies are used to retrieve atmospheric moisture fields into NWP models [20], [26], [29], [34]. Multispectral infrared images (e.g., Moderate-resolution Imaging Spectroradiometer), infrared (e.g., the Atmospheric Infrared Sounder) and higher frequency of microwave [e.g., Advanced Microwave Sounder Unit (AMSU)] sounders are widely used to obtain cloud characteristics and sounding information of atmosphere over land surfaces. Passive microwave observations (e.g., AMSR-E) also contain much information on bulk atmospheric moisture fields at higher frequencies; but these observations have not been well utilized over land surfaces. Because of much weaker and homogeneous water surface emission, methods to retrieve atmospheric moisture fields over water bodies were successful [2], [41], [44], [54]. Available methods over land employed mainly empirical approaches that used polarization differences based on regression equations [6], [7]. Those methods could result in large retrieval errors owing to brightness temperature depolarization attributed to the lack of homogeneity, roughness and dense vegetation. Compared with water surface (cold) emission, land surface emission exhibits strong temporal and spatial variability [3] because of different surface types, roughness conditions, and moisture content. These are major obstructions to microwave atmospheric retrieval over land.

In this paper, we have developed a system, referred to as the Coupled Atmosphere and Land Data Assimilation System (CALDAS), to physically introduce existing land and atmospheric moisture heterogeneities from AMSR-E observations into a mesoscale model. CALDAS assimilated AMSR-E lower frequency brightness temperature observations to better estimate land surface conditions, and merged them with higher frequency brightness temperature observations to estimate atmospheric moisture fields over land surfaces. To meet NWP model requirements, CALDAS was implemented on a standardized interface referred to as the Coupler, to handle effectively coupling and exchanges of data among individual system components (i.e., atmospheric model, model operators and assimilation methods).

The paper is organized as follows. In Section II, system development and its major components are explained briefly. In Section III, data sets and detailed methods of the system execution are outlined. The study domain and model setup are presented in Section IV. Based on numerical experiments, Section V investigates the applicability of microwave higher frequency observations in conjunction with lower frequency observations over the land surface. Further examinations have been carried out on the synchronized improvements of land and atmospheric initialization and its influences on model simulated land and atmosphere evolutions. Finally, in Section VI, conclusions from our research findings are presented.

II. System Development

CALDAS (Fig. 1) has three subsystems: 1) a land–atmosphere coupled mesoscale model [Advanced Regional Prediction System (ARPS)-Simple Biosphere model version 2 (SiB2)]; 2) a LDAS; and 3) a cloud microphysics data assimilation system (CMDAS). The combination of ARPS-SiB2-LDAS models are named as LDAS-A, and detailed information on LDAS-A was given in [46]. The following sections explain briefly the formulation of major components of the CALDAS system.

A. Mesoscale Model

Reference [52] compared the ARPS model performances against Weather Research and Forecasting (WRF) Model over the Indian region and showed that the ARPS model produced better results than WRF Model. Therefore, the ARPS was chosen as the atmospheric driver. The ARPS is a comprehensive regional- to storm-scale prediction system and its atmospheric prediction component is a 3-D, nonhydrostatic compressible model formulated in generalized terrain-following coordinates [56], [57].

B. Land Data Assimilation System

The LDAS consists of a land surface scheme, a physics-based radiative transfer model and a sequential assimilation algorithm. It improves the surface moisture heterogeneity through the assimilation of lower-frequency passive microwave observations, such as those at 6.9 and 10.65 GHz. The components of the LDAS are explained briefly in the following subsections.

1) Land Surface Scheme (Model Operator): The SiB2 was chosen as the LSM for ARPS and as the model operator for the LDAS. The SiB2 model simplified the soil depth into three soil sublayers: a surface soil layer that acts as a significant source of direct evaporation when moist; a root zone, which is the supplier of soil moisture to the roots and accounts for transpiration; and a deep soil layer that acts as a source

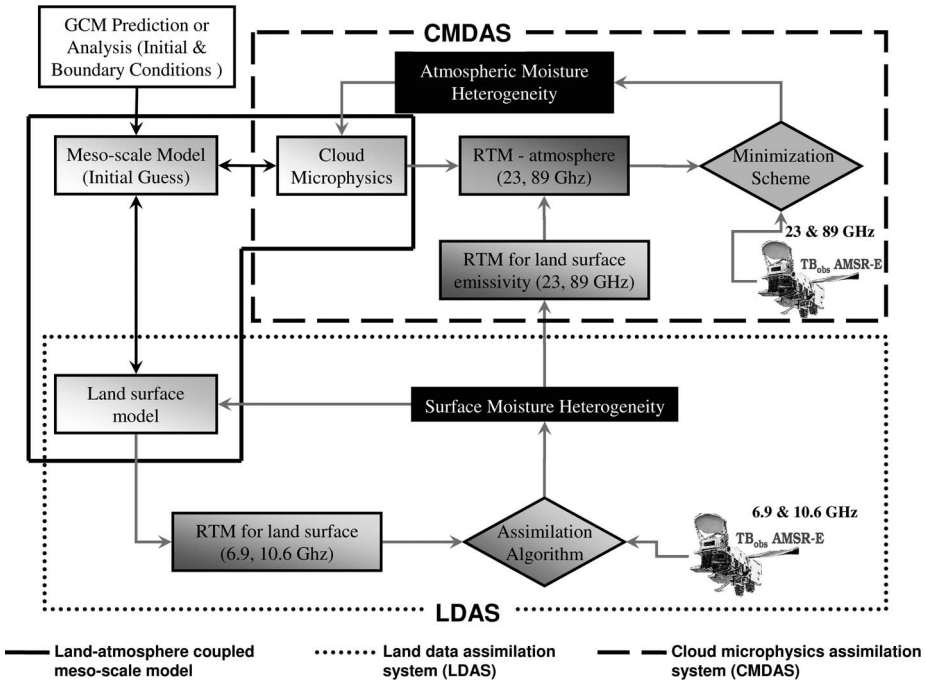


Fig. 1. Schematic diagram of Coupled Atmosphere and Land Data Assimilation System (CALDAS), its principal components, and the linkage between individual components.

for hydrological base flow and upward recharge of the root zone [49].

2) Radiative Transfer Model (Observation Operator): In the 1–15 GHz microwave region, the atmosphere is transparent even with the presence of clouds and moderate rainfall. Neglecting atmospheric and rainfall effects and assuming that reflection at the surface is much less than radiation from the surface and vegetation layer at lower frequencies, the brightness temperature T_b at the satellite level is given by [17]

$$T_b = (1 - R_p)T_s \left[\exp(-\tau_c) + [(1 - \omega_c)(1 - \exp(-\tau_c)T_c]\right]. \tag{1}$$

Here, R_p is the surface reflectivity and T_s is the surface physical temperature (K), ω_c is the single-scattering albedo of the canopy, τ_c is the vegetation optical thickness and T_c is the canopy temperature (K). The estimation of surface reflectivity for smooth surfaces is given in Appendix A. To account for surface roughness effects, the Advanced Integral Equation Model incorporating a shadowing effect [5], [33] was considered. To account for the volume scattering effects (due to heterogeneity of soil media), we adopted a method proposed by [38], [39].

3) Ensemble Kalman Filter (Assimilation Algorithm): The ensemble Kalman filter (EnKF) is a sequential estimation procedure for nonlinear models that updates the system whenever observations are available [14]. The EnKF represents the distribution of the system state with a collection of state vectors, called an ensemble, and replaces the covariance matrix with the sample covariance computed from the ensemble. In the field of hydrology, [24], [47] applied the EnKF to soil moisture estimation and found that it performed well against the variational assimilation method. An overview of the EnKF for soil moisture assimilation was given in Appendix B and in [46].

C. Cloud Microphysics Data Assimilation System (CMDAS)

A 1-D Variational CMDAS was developed by [41] over ocean and sea surfaces to improve the atmospheric moisture fields through assimilation of higher-frequency AMSR-E observations, such as 23.8 GHz and 89.0 GHz. In this paper, the model formulation follows [41] except in calculating boundary layer (land surface) emissivity. At this time, CMDAS considered atmospheric moisture variables such as those of cloud liquid water, WV, rain and snow as assimilation parameters, and cloud ice and hail were the by-products of data assimilation and Lin's ice microphysics [35]. For completeness, a brief explanation is given below.

1) Cloud Microphysics Scheme (Model Operator): Clouds are generally in a mixed-phase of cloud droplets and ice crystals. To be consistent with the atmospheric geophysical parameters and to establish realistic retrieval of moisture fields, Lin's ice microphysics [35] was chosen to be the model operator for CMDAS. Lin's scheme considers five classes of hydrometeors; cloud water, cloud ice, rain, snow, and hail. Exponential particle size distribution functions are hypothesized to describe the particle sizes for rain [40], for snow [19], and for hail [16].

2) Radiative Transfer Model (Observation Operator): At higher frequencies of AMSR-E, absorption and scattering effects dominate and increase with increasing hydrometeor liquid and solid phase. As a result, 89 GHz reaches saturation at very low atmospheric moisture states, and after that, the relationship between brightness temperature and cloud condensate decreases monotonically. WV is a polar molecule with an electric dipole and therefore produces rotation lines at 22.2 GHz. As a result, it exhibits strong absorptions (therefore strong emissions) at this frequency. Because of the applicability of soil RTM to wider frequency ranges, the same RTM used for

lower frequencies is used to estimate the surface emission and scattering effect on upwelling radiation at higher frequencies. To estimate the atmospheric absorption, emission and scattering effects on upwelling radiation at higher microwave frequencies, the four-stream fast model [36] was used. The model formulations are given in Appendix C.

3) Shuffled Complex Evolution (Minimization Scheme): The Shuffled Complex Evolution (SCE) technique developed by [12] was adopted as a minimization scheme in CMDAS. The SCE is a global minimization scheme, in which competitive evolution and complex shuffling are incorporated to ensure that information contained in the sample is efficiently and thoroughly exploited. The SCE algorithm is based on an iterative method, where a cost function is minimized by adjusting the state vector or assimilation parameters. The cost function $J(X)$ is given by

$$J(X) = \frac{1}{2}\left[(X_b - X)^T B^{-1}(X_b - X)\right] + \frac{1}{2}\left[\left[Y_0 - H(X)\right]^T R^{-1}\left[Y_0 - H(X)\right]\right]. \quad (2)$$

Here, B is the background error covariance matrix, and R is the observation error covariance matrix.

The estimation of background errors for cloud integrated condensates (liquid and ice components) in spatial and temporal scales is very difficult at this stage, because of the unavailability of detailed cloud observations. As a result, to explore potential of microwave cloud observations, the background error term (J_b) is assumed to be equal to zero. A further assumption was made for observations, where all observations have the same error that is equal to a unit matrix $(R = I)$. This method is a simplest data assimilation method similar to direct insertion method, which updates the system whenever the observation is available assuming that data are perfect (i.e., a hard update, by the corresponding observation at the same point without consideration of model and observational errors). Considering these assumptions, (2) can be written as

$$J(X) = \frac{1}{2}\sum_{i=1}^{N}\left[Y_0 - H(X_i)\right]^T\left[Y_0 - H(X_i)\right] \quad (3)$$

$$X_i = M(X_o) \quad (4)$$

where Y_o is the observations (23.8 and 89 GHz), N is the number of brightness temperature observations, H is the radiative transfer model that represents land and atmosphere, M is Lin's ice microphysical scheme, and X is the atmospheric state.

