Improving Predictions of Water and Heat Fluxes by Assimilating MODIS Land Surface Temperature Products into the Common Land Model

TONGREN XU AND SHAOMIN LIU

State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing, China

SHUNLIN LIANG

Department of Geography, University of Maryland, College Park, College Park, Maryland

JUN QIN

Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

(Manuscript received 1 April 2010, in final form 19 October 2010)

ABSTRACT

Four data assimilation scheme combinations derived from two strategies and two optimization algorithms [the ensemble Kalman filter (EnKF) and the shuffled complex evolution method developed at The University of Arizona (SCE-UA)] are developed based on the Common Land Model (CLM) to improve predictions of water and heat fluxes. The first strategy is constructed through adjusting the soil temperature, while the second strategy adjusts the soil moisture. Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) products are compared with ground-measured surface temperature, and assimilated into the CLM. The relationship equation between the MODIS LST products and CLM surface temperature is taken as the observation operator and the root-mean-square error (RMSE) is applied as the observation error. The assimilation results are validated by measurements from six observation sites located in Germany, the United States, and China. Results indicate that the developed data assimilation schemes can improve estimates of water and heat fluxes. Overall, strategy 2 is superior to strategy 1 when using the same optimization algorithm. The EnKF algorithm performs slightly better than the SCE-UA algorithm when using the same strategy. Strategy 2 combined with the EnKF algorithm performs best for water and heat fluxes, and the reductions in the RMSE are found to be 24.0 and 15.2 W m^{-2} for sensible and latent heat fluxes, respectively. The joint assimilation of the MODIS LST and soil moisture observations can produce better results for strategy 2 with the SCE-UA. Since preprocessing model parameters are used in this study, the uncertainties in the model parameters may have resulted in suboptimal assimilation results. Therefore, model calibrations should be conducted in the future.

1. Introduction

The accurate estimation of water and heat fluxes is crucial to climate change research, water resource planning and management, and water-saving agriculture applications. A number of methods have been developed to estimate water and heat fluxes. Ground measurements are relatively accurate at the patch scale. However, the monitoring network for ground measurements is not sufficient for global coverage, and there is a need to upscale observations from the patch to the regional scale. The remote sensing method can acquire land surface information accurately at the regional scale, which creates new opportunities for monitoring water and heat fluxes (Bastiaanssen et al. 1998; Su 2002; Liu et al. 2007). Since remote sensing data are instantaneous, estimating the daily, monthly, and annual flux values may cause errors. Over the last 20 yr, evolving land surface models have been able to produce continuous temporal and spatial variations of water and heat fluxes (Dickinson et al. 1986; Sellers et al. 1996; Dai et al. 2001, 2003). This has had a significant impact on our understanding of regional energy and water balances. However, the land model

Corresponding author address: Dr. Shaomin Liu, State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, No. 19, Xinjiekouwai St., Beijing 100875, China. E-mail: smliu@bnu.edu.cn

DOI: 10.1175/2010JHM1300.1

output is strongly contaminated by uncertainties within the model parameters, model structures, and forcing data; this has adversely affected the development and application of land surface models.

Considering these issues, developing a series of techniques combining land surface models with different types of observations is necessary. Data assimilation is one such technique and is one of the most advanced approaches to improving land surface model predictions (Margulis et al. 2002). Recently, there has been significant renewed interest in data assimilation techniques (Crow and Wood 2003; Liang 2004; Liang and Qin 2008). A number of past studies have developed several techniques to assimilate microwave brightness temperature into land surface and hydrological models with the goal of improving soil moisture estimates (Galantowiez et al. 1999; Reichle et al. 2002; Crow and Wood 2003; Yang et al. 2007; Qin et al. 2009).

Surface temperature has been established to be a key factor for water and energy balance studies. Surface temperature influences the partitioning of incoming radiant energy into ground, as well as sensible and latent heat fluxes. Overestimating the surface temperature results in excessive estimates of sensible and soil heat fluxes, which can impact the boundary layer development and atmospheric dynamics. Underestimating the results has corresponding adverse effects. Although latent heat flux is mainly affected by soil moisture, surface temperature is an important parameter too. Therefore, researchers have conducted numerous experiments to assimilate surface temperature observations into land surface models. Kumar and Kaleita (2003) and Huang et al. (2008) illustrated that remotely sensed radiometric temperatures can enhance the prediction of the soil temperature profiles. Crow et al. (2008) assimilated a thermal remote-sensing-based soil moisture proxy into a water balance model and improved the prediction of the root zone soil moisture. To obtain surface energy balance components, a small number of studies have developed a variational method of combining relatively simple models with surface temperature observations (Boni et al. 2001; Caparrini et al. 2004). Pipunic et al. (2008) conducted a comparative synthetic study to test the potential of assimilated remotely sensed data in improving water and heat flux predictions.