D. Coupler

To establish a standard system with optional plug-ins, and to test the combination of different models and data sources, the coupler was designed to be modular and extensible. It was also enabled on parallel computing platforms to satisfy the increasingly high-performance computing requirements of operational NWP models.

III. Data sets and Methods

A. Data sets

The following sections explain briefly the data sets used within the CALDAS system.

1) In Situ Data: The automated weather station installed at Gaize station includes a system for measuring soil moisture and temperature, a precipitation gauge and a pyranometer (MS-802). It works continuously to collect data with at temporal resolution of 1 h. Soil moisture content is measured with a Trime EZ sensor at depths of 3, 20, and 40 cm and soil temperature is measured with a Pt100 resistance thermometer at depths of 0, 5, 10, 20, 40, and 80 cm. In this paper, data of soil moisture at 3 cm and soil temperature at the surface were used to validate the models because the microwave sensors are sensitive to a few centimeters ($\sim$3 cm) of soil depth.

In July, 2008, Chinese and Japanese scientists jointly conducted an atmospheric experiment at Gaize station and during these periods, radiosonde soundings (Vaisala Radiosonde RS92-SGP) were launched every 6 h. The observations made during the period 6th July–16th July were used to validate the models [55].

2) Japan Aerospace Exploration Agency (JAXA) Soil Moisture Products: The 10–36 GHz algorithm [18] was applied to AMSR-E data to estimate the surface soil moisture products. Because the water content of vegetation affects the sensitivity of the microwave remote sensing of soil moisture, [18] used a method for simultaneously retrieving the soil moisture and vegetation water content from two indices, i.e., polarization index and index of soil wetness. These global moisture products are available at 0.1° spatial resolution, bi-daily.

3) AMSR-E Brightness Temperature Data: AMSR-E is a six-frequency dual-polarized total-power passive microwave radiometer that detects microwave emission from the Earth's surface and atmosphere. It measures horizontally and vertically polarized brightness temperatures separately at 6.925, 10.65, 18.7, 23.8, 36.5, and 89.0 GHz. Its individual measurements have spatial resolutions varying from 5.4 km at 89.0 GHz to 56 km at 6.925 GHz. Even though the size of the instantaneous field-of-view for each channel differs, active scene measurements are recorded at equal intervals of 10 km (5 km for the 89.0-GHz channels) along the scan. Calibrated brightness temperature (Level 1B) data were obtained from the Japan Aerospace Exploration Agency. We employed vertical polarization at 6.925 and 10.65 GHz to retrieve soil moisture heterogeneities, and the vertical polarization at 23.8 GHz and 89.0 GHz to retrieve WV and cloud condensate. Because the observations have different spatial resolutions depending on the frequency, we used the nearest neighborhood interpolation method to resample data to our model horizontal resolution.

4) MTSAT-1R Data: MTSAT-1R is a geostationary satellite located at 140 °E and records images in five wavebands such as those of visible (VIS) (0.55–0.80 μm), infrared channel 1 (IR1) (10.3–11.30 μm), infrared channel 2 (IR2) (11.5–12.50 μm), infrared channel 3 (IR3) (6.5–7.0 μm), and infrared channel 4 (IR4) (3.5–4.0 μm). Image resolutions for VIS and infrared (IR 1–4) channels at nadir are 1 and 4 km, respectively.

To investigate the observed cloud activity over the selected domain, split-window IR1 (MTSAT/1R1) observations were obtained since no other direct measurements were available. The MTSAT/1R1 brightness temperature observations are measurements of cloud-top temperatures under the assumption that the clouds are blackbodies. In addition, lower brightness temperatures in an infrared image indicate sufficiently thick clouds with the cloud tops radiating at the atmospheric temperature at higher altitudes.

5) Initial and Boundary Conditions: The initial and lateral boundary conditions are derived from data of the National Center for Environmental Prediction (NCEP). These NCEP FNL (Final) Operational Global Analysis data are available on 1.0° by 1.0° grids at 6-h intervals. This product is derived from the Global Forecast System that runs four times a day in near-real time at the NCEP. To obtain the corresponding initial and boundary conditions required to run the mesoscale model, analysis results available at 26 mandatory levels from 1000 to 10 mb were used. Variables (e.g., pressure, geopotential height, temperature, relative humidity and u and v winds) required for each mesoscale model vertical layer were interpolated from the 26 mandatory pressure levels. NCEP continuously maintains the archived time series of the NCEP/FNL data set to a near-current date but not in real time (http://dss.ucar.edu/datasets/ds083.2).

B. Methods

1) As shown in Fig. 1, the land–atmosphere mesoscale model (ARPS-SiB2) was set up using initial and boundary conditions from NCEP FNL global analysis data.
2) The Coupler integrated the ARPS model for a predefined period (10 min), prepared the forcing data to the land surface model (SiB2), and transfered the calculated atmospheric forcing data to SiB2.
3) At the beginning of SiB2 integration, the ensemble (50 members) of soil moisture profiles was generated by adding pseudorandom noise with prescribed statistics to the first estimated soil moisture contents in the surface, root, and deep layers. SiB2 was simulated independently for each ensemble member of the soil moisture profile, keeping the same model parameters and atmospheric forcing. At the end of the SiB2 calculation, the mean values of the soil moisture profile, soil temperature profile, and surface heat and moisture fluxes were computed from the ensemble of the forecast and fed back to the ARPS model as the lower boundary conditions of the atmospheric model. Procedures (2 and 3) were continued until the AMSR-E measurements are available.
4) At times when AMSR-E observations were available, the brightness temperatures at 6.9 and 10.65 GHz were perturbed with prescribed statistics to produce an ensemble of observations. The SiB2-driven ensemble of soil moisture profiles, surface, and canopy temperatures were used to obtain the simulated brightness temperatures (at 6.9 and 10.7 GHz) using the forward microwave radiation transfer model and the EnKF calculated the updated soil moisture profiles.
5) As soon as the LDAS finished the assimilations, the CMDAS was activated by the Coupler. The control variables (profiles of temperature, specific humidity, pressure, air density, mixing ratio of cloud water, rain water, hail, snow, and cloud ice) were obtained from ARPS as an initial state to run the model operator (Lin's ice microphysics). The four-stream fast model calculated the modeled brightness temperatures for 23.8 and 89 GHz at satellite level by considering the land surface as the lowest boundary. Land surface emissivity was calculated using assimilated soil moisture content, estimated soil and canopy temperature and calibrated roughness parameters.
6) The assimilation scheme (SCE) updated the assimilation parameters (integral of WV, cloud liquid water, snow, and rain) within the feasible domain and minimized the cost function by calculating errors between modeled and observed brightness temperatures, while the model operator (Lin's ice microphysics) maintained consistency between model and assimilated variables.
7) Finally, with the reinitialized land surface and atmospheric conditions, the Coupler integrated the ARPS-SiB2 model forward in time to predict land and atmospheric evolution until the next AMSR-E observations were available.

IV. Experimental Descriptions

The performance of the new system was investigated by considering three simulations: 1) ARPS run, where a one-way nesting procedure employed the land–atmosphere mesoscale model without any assimilation; 2) LDAS-A run, in which a ARPS run was accompanied by sequential land data assimilation; and 3) CALDAS run, in which a ARPS run was accompanied by a sequential land data assimilation and a cloud microphysics data assimilation. The results were compared with surface Automatic Weather Station measurements, satellite observations, and radiosonde soundings.

A. Study Domain

To validate CALDAS system and to assess its retrieval capabilities, the system was applied to a mesoscale area on the Tibetan Plateau. The study area is located in the western part of the Plateau, including the Gaize station (84.05° E, 32.3° N) and is bounded by the area (83.20° E–85.70° E, 30.6° N–33.2° N) (Fig. 2). This region is characterized by a wider flat valley and mountainous topography with heterogeneous soil moisture distributions that is favorable for the study of land–atmosphere interactions. The land-use type is bare land, or sparse vegetation without intense human activity, which ensured the applicability of the AMSR-E observations in this region. The other main reason for selecting this region was the availability of data sets for model validation.

B. Model Configuration and Parameter Settings

To capture small-scale atmospheric features related to the land surface effect, the ARPS domain horizontal resolution

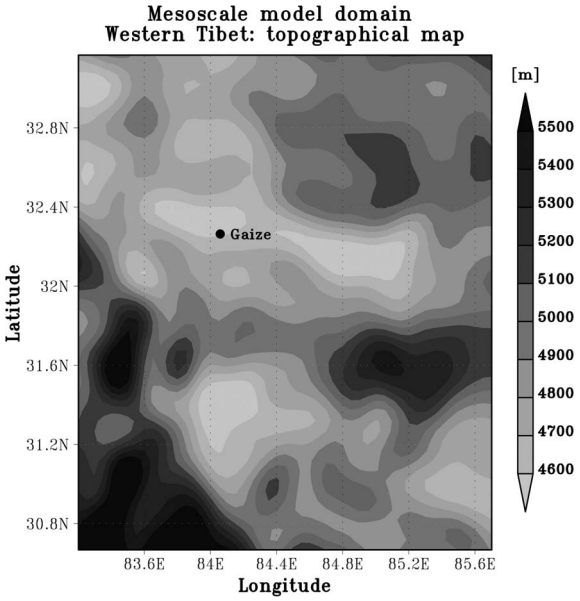


Fig. 2. Mesoscale model domain: topographical map (meters) including Gaize station.

was set to $0.05^\circ \times 0.05^\circ$. Total number of grids in an $x-y$-direction was 70×70 covering the domain area of 350×350 km^2. For the vertical grid, ARPS used a hyperbolic tangent function to stretch the grid interval from 40 m at the first level and 53 atmospheric layers in total ($\sim$18 km above the ground surface). The physical parameterization options were configured with a 1.5-order turbulent kinetic-energy-based closure scheme for sub-grid-scale turbulent mixing, latitude-dependent Coriolis parameters and Lin ice microphysics [35] as the microphysical processes. Geographical data sets such as soil type, land-use type and vegetation parameters, were obtained from ftp://aftp.fsl.noaa.gov/divisions/frd-laps/WRFSI/Geog_Data. The NCEP FNL data set was used to drive the initial and boundary conditions for model simulations. All three simulations were initialized with the same land and atmosphere initial conditions. The boundary conditions were introduced to the simulations at every 6 h. Spatially distributed radiative transfer model parameters used in this study were given in [46].

As integral values of moisture fields were assimilated by CMDAS, the assimilated integral values were distributed with predefined profiles. CLWC was assumed to have a parabolic distribution, with zero values above the top and below the bottom of cloud layers. In the cloud layers, air was saturated, and the remainder of WV was distributed above and below cloud layers. The distributions of rain and snow followed a skewed profile and began to form at the cloud tops, grew to their maximum at the cloud bottoms, and after that decreased because of evaporation and breakup of raindrop or snow flakes. To assimilate atmospheric moisture parameters, lower and upper limits of integrated values for the parameters have to be defined explicitly. Based on global positioning system observations at Gaize, lower and upper limits of IWV were set as $\sim$4 and $\sim$18 mm, respectively. Considering the nature of the clouds observed over the Tibetan Plateau (cumulus and cumulonimbus), the lower and upper bounds of ICLW have been set as 0 and 2.5 kgm^{-2}, respectively (http://www-das.uwyo.edu/~geerts/cwx/notes/chap08/moist_cloud.html). The upper bound for the other two parameters was selected arbitrarily since no information was available (Table I). In addition, the CALDAS considered only one cloud layer, and cloud depth has to be predefined before the assimilation takes place. In this paper, we investigated radiosonde observations (not shown), and according to [53] cloud layer top and base heights were set to 1500 and 5000 m (relative to the ground surface), respectively.

TABLE I
UPPER AND LOWER LIMITS OF ASSIMILATION PARAMETERS

Parameter	Unit	Lower limit	Upper Limit
Integrated Water Vapor (IWV)	kgm^{-2}	4.0	18.0
Integrated Cloud Liquid Water (ICLW)	kgm^{-2}	0.0	2.5
Integrated Rain Water (IRW)	kgm^{-2}	0.0	1.5
Integrated Snow Water (ISW)	kgm^{-2}	0.0	1.5

V. RESULTS AND DISCUSSIONS

Series of LDAS-A model simulations were performed to identify adverse effects of rainfall events on assimilated land surface conditions. Based on LDAS-A results, two simulation periods (1st–2nd July 2007 and 7th–8th July 2008) were selected to investigate performances of CALDAS.

A. Evaluation of Surface Soil Moisture Distribution

To assess reliability of assimilated land surface emission, simulated moisture contents were investigated with JAXA's soil moisture products because of unavailability of spatial independent measurements (e.g. infrared land observations were not available due to the presence of clouds over the model domain). Fig. 3(a)–(c) represent the spatial distribution of surface soil moisture for ARPS, CALDAS, and JAXA at 0730 UTC on 08th July 2008. The magnitude of soil moisture content simulated by ARPS was higher in most of the model grids (particularly in the west) and produced completely different spatial distribution when compared with the JAXA's product. On the other hand, the distribution of CALDAS soil moisture contents, particularly the dry and wet regions, was reasonably comparable to that of JAXA's product. Furthermore, the error statistics for model simulated surface soil moisture against JAXA's soil moisture are given in Table II (soil moisture contents in the mountainous region (above 5000 m) were excluded in the error calculations, because AMSR-E retrieval might contain biases due to slope and roughness effects on emitted microwave signal over very larger mountains). As shown in the table, CALDAS showed less mean bias error (MBE) and root mean square errors (RMSE) and had high correlation with JAXA's product compared against ARPS results. An important point to be noted from Fig. 3(b) and (c) and Table II is that JAXA's soil moisture contents were underestimated compared with CALDAS results. Quantitative validation of assimilated soil moisture content at the Gaize was shown in Fig. 9 and will be discussed later.

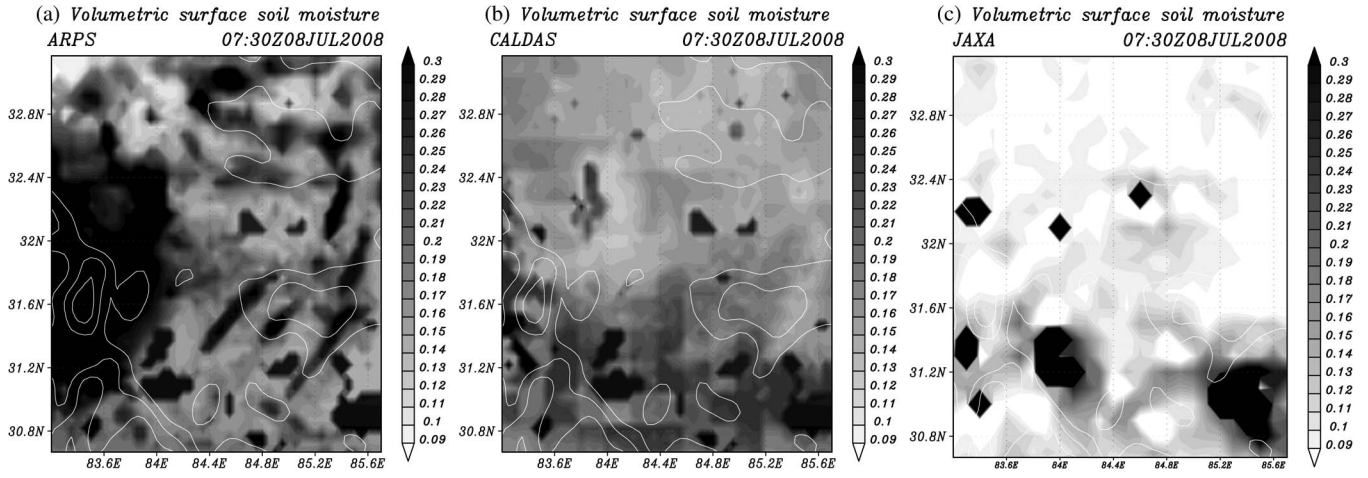


Fig. 3. Spatial distribution of simulated volumetric surface soil moisture (m^3/m^3) at 0730 UTC on 08th July 2008; (a) ARPS (no-assimilation), (b) CALDAS and (c) JAXA's product, contour lines depict topography.

TABLE II
ERROR STATISTICS FOR MODEL SIMULATED SURFACE SOIL MOISTURE WITH JAXA'S SOIL MOISTURE

Soil moisture	Unit	MBE	RMSE	Pearson corr. coeff.
ARPS	m^3/m^3	0.11	0.018	-0.005
CALDAS	m^3/m^3	0.07	0.007	0.540

B. Comparisons of Brightness Temperatures and Cloud Condensate With Satellite Observations

To investigate the performance of CALDAS, the assimilated brightness temperatures were analyzed. Fig. 4(a1)–(c1) and (a2)–(c2) compare the observed and assimilated brightness temperature at 89 and 23.8 GHz, respectively. The brighter grid cells (lower brightness temperatures) in Fig. 4(a1) indicated the existence of deep convective cloud and/or precipitation systems within these grid cells. As shown in Fig. 4(b1), the assimilated brightness temperatures were in reasonably good agreement when compared with the observed brightness temperatures as shown in Fig. 4(a1). The calculated absolute error distribution at 89 GHz, as shown in Fig. 4(c1), is less than 5 K for all over the model domain, except for a few grids cells. Similarly, as shown in Fig. 4(a2) and (b2), the assimilated brightness temperature at 23.8 GHz compares reasonably well with observed brightness temperature. The absolute difference of 23.8 GHz as shown in Fig. 4(c2) is less than 5 K in most of the model grid cells. In a few grid cells, the errors exceeded 10 K but were less than 15 K. When the error distribution of 23.8 GHz is compared with that of 89 GHz, 23.8 Ghz showed slightly higher values ($\sim$5–10 K) than at 89 GHz. These error differences between observed and assimilated brightness temperatures could result from the following: 1) the degree of sensitivity of a frequency to land surface emission, i.e., AMSR-E 23.8 GHz channel is more sensitive than the 89 GHz; 2) errors in the calculation of land surface emission (defined as surface emissivity multiplied by surface physical temperature). In RTM, the predicted surface temperature was used as a physical temperature, and when the model grid experiences overcast sky conditions the model's temperature will be lower than the actual temperature owing to less solar radiation inputs; and 3) some portion of the errors are attributed to larger contrasts in spatial resolution between AMSR-E frequencies (50 km-5 km), which therefore limits the accurate estimation of land surface emission at finer scales. For these reasons, the absolute errors were greater than 5 K in some of the grids for both frequencies, and errors at 23.8 GHz showed higher biases than 89 GHz.

The quantitative validations of individual or integrated atmospheric moisture variables (i.e., cloud liquid water, rain, snow, and hail) are difficult even at point locations because of unavailability of such data. A few variables (cloud liquid water and rain rate) from satellite outputs are available over ocean, but over land, these variables are still under investigation. Therefore, to evaluate the reliability of the assimilated cloud condensate, qualitative comparisons were performed between simulated vertically integrated condensate (summation of liquid and solid phases) and AMSR-E observation at 89 GHz. Fig. 5(a)–(c) shows the model simulation of vertically integrated cloud condensate at 0730 UTC on 08th July 2008 (immediately after the land and atmosphere coupled data assimilations) for ARPS, LDAS-A, and CALDAS, respectively. The distributions of total condensate of both ARPS and LDAS-A were similar, showed much higher values in the north, where the Gaize station is located, and were completely different from AMSR-E observation at 89 GHz [Fig. 5(d)]. On the other hand, the CALDAS simulation showed that atmospheric data assimilation significantly removed the predicted clouds, which were not observed by AMR-E at 89 GHz, particularly in the north and over the Gaize station when compared with ARPS and LDAS-A cases. The CALDAS also produced cloud activity as observed by AMSR-E, particularly in the south [Fig. 5(c)]. The assimilated cloud distributions were coherent and compared well with the observed cloud cell distributions of AMSR-E observations at 89 GHz.

To further investigate reliability of the vertical distribution of assimilated cloud parameters, two grid points were selected such that one point represented the Gaize station where satellite observations showed no clouds, and the other point (X) represented a grid cell located at [84E, 31.2N], where satellite