The purpose of most data assimilation experiments is to correct soil moisture and soil temperature profiles. Only a few studies have focused on obtaining accurate water and heat flux predictions by assimilating surface temperatures like those introduced above. However, these results are largely unverified, and more research is needed to determine if, and how well, assimilation of Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) products can improve the predictions of water and heat fluxes from land surface models. There is also a need to know whether the assimilation results are distinct by adjusting disparate state variables with different optimization algorithms. This study developed two data assimilation strategies that combine the MODIS LST products with the Common Land Model (CLM). Strategy 1 improves water and heat flux estimates by adjusting soil temperature, while strategy 2 achieves this goal by adjusting soil moisture. Two optimization algorithms, namely the ensemble Kalman filter (EnKF) and the shuffled complex evolution method developed at The University of Arizona (SCE-UA), were applied in the two strategies. Thus, four data assimilation scheme combinations derived from the two strategies and two optimization algorithms were developed. The assimilation results were validated at six observation sites with different land cover types (grassland, cropland, and orchard) located in Germany, the United States, and China.

This paper is organized as follows. Section 2 introduces the model operator, observation operator, and the two data assimilation strategies and two optimization algorithms. Section 3 discusses the experiment, including the description of observation sites, MODIS data, and model input data. Section 4 presents the results, including (i) the determination of the model and observation errors and (ii) the results of assimilating the MODIS LST products. Section 5 provides the summary and discussions.

2. Methodology

The developed data assimilation system consists of a model operator for calculating the diurnal variations in water and heat fluxes, an observation operator for predicting surface temperature from model state variables, and an optimization scheme for optimizing the state variables, all of which are described below.

a. Model operator

This study used the CLM as the model operator. The CLM was proposed by Dai et al. (2001, 2003), and it combines physical, hydrological, and biological processes that can simulate soil temperature, soil moisture, water and heat fluxes, and other variables. In the CLM, every surface grid cell is divided into up to 5 tiles with the following fractions: dominant vegetation, secondary vegetation, bare soil, wetland, and inland water. The CLM has 1 vegetation layer, 10 unevenly spaced vertical soil layers, and up to 5 snow layers (depending on snow depth). In addition, it includes a canopy photosynthesis-conductance model (two broadleaf models) used to

quantify the transfer of CO_2 and water vapor into and out of vegetation.

In the CLM, water and heat fluxes are calculated by solving the soil–vegetation–atmosphere energy balance equation. In the case of a nonvegetated surface, the energy balance equation is as follows:

$$R_{n,g}(T_g) - H_g(T_g) - LE_g(T_g) - G_g(T_g) = 0. \quad (1)$$

where $R_{n,g}$ is the net radiation at the soil surface (W m^{-2}); H_g, LE_g, and G_g are the sensible, latent, and soil heat fluxes, respectively, at the soil surface (W m^{-2}); and T_g is the temperature at the soil surface (K). The sensible and latent heat fluxes are dependent on the temperature and moisture at the soil surface and can be obtained as follows:

$$H_g = -\rho_{\text{atm}} C_p(\theta_{\text{atm}} - T_g)/r_{\text{ah}}. \quad (2)$$

Here, ρ_{atm} is the density of atmospheric air (kg m^{-3}), C_p is the specific heat of air at constant pressure (equal to 1012 J kg^{-1} K^{-1}), θ_{atm} is the air temperature (K), and r_{ah} is the aerodynamic resistance to heat transfer (s m^{-1}):

$$LE_g = -\rho_{\text{atm}}(q_{\text{atm}} - q_g)/r_{\text{aw}}, \quad (3)$$

$$q_g = h_r q_{\text{sat}}^{T_g}, \quad \text{and} \quad (4)$$

$$h_r = \exp\left(\frac{\psi g}{R_w T_g}\right), \quad (5)$$

where q_{atm} is the atmospheric specific humidity (kg kg^{-1}), q_g is the specific humidity of the soil surface (kg kg^{-1}), r_{aw} is the aerodynamic resistance to water vapor transfer (s m^{-1}), $q_{\text{sat}}^{T_g}$ is the saturated specific humidity at the soil temperature T_g (kg kg^{-1}), h_r is the relative humidity of the ground surface air (–), Ψ is the soil matrix potential (mm), g is the gravity constant (m s^{-2}), and R_w is the gas constant for water vapor (J kg^{-1} K^{-1}).