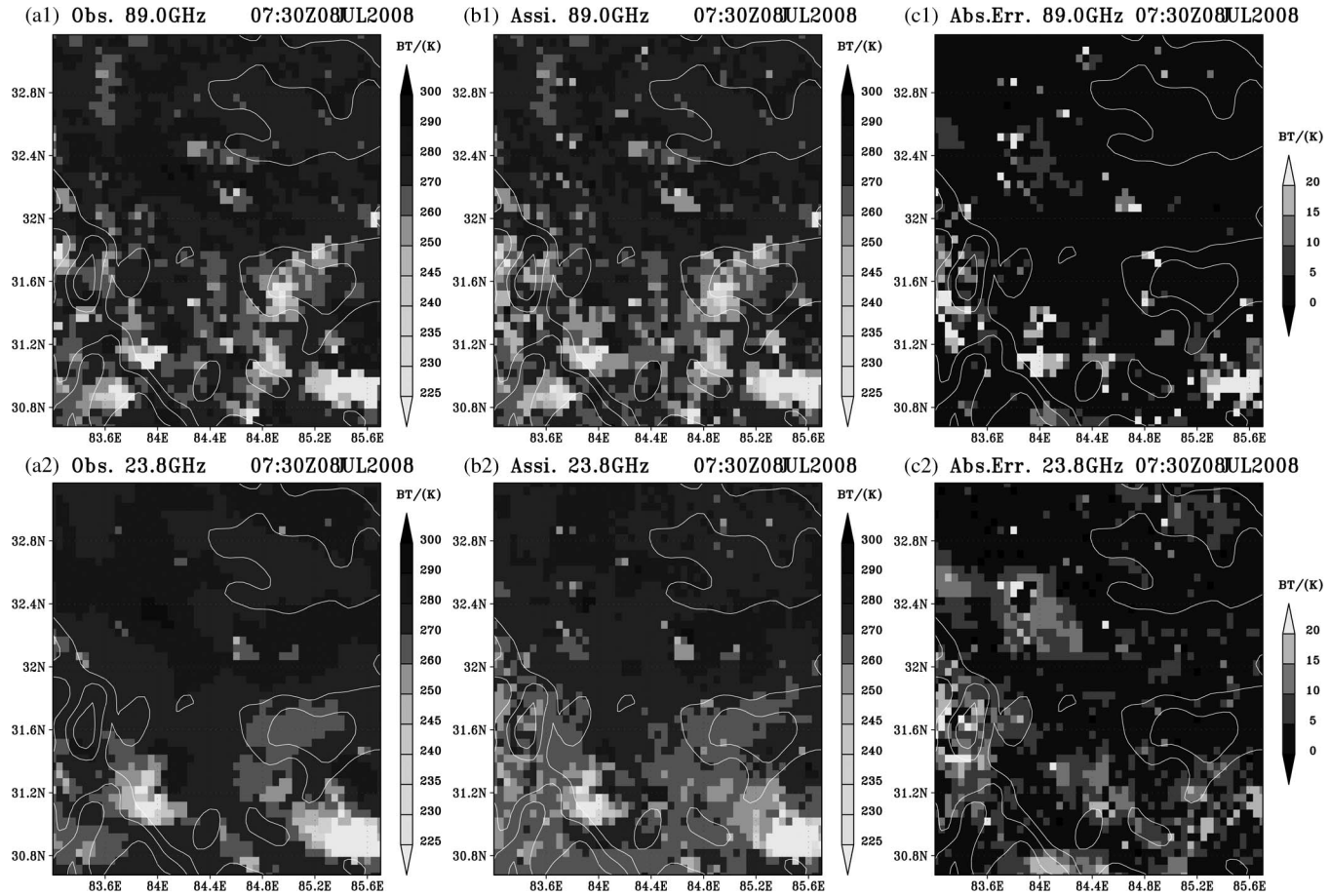


Fig. 4. Spatial distribution of brightness temperatures (K) at 0730 UTC on 08th July 2008; (a1) and (a2) observation by AMSR-E, (b1) and (b2) assimilation by CALDAS, (c1) and (c2) absolute error between observation and assimilation at 89 GHz and 23 GHz, respectively.

observation indicated clouds. Fig. 6 shows variation in model predicted cloud condensates with pressure height. As shown in Fig. 6(a), at the Gaize model grid point, ARPS and LDAS-A model simulations indicated the existence of cloud between ∼500 hPa to ∼300 hPa. On the other hand, the assimilation of cloud parameters in the CALDAS simulation effectively removed the predicted cloud and showed almost zero values of cloud condensate for the entire atmospheric column. Fig. 6(b) is for the point X, where AMSR-E observed cloud activity. The ARPS model did not predict any clouds, whereas LDAS-A showed very thin and light cloud between ∼350 hPa to ∼300 hPa. However, CALDAS predicted tall and dense clouds from ∼500 hPa to ∼300 hPa. Therefore, CALDAS as a multifrequency assimilation system is very promising and has the potential to reasonably retrieve atmospheric information over land surfaces. However, the quantitative information on each atmospheric moisture variable and microphysical properties has to be investigated using reliable data sets to asses the full capabilities of the CALDAS model.

To investigate the effect of assimilations during the model forecast, geostationary meteorological satellite infrared cloud-top temperature observations (MTSAT/1R1) were used. The model-simulated cloud-top temperature was estimated based on the study by [46]. Fig. 7(a) and (b) shows the diurnal variation in spatial correlations and RMSE calculated from cloud top temperatures simulated by models and observed by MTSAT/1R1 on 08th July 2008, respectively. Because Gaize is located at the periphery of MTSAT's field of view, MTSAT experiences a slight shift in observed cloud positions compared with AMSR-E observations. To eliminate this shift, MTSAT/1R1 and model cloud top temperatures were up-scaled to 0.5°. Negative highs in the spatial correlation indicate mismatch between MTSAT/1R1 observations and model simulations of cloud positions. Positive highs indicate coherences between MTSAT/1R1 observation and model simulations. Prior to land and atmospheric assimilation, all three models showed relatively similar and lower (∼0.2) spatial correlations, which became negative after 0600 UTC. At the time when CALDAS performed both land and atmospheric assimilation, the correlations were negative and RMSEs were maximum. However, immediately after reinitialization, the CALDAS results showed rapid increase (from −0.1 to +0.5) in the spatial correlation, and rapid decrease in RMSE (from 53 to 38 K). One and half hours after the assimilation, correlation reached a maximum of +0.72 and RMSE reached a minimum of 33 K. Whereas the ARPS and LDAS-A model simulations showed negative or very low spatial correlations and higher RMSEs with MTSAT/1R1 observations. These results indicates that CALDAS was reinitialized with improved cloud distributions, which coincided well with MTSAT/1R1 cloud observations and realistic assimilation of cloud condensate enhanced simulated cloud activity over the model domain.

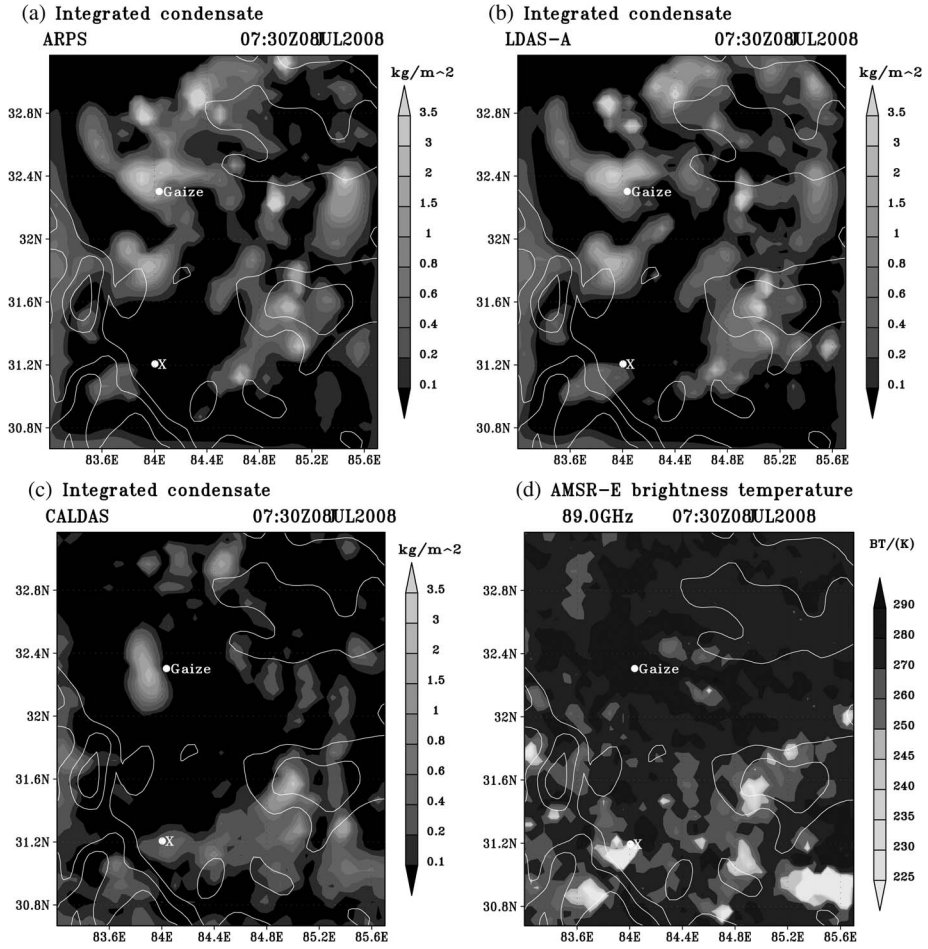


Fig. 5. Spatial distribution of vertically integrated condensate (liquid plus solid phases) ($\mathrm{Kg/m^2}$) at 0730 UTC on 08th July 2008; (a) ARPS, (b) LDAS-A, (c) CALDAS, and (d) AMSR-E brightness temperature at 89 GHz.

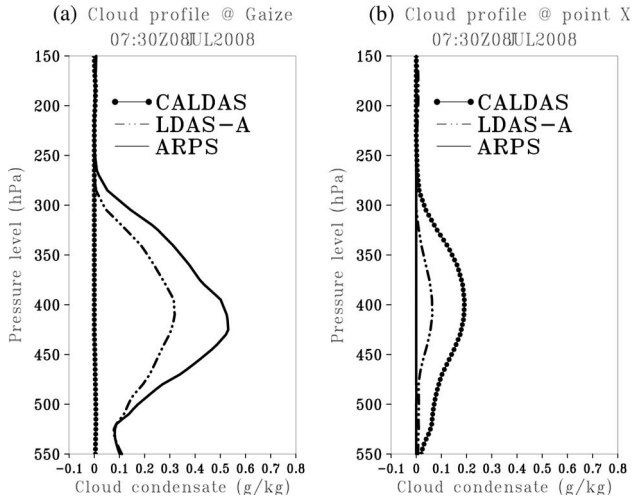


Fig. 6. Profiles of simulated cloud condensate (g/kg) from ARPS, LDAS-A, and CALDAS at 0730 UTC on 08th July 2008; (a) at Gaize station (b) at a grid cell (marked as X in Fig. 5) located at [84° E, 31.2° N].

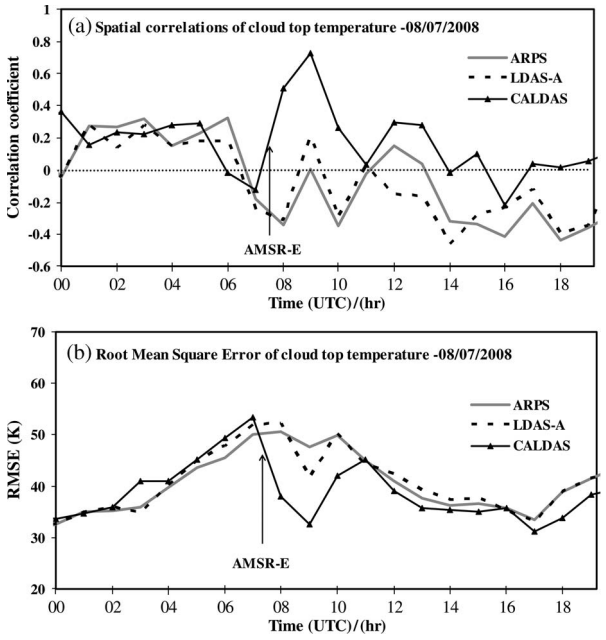


Fig. 7. Hourly variation of error statistics calculated from model simulated cloud top temperatures and MTSAT/1R1 cloud top temperatures on 08th July 2008; (a) spatial correlations, (b) root mean square errors (RMSE).