In the case of a vegetated surface, the sensible and latent heat fluxes are partitioned into canopy and soil fluxes that depend on the canopy and soil temperatures. The energy balance equation can be described as follows:

$$R_{n,c} - H_c(T_g, T_c) - LE_c(T_g, T_c) = 0 \quad \text{and} \quad (6)$$

$$R_{n,g} - H_g(T_g, T_c) - LE_g(T_g, T_c) - G_g(T_g, T_c) = 0, \quad (7)$$

where $R_{n,c}$ is the net radiation of the vegetation canopy (W m^{-2}); H_c and LE_c are the canopy sensible and latent heat fluxes, respectively (W m^{-2}); and T_c is the canopy temperature (K). Because of the coupling between the canopy temperature and canopy fluxes, the Newton–Raphson iteration is used to solve them simultaneously using the soil temperature from the previous time step. In this case, the sensible heat flux from the canopy and soil can be obtained as follows:

$$H_c = -\rho_{\text{atm}} C_p[c_a^h \theta_{\text{atm}} + c_g^h T_g - (c_a^h + c_g^h)T_c]\frac{c_v^h}{c_a^h + c_v^h + c_g^h}$$

and (8)

$$H_g = -\rho_{\text{atm}} C_p[c_a^h \theta_{\text{atm}} + c_v^h T_c - (c_a^h + c_v^h)T_g] \times \frac{c_g^h}{c_a^h + c_v^h + c_g^h}, \quad (9)$$

Here, c_a^h, c_g^h, and c_v^h are the sensible heat conductance from the canopy air to the atmosphere, ground to canopy air, and leaf surface to canopy air, respectively (m s^{-1}). The latent heat flux from the canopy and soil can be obtained as follows:

$$LE_c = -\rho_{\text{atm}}[c_a^w q_{\text{atm}} + c_g^w q_g - (c_a^w + c_g^w)q_{\text{sat}}^{T_c}]\frac{c_v^w}{c_a^w + c_v^w + c_g^w}$$

and (10)

$$LE_g = -\rho_{\text{atm}}[c_a^w q_{\text{atm}} + c_v^w q_{\text{sat}}^{T_c} - (c_a^w + c_v^w)q_g] \times \frac{c_g^w}{c_a^w + c_v^w + c_g^w}, \quad (11)$$

where $q_{\text{sat}}^{T_c}$ is the saturated specific humidity at the canopy temperature T_c (kg kg^{-1}) and c_a^w, c_g^w, and c_v^w are the latent heat conductance from the canopy air to the atmosphere, soil to canopy air, and leaf surface to canopy air, respectively (m s^{-1}).

b. Observation operator

In this study, the component temperature decomposition method was selected to relate surface temperatures from the CLM to the MODIS LST products (Anderson et al. 2005):

$$T_{\text{RAD}}^4(\theta) \approx f(\theta)T_c^4 + [1 - f(\theta)]T_s^4, \quad (12)$$

where T_{RAD} is the surface radiometric temperature (K), T_c is the canopy radiometric temperature (K), T_s is the ground radiometric temperature (K), f is the fractional vegetation cover (–), and θ is the view zenith angle of sensors (°). The fractional vegetation cover can be expressed as follows:

$$f(\theta) = 1 - \exp[-0.5F/\cos(\theta)], \quad (13)$$

where F is the leaf area index (LAI) from the MODIS LAI products. The MODIS LST products can be expressed

as a hybrid of the canopy and soil temperatures from the CLM, so Eq. (12) can be rewritten as

$$T_{\mathrm{LST}}(\theta) = \langle \{ f(\theta)\varepsilon_C T_{\mathrm{Creal}}^4 + [1 - f(\theta)]\varepsilon_S T_{\mathrm{Sreal}}^4 \} / \varepsilon_{\mathrm{CS}} \rangle^{0.25}, \tag{14}$$

where T_{LST} is the MODIS LST product (K); T_{Creal} and T_{Sreal} are the canopy and soil temperatures (K), respectively, from CLM; ε_{CS} is the site emissivity (mixed emissivity of soil and canopy of test site), with values of 0.987 for grassland and cropland (Wang et al. 2008) and 0.978 for orchards (Valor and Caselles 1996); and ε_C and ε_S are the pure emissivities of the canopy and soil with values of 0.99 and 0.96, respectively (Valor and Caselles 1996).

Equation (14) was determined to be the observation operator of the data assimilation system. This observation operator relates CLM state variables to the MODIS LST products, which can be applied to both point and regional-scale assimilation tests.

c. Data assimilation strategies

In this study, two data assimilation strategies were developed to assimilate the MODIS LST products. Strategy 1 improves the estimation of water and heat fluxes by adjusting soil temperature, while strategy 2 adjusts the soil moisture.

1) STRATEGY 1

Equations (2)–(5), (8), and (9) indicate that soil temperature is a key variable for water and heat fluxes. The dynamics of soil temperature are assumed to obey the following heat diffusion equation:

$$c\frac{\partial T}{\partial t} = -\frac{\partial}{\partial z}\left(\lambda \frac{\partial T}{\partial z}\right), \tag{15}$$

where c is the volumetric soil heat capacity (J m^{-3} K^{-1}), T is the soil temperature (K), t is time (–), λ is the thermal conductivity of soil (W m^{-1} K^{-1}), and z is the soil depth (m). In this strategy, the CLM can be considered to be a ''black box'' system and is therefore described as follows:

$$X_{k+1} = M(X_k, \alpha_{k+1}, \beta_{k+1}), \tag{16}$$

where X_{k+1} and X_k represent the soil temperature profiles (K) at times $k + 1$ and k, respectively; M(–) represents the model operator; α_{k+1} represents the forcing data at time $k + 1$; and β_{k+1} represents surface parameters at time $k + 1$.

Following this strategy, the state vector contains the soil temperature profile, and the observation vector contains the surface temperature. From Eq. (16), the soil temperature profile at time $k + 1$ can be changed by adjusting the soil temperature profile at time k.

2) STRATEGY 2

As determined in Eqs. (3), (4), (10), and (11), soil moisture is a key factor of latent heat flux. In the CLM, when soil moisture increases, latent heat flux increases and surface temperature decreases, and vice versa. The other components of the energy balance equation change accordingly. In the CLM, liquid soil water is assumed to be governed by the following equation (Dai et al. 2003):

$$\frac{\partial W_{\mathrm{liq}}}{\partial t} = -\frac{\partial q}{\partial z} - f_{\mathrm{root}} E_{\mathrm{tr}} + M_{\mathrm{il}}, \tag{17}$$

where W_{liq} is the mass of the soil water (kg), t is time, q is the water flow within the soil (kg m^{-2} s^{-1}), f_{root} is the root fraction (–), E_{tr} is transpiration (kg m^{-2} s^{-1}), and M_{il} is the mass rate of the melting (positive) or freezing (negative) of soil ice. The vertical water flow within the soil is described by Darcy's law.

In this strategy, the state vector contains the soil moisture profile, and the observation vector contains the surface temperature. Therefore, soil moisture can be adjusted by changing the soil moisture at the previous time step.

d. Optimization algorithms

In this study, two different optimization algorithms were applied for the two strategies: the EnKF and SCE-UA.

1) ENKF

The Kalman filter (KF) method was first proposed by Kalman (1960) and is used for linear dynamic systems. The extended Kalman filter (EKF) was proposed (Jazwinski 1970) for nonlinear systems. It is based on the first-order linearization and is cumbersome when applied to complex problems. Evensen (1994) proposed the EnKF to circumvent the problematic integration of the EKF for more complex problems. According to Epstein's theory of stochastic dynamic prediction, the EnKF calculates the state prediction error covariance through the Monte Carlo method. It has been widely applied due to its ease of implementation. A brief algorithm summary is provided to explicate the implementation of the EnKF (Huang et al. 2008).