Fig. 8 compares cloud top temperature distributions at 0900 UTC on 8th July 2008, when the spatial correlation was highest and RMSE was lowest in the CALDAS simulation. Very active

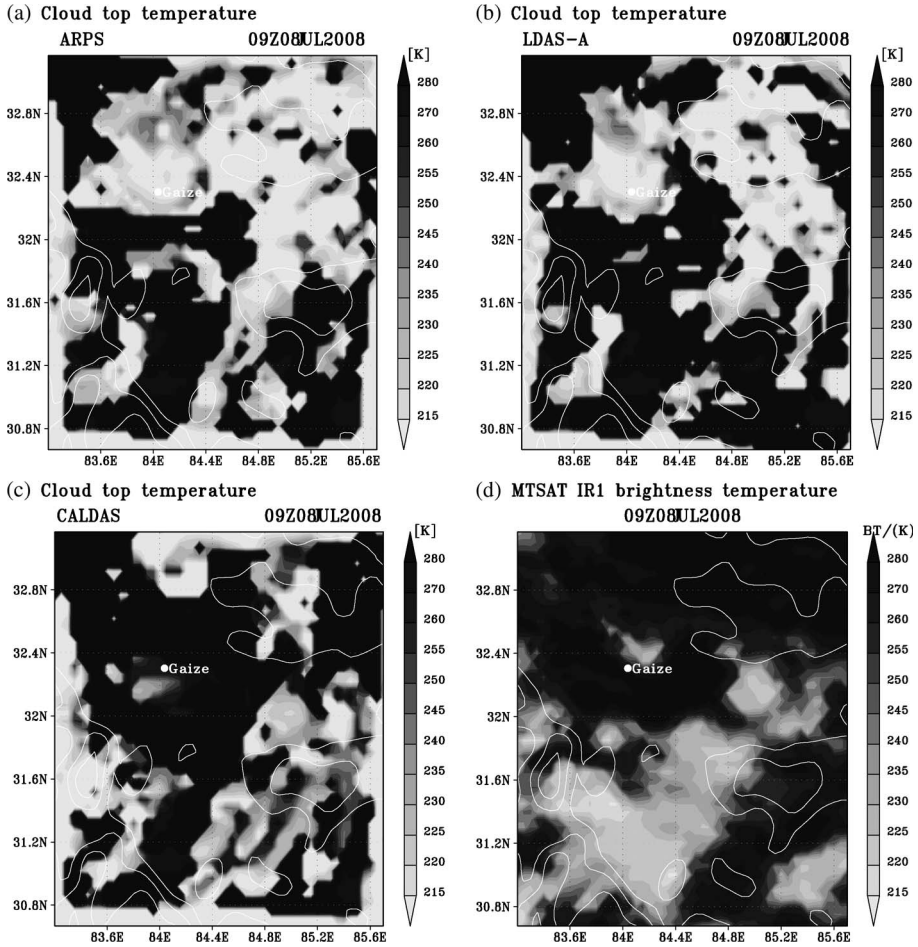


Fig. 8. Comparison of model simulated cloud top temperature with satellite observation at 0800 UTC on 08th July 2008 (1.5 h after assimilation); (a) ARPS case, (b) LDAS-A case, (c) CALDAS case, and (d) Geo-stationary meteorological satellite (MTSAT) infrared (1R1) brightness temperature (K), contour lines depict topography.

clouds were simulated by both ARPS and LDAS-A models over the Gaize station and to the north, whereas MTSAT/1R1 showed high cloud activity to the south [Fig. 8(a), (b), and (d)]. As shown in Fig. 8(c), the CALDAS simulation improved most of the model's very active clouds to the north and better represented cloud spatial distribution to the south as observed by the MTSAT/1R1 channel. Although CALDAS significantly reduced most of the cloud activity to the north, by one and half hours after the assimilation, the number of cloud cells increased in that region in contrast to results immediately after the assimilation [Fig. 5(c)]. The increments in cloud activity to the north in the CALDAS simulation at 0900 UTC could be due to modeled dynamic (wind) fields, particularly the existence of dynamic convergence in this region due to inaccurately predicted clouds before the assimilation. In the cloud microphysics data assimilation, wind fields were not modified in accordance to the assimilated moisture fields, and the same conditions were kept before and after the assimilation. Therefore, the presence of dynamic convergence could have encouraged cloud formation to the north. On the other hand, improvements in cloud activity simulated by CALDAS are noteworthy, particularly over the Gaize station and its surrounding areas. The removal of model-misrepresented cloud condensate in CALDAS contributed significantly to positive spatial correlations after assimilation as shown in Fig. 7, whereas the other two models showed negative correlations with MTSAT/1R1 observations. In the south, cloud activity from 0730 UTC to 0900 UTC also increased and compared well with MTSAT/1R1 observations in the CALDAS simulation. The replacement of air gases with cloud condensate (has lower density than air gases) can induce buoyancy or upward motion. This buoyancy could have further enhanced cloud activity over the southern region. As a result, the CALDAS simulated higher cloud activity, and produced a high correlation with MTSAT observations at 0900 UTC.

C. Validations of Model Surface Soil Moisture and Rainfall With In-Situ Observations

In the following sections, CALDAS simulations are investigated from perspective of land surface. Fig. 9(a1) and (b1) compare surface soil moisture contents simulated by APRS, LDAS-A, and CALDAS with surface soil moisture content recorded at the Gaize station. The sudden drops in the soil moisture trend correspond to the assimilation response, and sudden peaks indicate rainfall occurrences. The figures are separated into several periods, and separation lines (dashed vertical lines) indicate the time when AMSR-E observations

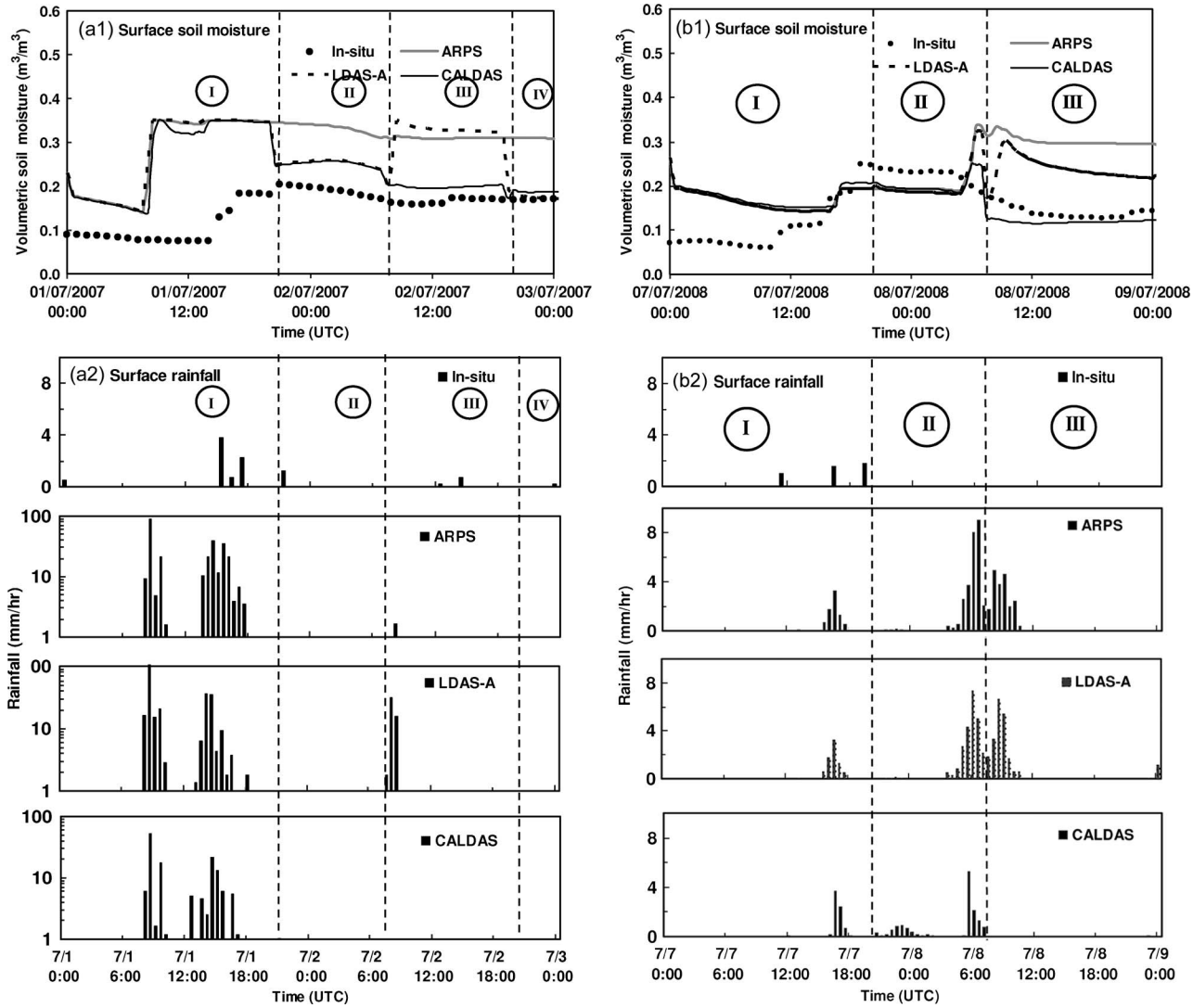


Fig. 9. Comparison of observed and simulated surface variables by ARPS, LDAS-A, and CALDAS at the Gaize station; (a1) and (b1) for surface volumetric soil moisture (m^3/m^3) at 3 cm from 1st to 2nd July 2007 and from 7th to 8th July 2008, respectively, (a2) and (b2) for surface rainfall (mm/h) from 1st to 2nd July 2007 and from 7th to 8th July 2008, respectively. Dashed vertical lines on the figures depicted the time of assimilations.

were assimilated in both LDAS-A and CALDAS models. As shown in Fig. 9(a1), period I corresponds to a free model run where no assimilation was performed. During this period the atmospheric model predicted a rainfall event that exceeded 100 mm/h, which is unrealistic in this region. All soil moisture simulations were affected severely by this rainfall and showed saturated moisture states. In the ARPS model, this rainfall induced biases that persisted until the end of model execution. This situation causes biases in partitioning of turbulence fluxes and land–atmosphere interaction processes, which could then enforce negative impacts on model forecast and water and energy budgets evolution.