At the beginning of the algorithm, the first-guest value X_0, model parameters β_{k+1}, and background error covariance P_0 are determined according to prior knowledge. The initial state variable ensemble can be obtained by adding random noises to X_0:

$$X_{i,0} = X_0 + u_i \quad u_i \sim N(0, P_0), \tag{18}$$

where μ_i is the background error vector, which conforms to the Gaussian distribution with a zero mean and a covariance matrix of P_0. CLM state variables then proceed by adding i (i represents the ensemble member) number of random noises to Eq. (16), which conforms to the Gaussian distribution. This is expressed by the following equation:

$$X^f_{i,1} = M(X_{i,0}, \alpha_1, \beta_1) + w_i \quad w_i \sim N(0, Q_1), \tag{19}$$

where $X^f_{i,1}$ represents the forecasted state variables of the ith member at time 1, the superscript f means the forecasted state variables, w_i is the model error vector and conforms to a Gaussian distribution with a zero mean and a covariance matrix $\mathbf{Q}$, and Q is the model error.

When no MODIS LST exists, CLM state variables proceed using the following equation:

$$X^f_{i,k+1} = M(X^f_{i,k}, \alpha_{k+1}, \beta_{k+1}) + w_i \quad w_i \sim N(0, Q_k), \tag{20}$$

where $X^f_{i,k}$ and $X^f_{i,k+1}$ represent the forecasted state variables of the ith member at times k and $k + 1$, respectively. The $X^f_{i,k+1}$ remains a single function of $X^f_{i,k}$. When using the MODIS LST products at the time $k+1$, the observation operator predicts the surface temperature as follows:

$$Y_{i,k+1} = \mathbf{H}(X^f_{i,k+1}) + v_i \quad v_i \sim N(0, R_{k+1}), \tag{21}$$

where $Y_{i,k+1}$ is the surface temperature of the ith member at the time $k + 1$, $\mathbf{H}(-)$ represents the observation operator, v_i is the observation error that conforms to a Gaussian distribution with a zero mean and a covariance matrix $\mathbf{R}_{k+1}$, and R_{k+1} is the observation error. The calculation of the model and observation errors is introduced in section 4. The number of state variables is updated as follows:

$$X^a_{i,k+1} = X^f_{i,k+1} + \mathbf{K}_{k+1}(Y^o_{k+1} - Y_{i,k+1}), \tag{22}$$

$$\mathbf{K}_{k+1} = \mathbf{P}^f_{k+1}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f_{k+1}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{k+1})^{-1}, \tag{23}$$

$$\mathbf{P}^f_{k+1} = \frac{1}{N-1}\sum_{i=1}^{N}(X^f_{i,k+1} - \overline{X}^f_{k+1})(X^f_{i,k+1} - \overline{X}^f_{k+1})^{\mathrm{T}}, \tag{24}$$

$$\mathbf{P}^f_{k+1}\mathbf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}(X^f_{i,k+1} - \overline{X}^f_{k+1}) \times [\mathbf{H}(X^f_{i,k+1}) - \mathbf{H}(\overline{X}^f_{k+1})]^{\mathrm{T}}, \quad \text{and} \tag{25}$$

$$\mathbf{H}\mathbf{P}^f_{k+1}\mathbf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}[\mathbf{H}(X^f_{i,k+1}) - \mathbf{H}(\overline{X}^f_{k+1})] \times [\mathbf{H}(X^f_{i,k+1}) - \mathbf{H}(\overline{X}^f_{k+1})]^{\mathrm{T}}, \tag{26}$$

where $X^a_{i,k+1}$ represents the analyzed state variables of the ith member at the time $k + 1$, $\mathbf{K}_{k+1}$ is the Kalman gain matrix at the time $k + 1$, Y^0_{k+1} is the MODIS LST product at the time $k + 1$, $\mathbf{P}^f_{k+1}$ is the forecasted background error covariance matrix at the time $k + 1$, $\mathbf{H}^{\mathrm{T}}$ is the transposed matrix of the observation operator, N is the number of ensembles, $\overline{X}^f_{i,k+1}$ is the mean value of forecasted state variables at the time $k + 1$, $[\cdot]^{\mathrm{T}}$ represents the transposed matrix, $\mathbf{H}(X^f_{i,k+1})$ is the surface temperature of the ith member at the time $k + 1$, and $\mathbf{H}(\overline{X}^f_{k+1})$ is the mean value of the CLM surface temperature at the time $k + 1$.