At 2030 UTC on 1st July 2007 (boundary between period I and II) when satellite observations were available, LDAS-A and CALDAS significantly improved rainfall-induced biases and reinitialized the land surface conditions. During period II, model soil moisture simulations by both LDAS-A and CALDAS compared well with observed soil moisture trends. When the next observations were available, the assimilation by LDAS-A brought the soil moisture close to the observed value (boundary between period II and III). However, immediately after the second land data assimilation (during period III), LDAS-A surface conditions were affected severely by instantaneous model rainfall events [Fig. 9(a1)]. To remove biases in surface conditions, LDAS-A model requires the next satellite observations. As expected, the next observations (at the boundary between period III and IV) realigned the simulated soil moisture that compared well with observed soil moisture. Similarly, for 2008 as shown in Fig. 9(b1), during period III, LDAS-A surface conditions were also affected by the model rainfall event. In this case, the model predicted rainfall events before the land data assimilation was executed, and the rainfall event continued after the assimilation. Conversely, as shown in Fig. 9(a2) and (b2), the introduction of atmospheric data assimilation in CALDAS immediately removed the predicted rainfall events that occurred during or immediately after the land data assimilation. The successful removal of inaccurate rainfall events effectively improved the simulated surface soil

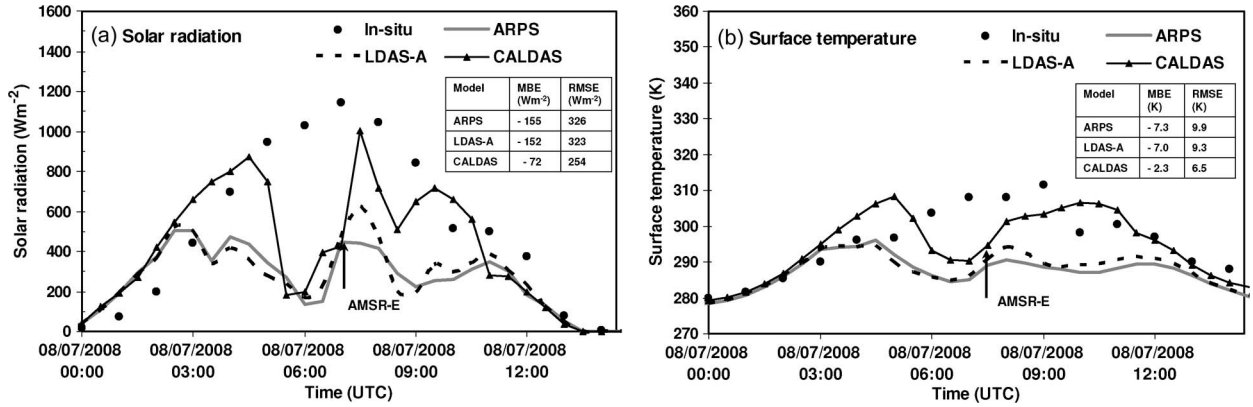


Fig. 10. Comparison of observed and simulated surface variables by ARPS, LDAS-A, and CALDAS at Gaize station on 08th July 2008; (a) solar radiation (Wm^{-2}), and (b) surface temperature (K).

moisture content. As a result the surface soil moisture contents simulated by CALDAS in both cases [Fig. 9(a1) and (b1)] compared well with observed soil moisture contents and the trends at the Gaize station, particularly during period III. Therefore, the improved land surface conditions will guide LSM to represent land surface processes accurately during model forecasts.

D. Evaluation of Model Forcing Data With In-Situ Observations

Further investigations were conducted to assess the effect of improved cloud distribution on model simulated solar radiation, which is another major forcing for accurate land surface modeling. Fig. 10(a) shows that solar radiation simulated by CALDAS compares well with the observed radiation and has smaller MBE and RMSE, whereas ARPS and LDAS-A showed much lower solar radiation compared with observation and have higher MBE and RMSE. The improvements in CALDAS simulation can be separated into two time slots: 1) improvements from 0300 UTC to 0500 UTC resulting from the previous nighttime assimilation of AMSR-E observations (not shown); and 2) improvements after the second assimilation, which took place around 0730 UTC. Hence, the CALDAS improved cloud representation over the Gaize station, solar radiation simulated by CALDAS peaked immediately after the assimilation and showed reasonable agreement with observations. One hour after the assimilation, a decrease in solar radiation was observed in the CALDAS simulation because of uncertainties in estimated cloud distribution closer to Gaize station [Fig. 5(c)]. However, two hours after the assimilation it again increased and followed the observed trends, whereas in the ARPS and LDAS-A models, because of the presence of clouds, both models underpredicted the solar radiation compared with observations significantly. The improvement in soil moisture and solar radiation resulted in improvements in CALDAS-simulated surface temperature with MBE equals to -2.3 K, RMSE equals to 6.5 K [Fig. 10(b)]. The other two simulations by ARPS and LDAS-A resulted to underestimations (MBE ~ -7 K and RMSE $\sim$9 K) due to lower solar radiation input and higher soil moisture content in the Gaize model grid cell.

E. Evaluation of Land-Atmosphere Interactions

Improving land surface conditions (i.e., soil moisture and surface temperature) and land surface forcing (precipitation and radiation) will eventually improve the estimation of turbulent heat fluxes and convective processes, which will in turn improve land–atmosphere interactions. To verify the mechanism of land–atmosphere interactions, radiosonde soundings were compared with the simulated soundings. Fig. 11(a) compares observed potential temperature soundings with simulated soundings from ARPS, LDAS-A, and CALDAS simulations at 0500 UTC on 7th July 2008, which corresponds to 1100 Local Time (LT) at Gaize station. During this time, the land surface was heated up by the solar radiation and turbulent mixing started to occur closer to the land surface, as shown by observed soundings. This observed trend was well captured by CALDAS potential temperature soundings, and compared well with observed soundings. The ARPS and LDAS-A models underestimated potential temperature soundings from the surface to $\sim$540 hPa, and the difference between simulated and observed potential temperature close to the surface was $\sim$4 K. At this time, the improvement in land surface conditions alone did not result to any improvement in potential temperature profile. The next observed sounding data was available after 6 h, at 1100 UTC (1700 LT). At this time, as shown in Fig. 11(b), the observed potential temperature soundings indicated a well-mixed turbulent boundary layer with a well-defined top at $\sim$480 hPa. During this period, highly active interactions took place between land and atmosphere. However, high moisture contents and overcast sky conditions (very week solar radiation) in ARPS and LDAS-A (slightly better than ARPS near the surface) models estimated very weak turbulent fluxes as the lower boundary conditions, which might have suppressed land–atmosphere interactions in both models. The actual well-mixed turbulent boundary layer was captured well by the CALDAS simulation. The simulated potential temperature soundings nearly mirrored observed soundings and has less biases (MBE $= -0.58$ K, RMSE $= 1.06$ K) compared with other two simulations. The improvement extended from the surface to the middle atmospheric layers ($\sim$480 hPa) in the CALDAS simulation. These results indicated that CALDAS represented well the actual mechanisms of land–atmosphere interactions.

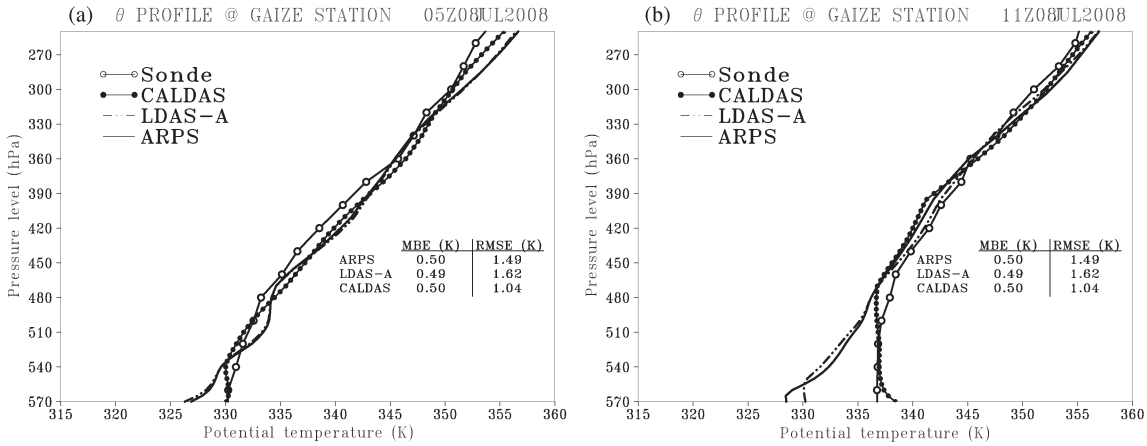


Fig. 11. Comparison of observed potential temperature (K) soundings with ARPS, LDAS-A and CALDAS model soundings at (a) 0500 UTC, and (b) 1100 UTC on 08th July 2008.

VI. Summary and Conclusion

Accurate representation of land surface conditions will improve the predictability of weather and climate models. To better represent land surface heterogeneities in NWP applications, land data assimilation was coupled with a mesoscale model to directly assimilate AMSR-E lower-frequency brightness temperature observations. However, because of limited observations by AMSR-E (bi-daily), reinitialized land surface conditions in land–atmosphere coupled models often suffer from erroneously predicted atmospheric forcing, particularly precipitation and solar radiation. Although severely affected land surface conditions can be improved in the next satellite overpass, because of the non-existence of future satellite data, this situation is crucial in the case of numerical weather forecasting.

To overcome this operational pitfall and to enhance the predictability of a mesoscale model, we developed a system that coupled atmosphere and land data assimilation systems with a mesoscale model (CALDAS). CALDAS merged information on AMSR-E's lower frequency observations with that of higher frequencies, and therefore facilitated passive microwave remote sensing to obtain atmospheric information over highly emissive and varying land surfaces. Numerical experiments in the Tibetan Plateau showed that CALDAS reasonably retrieved land and atmospheric moisture fields from AMSR-E multifrequency observations. CALDAS improved the spatial distribution of predicted clouds effectively, and predicted clouds correlated highly with MTSAT/1R1 observations even 3.5 h after the assimilation. During this period, the other two models (ARPS and LDAS-A) showed high negative correlations with the MTSAT/1R1 cloud observations. Improvements in cloud simulations resulted in improvements in model forcings (i.e., precipitation and solar radiation) to the land surface model. With improved atmospheric forcing, the land surface model better estimated surface soil moisture and surface temperature, and maintained them closer to observed conditions during model forecast. As a result, CALDAS captured existing land surface processes and the feedback mechanism of land–atmosphere interactions, as confirmed by surface observations and radiosonde soundings.

These results are encouraging in terms of producing reliable regional water and energy budgets, particularly in ungauged regions, because the system requires only satellite data as inputs. Though the results showed that the investigated system is capable of improving land surface variables, atmospheric moisture information, and land–atmosphere interactions, further improvements in system performance and applicability are needed. 1) Dynamic field adjustments are crucial in cloud microphysics data assimilation to obtain reliable model forecast, which was not addressed in this study. 2) Cloud parameters (i.e., cloud profiles, heights, and integrated condensates) were predefined in CALDAS. Use of other satellite observations, e.g., AMSU, will provide more detailed atmospheric information that can overcome these assumptions. 3) Integration of similar satellite information (e.g., Tropical Rainfall Measuring Mission Microwave Imager), Global Precipitation Measurement Microwave Imager, and Global Change Observation Mission is feasible in the system. This will reduce the AMSR-E temporal sampling problem and should enhance model predictability. 4) Consideration of remotely sensed surface temperature into the system also has the potential to improve land and atmospheric moisture estimation. 5) Further validation of the results, such as cloud solid and liquid phases, will have to wait for accurate and unambiguous 3-D cloud observations, which will be the direction of future intensive observations.