2) SCE-UA

Although the EnKF works well for some nonlinear systems, it may perform poorly in some cases where the true posterior is non-Gaussian (Evensen 1997). As there may be a highly nonlinear relationship between soil temperature (and soil moisture) and surface temperature, the initial Gaussian probability density function (PDF) can lead to a non-Gaussian PDF. Therefore, we applied another type of optimization algorithm to our data assimilation tests.

In our data assimilation strategies, our objective was to minimize the differences in the simulated and observed surface temperatures, which can be expressed by the following equation:

$$J(x) = |T^{\mathrm{obs}}_s - T^{\mathrm{sim}}_s|, \tag{27}$$

where J is the cost function, T^{obs}_s is the observed surface temperature (MODIS LST products), and T^{sim}_s is the surface temperature calculated with Eq. (14). To minimize this cost function, the SCE-UA (Duan et al. 1993) was used to obtain the optimal control vector, which does not require the derivatives of the model and can avoid being corrupted by small pits and bumps on the function surface. The SCE-UA search routine is a global optimization strategy that combines the strength of the simplex method with the concept of a controlled random search, competitive evolution, and the strategy of complex shuffling. The synthesis of these four concepts makes the SCE-UA method more effective, robust, flexible, and efficient, although it is less sensitive to the initial values of parameters than the simplex method.

When no MODIS LST exists, the SCE-UA algorithm is not launched. When confronted with the MODIS LST, the SCE-UA algorithm is applied to minimize the cost function Eq. (27) in the current step. In the SCE-UA algorithm, the upper and lower bounds of the state variables should be determined by the following equations:

$$X_{\mathrm{lower}} = \max(X_{\min}, X_{\mathrm{sim}} - \Delta X) \quad \text{and} \tag{28}$$

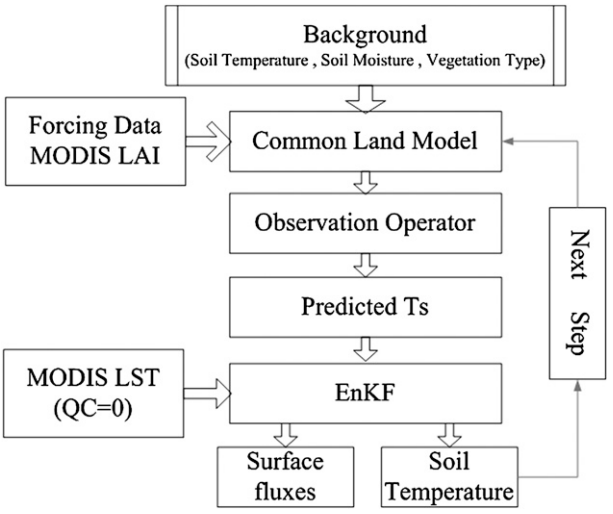

FIG. 1. Flowchart of data assimilation strategy 1.

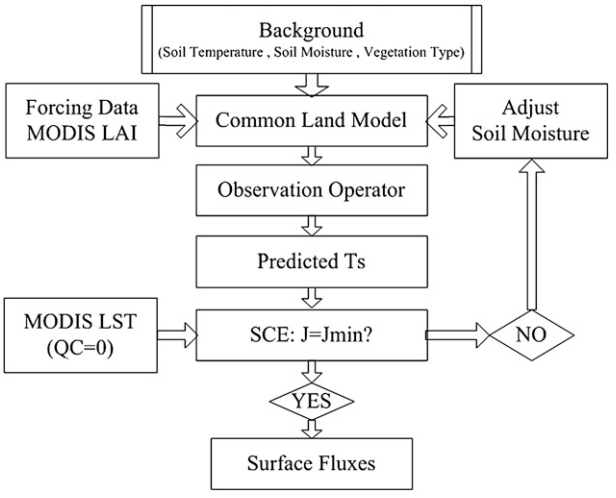

FIG. 2. Flowchart of data assimilation strategy 2.

$$X_{\text{upper}} = \min(X_{\max}, X_{\text{sim}} + \Delta X), \tag{29}$$

where X_{lower} and X_{upper} are the lower and upper bounds of the state variables, respectively; $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the state variables, respectively; X_{sim} is the simulated state variable; and ΔX is the uncertainty in the state variables.