Appendix

A. RTM Model for Soil

The calculation of surface reflectivity depends critically on surface moisture conditions and degree of roughness. When the land surface is smooth, Fresnel power reflectivity is used. The horizontally (R_h) and vertically (R_v) polarized Fresnel power reflectivity is calculated as

$$R_h = \left| \frac{\cos\theta - \sqrt{\epsilon_r - \sin^2\theta}}{\cos\theta + \sqrt{\epsilon_r - \sin^2\theta}} \right|^2 \tag{5}$$

$$R_v = \left| \frac{\epsilon_r \cos\theta - \sqrt{\epsilon_r - \sin^2\theta}}{\epsilon_r \cos\theta + \sqrt{\epsilon_r - \sin^2\theta}} \right|^2 . \tag{6}$$

Here, ϵ_r is the dielectric constant of the soil-water mixture, which is given by [8]

$$\epsilon_r = \left[1 + (1-\omega_s)\left(\epsilon_s^{\alpha} - 1\right) + w^{\beta}\epsilon_{f_\omega}^{\alpha} - w\right]^{\frac{1}{\alpha}} \tag{7}$$

where ω_s is soil porosity, w is soil water content, ϵ_s is the dry soil dielectric constant, ϵ_{f_ω} is the dielectric constant of free water, $\alpha = 0.65$, and β is a coefficient dependent on soil texture. β is calculated from the soil texture as [51]

$$\beta = 1.09 - 0.0011 * \%sand + 0.0018 * \%clay. \tag{8}$$

Further information can be found in [46].

B. EnKF Formulation for Soil

Consider $X = [w_1, w_2, w_3]^T$ as a state variable and the first estimate, where w_1, w_2 and w_3 are the soil moisture contents of the surface, root, and deep soil layers, respectively. The first estimate is used to create an ensemble of size (N) by adding pseudorandom noise with known statistics. By dropping the time notation, each member of state variable X_i is given by

$$X_i = \bar{X} + e_i \qquad e_i(i = 1, 2 \ldots N) \sim N(0, P) \tag{9}$$

where e_i is the random error vector of each member obtained from a multivariate Gaussian distribution with zero mean and error covariance matrix P, and $\bar{X}$ is the expectation of the first estimate X. In the forecast step, the forecast state member X_i^f is determined from the nearest analysis state member X_i^a according to

$$X_i^f = M\left(X_i^a\right) + u_i \qquad u_i \sim N(0, Q) \tag{10}$$

where M is the model operator and u_i is the model error vector of each member, obtained from a multivariate Gaussian distribution with zero mean and error covariance matrix Q.

In the analysis step, the AMSR-E observation data are perturbed by adding a random observation error and each member of the analyzed state variable X_i^a is updated as

$$X_i^a = X_i^f + K\left[(Y_o + v_i) - H\left(X_i^f\right)\right] \qquad v_i \sim N(0, R) \tag{11}$$

where K is a Kalman gain matrix, H is the observation operator, Y_o is the observation, R is the observation error covariance, and v_i is a random error vector of the observation with zero mean and covariance matrix R.

C. RTM for Atmosphere

The radiative transfer equation for polarized waves in a plane-parallel and horizontally homogeneous scattering layer that composed of spherical particles can be expressed by [50]

$$\mu \frac{d}{d\tau}\begin{bmatrix} I_V(\tau,\mu) \\ I_H(\tau,\mu) \end{bmatrix} = \begin{bmatrix} I_V(\tau,\mu) \\ I_H(\tau,\mu) \end{bmatrix} - \frac{\omega_0}{2}\int_{-1}^{1}\begin{bmatrix} P_{VV} & P_{VH} \\ P_{HV} & P_{HH} \end{bmatrix} \times \begin{bmatrix} I_V(\tau,\mu) \\ I_H(\tau,\mu) \end{bmatrix} d\mu' - (1-\omega_0)B(\tau)\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{12}$$

where $I_p(\tau, \mu)$ is the radiance at optical depth τ in direction μ (the cosine of zenith angle) for polarization P, ω_0 is the single scattering albedo, $B\tau$ is the Plank function at τ, and $P_{i,j}$ $(i, j = H$ or V) is the scattering phase function.

The exact solution of $I_p(\tau, \mu)$ can be obtained by solving (12) using the discrete ordinate method with sufficient (e.g., 32) streams. However, this is computationally intensive because 1) the phase functions are expressed in summations of infinite terms and need to be calculated for many directions, and 2) the eigenvalue problem for a general solution of (12) needs to be solved numerically for streams larger than 4. To reduce the computation burden, the four-stream fast model was selected [36] and this model introduces two assumptions: 1) no cross polarizations exist (i.e., $P_{HV} = P_{VH} = 0$); and 2) the scattering phase matrix can be approximated by the Henyey–Greenstein scattering phase function [23].

ACKNOWLEDGMENT

The authors would like to thank K. Tamagawa and T. Ohta, Data Archiving Managers of the Coordinated Energy and Water Cycle Observation Project, The University of Tokyo, for their continued support in the model validations; Japan Aerospace Exploration Agency for providing the AMSR-E brightness temperature observations; and Kochi University for providing the MTSAT/1R data set. The surface and radiosonde data were provided by the Japan–China JAICA project. This research was carried out as part of the Research Program on Climate Change Adaptation (RECCA) and verification experiment for AMSR.AMSR-E and AMSR-2 (GCOM-W).

REFERENCES

[1] A. C. M. Beljaars, P. Viterbo, M. J. Miller, and A. K. Betts, "The anomalous rainfall over the United States during 1993: Sensitivity to land surface parameterization and soil moisture anomalies," *Mon. Weather Rev.*, vol. 124, no. 3, pp. 362–383, 1996.

[2] L. P. Bobylev, E. V. Zabolotskikh, L. M. Mitnik, and M. L. Mitnik, "Atmospheric water vapor and cloud liquid water retrieval over the arctic ocean using satellite passive microwave sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 283–294, Jan. 2010.

[3] S. Bonafoni, V. Mattioli, P. Basili, P. Ciotti, and N. Pierdicca, "Satellite-based retrieval of precipitable water vapor over land by using a neural network approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3236–3248, Sep. 2011.

[4] S. Boussetta, T. Koike, K. Yang, T. Graf, and M. Pathmathevan, "Development of a coupled land-atmosphere satellite data assimilation system for improved local atmospheric simulations," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 720–734, 2008.

[5] K. Chen, T. D. Wu, L. Tsang, Q. Li, J. C. Shi, and A. K. Fung, "Emission of rough surfaces calculated by the integral equation method with comparison to a three dimensional moment method simulation," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 90–101, Jan. 2003.

[6] M. N. Deeter, "A new satellite retrieval method for precipitable water vapor over land and ocean," *Geophys. Res. Lett.*, vol. 34, pp. L02815-1–L02815-5, Jan. 2007, doi:10.1029/2006GL028019.

[7] M. N. Deeter and J. Vivekanandan, "New dual-frequency microwave technique for retrieving liquid water path over land," *J. Geophys. Res.*, vol. 111, pp. D15209-1–D15209-11, Aug. 2006, doi:10.1029/2005JD006784.

[8] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes, "Microwave dielectric behavior of wet soil—Part II: Dielectric mixing models," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-23, no. 1, pp. 35–46, Jan. 1985.

[9] C. S. Draper, J.-F. Mahfouf, and J. P. Walker, "An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme," *J. Geophys. Res.*, vol. 114, pp. D20104-1–D20104-13, Oct. 2009, doi:10.1029/2008JD011650.

[10] M. Drusch, "Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set," *J. Geophys. Res.*, vol. 112, pp. D03102-1–D03102-14, Feb. 2007.

[11] M. Drusch, E. F. Wood, and T. Jackson, "Vegetative and atmospheric corrections for the soil moisture retrieval from passive microwave remote sensing data: Results from the Southern Great Plains Hydrology Experiment 1997," *J. Hydrometeorol.*, vol. 2, no. 2, pp. 181–192, Apr. 2001.

[12] Q. Duan, S. Sorooshian, and V. Gupta, "Effective and efficient global optimization for conceptual rainfall-runoff models," *Water Resour. Res.*, vol. 28, no. 4, pp. 1015–1031, 1992.

[13] D. Entekhabi, E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, N. Martin, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, and J. Van Zyl, "The Soil Moisture Active and Passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.

[14] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dyn.*, vol. 53, no. 4, pp. 343–367, 2003.

[15] G. Evensen, Ed., *Data Assimilation, The Ensemble Kalman Filter*, 2nd ed. New York: Springer-Verlag, 2009.

[16] B. Federer and A. Waldvogel, "Hail and raindrop size distributions from a Swiss multicell storm," *J. Appl. Meteorol.*, vol. 14, no. 1, pp. 91–97, Feb. 1975.

[17] H. Fujii and T. Koike, "Development of a TRMM/TMI algorithm for precipitation in the Tibetan Plateau by considering effects of land surface emissivity," *J. Meteorol. Soc. Jpn.*, vol. 79, no. 1B, pp. 475–483, 2001.

[18] H. Fujii, T. Koike, and K. Imaoka, "Improvement of the AMSR—E Algorithm for soil moisture estimation by Introducing a fractional vegetation coverage data set derived from MODIS data," *J. Remote Sens. Soc. Jpn.*, vol. 29, no. 1, pp. 282–292, Jan. 2009.

[19] K. S. S. Gunn and J. S. Marshall, "The distribution with size of aggregate snowflakes," *J. Meteorol.*, vol. 15, no. 5, pp. 452–461, Oct. 1958.

[20] R. Guo, G. Zhang, Z. Li, and L. Zhang, "ATOVS data assimilation in two different heavy rainfalls," in *Proc. ICIST*, 2011, pp. 298–305.

[21] Y.-R. Guo, Y.-H. Kuo, J. Dudhia, D. Parsons, and C. Rocken, "Four-dimensional variational data assimilation of heterogeneous mesoscale observations for a strong convective case," *Mon. Weather Rev.*, vol. 128, no. 3, pp. 619–649, 2000.

[22] S.-Y. Ha, Y.-H. Kuo, Y.-R. Guo, and G.-H. Lim, "Variational assimilation of slant-path wet delay measurements from a hypothetical grond-based gps network. Part 1: Comparison with precipitable water assimilation," *Mon. Weather Rev.*, vol. 131, no. 11, pp. 2635–2655, 2003.

[23] L. C. Henyey and J. L. Greenstein, "Diffuse radiation in the galaxy," *Astrophys. J.*, vol. 93, pp. 70–83, Jan. 1941.

[24] C. L. Huang, X. Li, and L. Lu, "Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 889–900, 2008.

[25] D. Huang, Y. Liu, and W. Wiscombe, "Determination of cloud liquid water distribution using 3D cloud tomography," *J. Geophys. Res.*, vol. 113, pp. D13201-1–D13201-13, Jul. 2008, doi:10.1029/2007JD009133.

[26] J. Huang, P. Minnis, B. Lin, Y. Yi, T.-F. Fan, S. Sun-Mack, and J. K. Ayers, "Determination of ice water path in ice-over-water cloud systems using combined MODIS and AMSR-E measurements," *Geophys. Res. Lett.*, vol. 33, pp. L21801-1–L21801-5, Nov. 2006, doi:10.1029/2006GL027038.