In this study, $X_{\min}$ and $X_{\max}$ were 260 and 320 K, respectively, for soil temperature profiles and were used as the wilting point and porosity for soil moisture profiles (Pipunic et al. 2008). The uncertainty in state variables, ΔX, can be determined by the root-mean-square-error (RMSE) values of soil temperature and moisture as calculated in section 4.

Figures 1 and 2, respectively, depict flowcharts of strategy 1 with the EnKF and strategy 2 with the SCE-UA, which can improve the estimation of water and heat fluxes by adjusting the model state variables.

3. Data collection and processing

a. Site description

The data assimilation experiments were conducted at six observation sites: Lindenberg, Germany; Bondville, Illinois, and Goodwin, Mississippi, United States; and TongYu, GuanTao, and MiYun, China. These sites are covered with grasses, crops, and orchards; the details are given in Table 1. Data from the Lindenberg and TongYu sites are sourced from the Coordinated Enhanced Observing Period (CEOP), and the details are shown online (www.ceop.net). Data from the Bondville and Goodwin sites are sourced from FLUXNET (information available online at www.fluxnet.ornl.gov/fluxnet/index.cfm). Details for the GuanTao and MiYun sites are summarized in Table 2. The surfaces for most of the sites are smooth and even, and the MiYun site is located in a valley (approximately 500–1000 m × 2420 m) in the Hai River basin in China. In general, the meteorological data were typically measured using a set of instruments mounted on a tower for 30 min.

Continuous flux data measured by the eddy covariance (EC) system were used as validation data in this study, and the data were processed carefully by different groups of primary investigators (PIs; i.e., Baldocchi et al. 2001). Taking the EC data measured in the GuanTao and MiYun sites as an example, the procedure for processing EC data includes spike detection, lag correction of H_2O–CO_2 relative to the vertical wind component, sonic virtual temperature correction, coordinating rotation using the planar fit method, and corrections for air density fluctuation [Webb–Pearman–Leuning (WPL) correction; Webb et al. (1980)], frequency response correction, etc. An EdiRe (University of Edinburgh, information online at http://www.geos.ed.ac.uk/abs/research/micromet/EdiRe) software package was used for this purpose. Furthermore, the half-hourly flux data were screened via a two-step process: 1) measurements during the periods of precipitation (including 1 h before and after) were rejected and 2) measurements at night with friction velocity $u_* < 0.1$ m s^{-1} were rejected (Blanken et al. 1998).

Although the EC flux system is accepted as one of the best methods for measuring surface fluxes (Baldocchi et al. 2001), it has limitations. One of the biggest concerns is the ''energy imbalance'' in applications of the EC data, as observed in most experiments (Mauder et al. 2006; Oncley et al. 2007). The energy balance closure of

TABLE 1. Summary of the six observation sites.

Site name	Country	Lat	Lon	Elevation (m)	Land cover	Year
Lindenberg	Germany	52.17°N	14.12°E	112.0	Grass	2005
Goodwin	United States	34.25°N	89.97°W	70.0	Grass	2005
Bondville	United States	40.01°N	88.29°W	300.0	Cropland	2006
TongYu	China	44.42°N	122.87°E	184.0	Cropland	2004
GuanTao	China	36.52°N	115.13°E	30.0	Cropland	2009
MiYun	China	40.63°N	117.32°E	352.0	Orchard	2008

the EC system was assessed at the test sites, and the energy balance ratio (EBR) is defined as

$$\mathrm{EBR} = (H + LE)/(R_n - G_0), \tag{30}$$

where H and LE are the sensible and latent heat fluxes, respectively; R_n is the net radiation; and G_0 is the surface soil heat flux calculated following the method proposed by Yang and Wang (2008). Table 3 summarizes the EBR at the six sites and also demonstrates that an energy imbalance was also observed at the test sites.

b. MODIS data

Two MODIS sensors, on board the *Terra* and *Aqua* satellites have been launched for global studies of the atmosphere, land, and ocean processes. The *Terra* overpass time is approximately 1030 local solar time (LST) in its descending mode and 2230 LST in its ascending mode. The *Aqua* overpass time is approximately 1330 LST in its descending mode and 0130 LST in its ascending mode. In this study, the standard MODIS products (MOD11A1, MYD11A1, and MOD15A2) were used (collection 5) and are available online (https://wist.echo.nasa.gov/api/).