[27] T. Iwabuchi, Y.-R. Guo, C. Rocken, T. Van Hove, and Y.-H. Kuo, "Impact of ground-based GPS retrievals on moisture field and rainfall forecast in WRF/3DVAR," in *Proc. 15th WRF/MM5 Users' Workshop*, Boulder, CO, Jun. 27–30, 2005, pp. 351–360.

[28] Y. P. Kerr, P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.-J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. E. Juglea, M. R. Drinkwater, A. Hahne, M. Martín-Neira, and S. Mecklenburg, "The SMOS mission: New tool for monitoring key elements of the global water cycle," *Proc. IEEE*, vol. 98, no. 5, pp. 666–687, May 2010.

[29] J. Key and R. G. Barry, "Cloud cover analysis with arctic AVHRR data. 1. Cloud detection," *J. Geophys. Res.*, vol. 94, no. D15, pp. 18 521–18 535, 1989.

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

[31] S. V. Kumar, R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, "Role of subsurface physics in the assimilation of surface soil moisture observations," *J. Hydrometeorol.*, vol. 10, no. 6, pp. 1534–1547, 2009.

[32] Y.-H. Kuo, X. Kuo, Y. Zuo, and Y.-R. Guo, "Variational assimilation of precipitable water using a nonhydrostatic mesoscle adjoint model. Part 1: Moisture retrieval and sensitivity experiments," *Mon. Weather Rev.*, vol. 124, pp. 122–147, 1996.

[33] D. N. Kuria, T. Koike, H. Lu, T. Tsutsui, and T. Graf, "Field-supported verification and improvement of a passive microwave surface emission model for rough, bare, and wet soil surfaces by incorporating shadowing effects," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1207–1216, May 2007.

[34] S. Lee, B. H. Kahn, and J. Teixeira, "Characterization of cloud liquid water content distributions from CloudSat," *J. Geophys. Res.*, vol. 115, pp. D20203-1–D20203-13, Oct. 2010, doi:10.1029/2009JD013272.

[35] Y. L. Lin, R. D. Farley, and H. D. Orville, "Bulk parameterization of the snow field in a cloud model," *J. Clim. Appl. Meteorol.*, vol. 22, no. 6, pp. 1065–1089, Jun. 1983.

[36] G. Liu, "A fast and accurate model for microwave radiance calculations," *J. Meteorol. Soc. Jpn.*, vol. 76, no. 2, pp. 335–343, 1998.

[37] Y. Liu, P. H. Daum, S. K. Chai, and F. Liu, "Cloud parameterizations, cloud physics, and their connections: An overview," in *Recent Research Developments in Geophysics*, S. G. Pandalai, Ed. Trivandrum, India: Research Signpost, 2002, pp. 119–142.

[38] H. Lu, T. Koike, T. Ohta, H. Fujii, and H. Tsutsui, "Improving the AMSR-E soil moisture algorithm of the University of Tokyo through field experiments and parameters optimization," in *Proc. IEEE IGARSS*, 2008, vol. 2, pp. 65–68.

[39] H. Lu, T. Koike *et al.*, "Development of a physically-based soil moisture retrieval algorithm for spaceborne passive microwave and its application to AMSR-E," *J. Remote Sens. Soc. Jpn.*, vol. 29, no. 1, pp. 253–262, 2009.

[40] J. S. Marshall and W. M. Palmer, "The distribution of raindrops with size," *J. Atmos. Sci.*, vol. 5, no. 4, pp. 165–166, Aug. 1948.

[41] C. R. Mirza, T. Koike, K. Yang, and T. Graf, "Retrieval of atmospheric integrated water vapor and cloud liquid water content over the ocean from satellite data using the 1-D-Var Ice Cloud Microphysics Data Assimilation System (IMDAS)," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 119–129, Jan. 2008.

[42] K. Mitchell, C. Marshall, D. Lohmann, M. Ek, Y. Lin, P. Grunmann, P. Houser, E. Wood, J. Schaake, D. Lettenmaier, D. Tarpley, W. Higgins, R. Pinker, A. Robock, B. Cosgrove, J. Entin, and Q. Duan, "The collaborative GCIP Land Data Assimilation (LDAS) project and supportive NCEP uncoupled land-surface modeling initiatives," in *Proc. 15th Amer. Meteorol. Soc. Hydrol. Conf.*, Long Beach, CA, Jan. 9–14, 2000, pp. 1–4, Preprints.

[43] S. Nicholson, "Land surface processes and Sahel climate," *J. Hydrol.*, vol. 38, no. 1, pp. 117–139, 2000.

[44] C. O'Dell, F. Wentz, and R. Bennartz, "Cloud liquid water path from satellite-based passive microwave observations: A new climatology over the global oceans," *J. Clim.*, vol. 21, no. 8, pp. 1721–1739, 2008.

[45] J. S. Pal and E. A. B. Eltahir, "Teleconnections of soil moisture and rainfall during the 1993 Midwest summer flood," *Geophys. Res. Lett.*, vol. 29, no. 18, pp. 1865–1859, Sep. 2002.

[46] M. Rasmy, T. Koike, S. Boussetta, H. Lu, and X. Li, "Development of a satellite land data assimilation system coupled with a mesoscale model in the Tibetan Platea," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2847–2862, Aug. 2011.

[47] R. H. Reichle and D. B. McLaughlin, "Hydrologic data assimilation with the ensemble Kalman filter," *Mon. Weather Rev.*, vol. 130, no. 1, pp. 103–114, Jan. 2002.

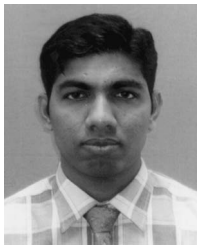
[48] K. Scipal, M. Drusch, and W. Wagner, "Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system," *Adv. Water Resour.*, vol. 31, no. 8, pp. 1101–1112, Aug. 2008.

[49] P. J. Sellers, S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, and G. J. Collatz, "A revised land surface parameterization (SiB2) for atmospheric GCMs, Part II: The generation of global fields of terrestrial biophysical parameters from satellite data," *J. Clim.*, vol. 9, no. 4, pp. 706–737, Apr. 1996.

[50] L. Tsang and J. A. Kong, "Theory of thermal microwave emission from a bounded medium containing spherical scatters," *J. Appl. Phys.*, vol. 48, no. 8, pp. 3593–3599, Aug. 1977.

[51] F. T. Ulaby, R. K. Moore, and A. K. Fung, Eds., *Microwave Remote Sensing: Active and Passive*, vol. III, *Volume Scattering and Emission Theory*. Dedham, MA: Artech House, 1986.

[52] S. S. Vaidya, "Simulation of weather systems over Indian region using mesoscale models," *Meteorol. Atmos. Phys.*, vol. 95, no. 1/2, pp. 15–26, Jan. 2007.
[53] J. Wang, W. B. Rossow, and Y. Zhang, "Cloud vertical structure and its variations from a 20 year global rawinsonde data set," *J. Clim.*, vol. 13, no. 17, pp. 3041–3056, Sep. 2000.
[54] F. J. Wentz and T. Meissner, "Algorithm theoretical basis document: AMSR-E ocean algorithms supplement 1," in *Proc. Remote Sens. Syst. Conf.*, Santa Rosa, CA, 2007.
[55] X. Xiangde, Z. Renhe, T. Koike, L. Chungu, and S. Xiaohui, "A new integrated observational system over the Tibetan plateau," *Bull. Amer. Meteorol. Soc.*, vol. 89, no. 10, pp. 1492–1496, Oct. 2008.
[56] M. Xue, K. K. Droegemeier, and V. Wong, "The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification," *Meteorol. Atmos. Phys.*, vol. 75, no. 3/4, pp. 161–193, 2000.
[57] M. Xue, K. K. Droegemeier, V. Wong *et al.*, "The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications," *Meteorol. Atmos. Phys.*, vol. 76, no. 3/4, pp. 143–165, 2001.



Mohamed Rasmy received the B.Sc.Eng. (Hons.) degree from the University of Peradeniya, Peradeniya, Sri Lanka, in 2003, and the M.Sc., and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 2006 and 2010, respectively.

He is currently a Research Associate with the River and Environmental Engineering Laboratory, Department of Civil Engineering, University of Tokyo. His research interests are water resources, microwave remote sensing, land and atmospheric data assimilations, and numerical weather prediction.

Toshio Koike received the B.Eng., M.Eng., and D.Eng. degrees from the University of Tokyo, Tokyo, Japan, in 1980, 1982, and 1985, respectively.

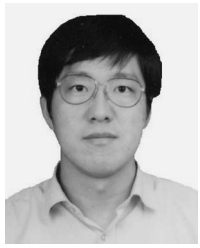
He was a Research Associate at the University of Tokyo in 1985 and was appointed as an Assistant Professor at the same university in 1986. He was also appointed an Associate Professor at Nagaoka University of Technology, Nagaoka, Japan, in 1992. He has been a Professor at the River and Environmental Engineering Laboratory, Department of Civil Engineering, University of Tokyo since 2000. He is the Lead Scientist of the Coordinated Enhanced Observing Period project. His research interests are in hydrology, water resources, satellite remote sensing, climate change, and Asian monsoons.

David Kuria received the B.Sc (Surveying) degree with honors from the University of Nairobi, Nairobi, Kenya, in 1998, the M.Sc. degree (photogrammetry and geoinformatics) from the Stuttgart University of Applied Sciences, Stuttgart, Germany, and Ph.D. degree from the University of Tokyo, Tokyo, Japan, in 2003 and 2007, respectively.

He is currently a Lecturer in the Department of Geomatic Engineering and Geospatial Information Systems of the Jomo Kenyatta University of Agriculture and Technology, Kenya. His reserarch interests are in microwave remote sensing, webmapping, climate research, and geospatial application development.

Cyrus Raza Mirza received the B.Eng. (with honors) degree from University of Engineering and Technology, Taxila, Pakistan, in 1997, the M.Eng. degree from the University of Engineering and Technology, Lahore, Pakistan, in 2000, and the D.Eng. degree from the University of Tokyo, Tokyo, Japan, in 2005.

His research interests are cloud microphysics, data assimilation, atmospheric modeling, ocean-atmospheric interactions, prediction of precipitation, and prediction of ungauged basins.



Xin Li received the B.Sc degree from Nanjing University, Nanjing, China, in 1992, and Ph.D. degree from the Chinese Academy of Sciences, Lanzhou, China, in 1998.

He is currently a Professor at the Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences. His primary research interests include land data assimilation, application of remote sensing and Geographic Information System in hydrology and cryosphere science, and integrated watershed modeling.

Kun Yang received the bachelor and master degrees from Tsinghua University, Beijing, China, in 1994 and 1997, respectively, and the Ph.D. degree from the University of Tokyo, Tokyo, Japan, in 2000.

He is with Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China. He is currently the Chair of GEWEX/CEOP Water and Energy Budget Study Project and the Vice Director of the Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Chinese Academy of Sciences. His research interests are land surface processes, observation, modeling and data assimilation, land-atmosphere interactions, regional atmospheric modeling, and radiation modeling.