The MODIS LST products provide the temperature and emissivity values for each pixel. The generalized split-window algorithm is used to retrieve the temperature from the MODIS thermal and middle-infrared spectral regions (Wan and Dozier. 1996). Many validation results have indicated that the MODIS LST at 1-km resolution has an accuracy greater than 1 K in the range of 263–300 K (Wan et al. 2002), although some cases may have large errors (Wang et al. 2008). MOD11A1 and MYD11A1 provide two land surface temperature values (day and night) daily. Therefore, four MODIS LST observations are available daily with clear skies that can be assimilated into the developed assimilation schemes. MOD15A2 provides LAI values that are 8-day composites, and the values are incorporated into CLM directly. All these products are stored in the hierarchical data format (HDF), which is a sinusoidal projection with a spatial resolution of 1 km.

In addition, the MODIS products provide satellite view time and quality control flag data. As the MODIS data are often contaminated by clouds, only clear-sky products (QC = 0) are used. Table 4 shows the MODIS LAI data at the six observation sites.

TABLE 2. Summary of instruments at GuanTao and MiYun sites.

Parameters	Instruments		Height (m)	
	GuanTao	MiYun	GuanTao	MiYun
Wind speed/direction	WS03001; R. M. Young		12.7	10.66 and 30.56
Air temp/humidity	HMP45C; Vaisala		4 and 12.5	10.66 and 30.56
Air pressure	CS100; Campbell Scientific	AV-410; Avalon		
Precipitation	TE525; Campbell Scientific	Young-52203; R. M. Young		
Radiation	CNR-1; Kipp and Zonen		14.3	30.76
Soil heat flux	HFP01; Campbell Scientific	HFT-3; Hukseflux	0.02	0.02
IR temp	IRTC-3; Avalon		15.7	30.56
Soil temp	107; Campbell Scientific		0.02, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0	0.02, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0
Soil moisture	ECH$_2$O-10; Decagon		0.02, 0.05, 0.1, 0.2, 0.4, 0.6, 1.0	0.02, 0.05, 0.1, 0.2, 0.4, 0.6, 1.0
Fluxes	Eddy covariance system (CSAT3, Campbell Scientific; LI-7500, LI-COR)		15.6	26.66
	Large aperture scintillometer (LAS, Kipp and Zonen)		35.86 m (effective height) and 2420 m (pathlength)	15.6 and 2760 m

TABLE 3. Summary of the EBRs at six sites.

	Lindenberg	Goodwin	Bondville	TongYu	GuanTao	MiYun
EBR	0.79	0.96	0.78	0.95	0.87	0.78

c. Model input data

The input data for the CLM include land surface type, soil and vegetation parameters, and forcing data. The CLM is designed to handle a variety of data sources. Therefore, preprocessing of the data is necessary to take advantage of the CLM datasets. Land cover types are based on the International Geosphere–Biosphere Programme (IGBP) classification system. Soil texture is sourced from a global database according to the percentage of sand and clay. All the datasets are available at a spatial resolution of 30 s. Thus, the soil thermal and hydraulic properties can be calculated (Dai et al. 2001). CLM contains both time-invariant and time-varying vegetation parameters. The former are constant values related to different vegetation types. LAI is a key parameter of CLM and is sourced from the MODIS LAI products. The input forcing data used in this study were taken from a continuous series of half-hourly meteorological data at the six observation sites. They consist of wind speed, air temperature, relative humidity, air pressure, precipitation, incoming shortwave radiation, and incoming longwave radiation. Given the half-hourly temporal resolution of meteorological forcing data, the model was run at half-hourly time steps for all the experiments. The model state variables (soil temperature and soil moisture) were initialized by ground measurements at the six sites.

4. Results and analysis

a. Quantifying model and observation errors

There are three primary types of factors that contaminate the model outputs: the model structure, uncertainties in model parameters, and forcing data. When observations are available on a continuous temporal scale, model errors can be specified arbitrarily, as their influence diminishes quickly during the process of incorporating the measurements. However, forecasted errors are governed by the model errors as observations are intermittent. One of the problems associated with assimilating remote sensing measurements is that the data are often cloud contaminated, such as MODIS data, and therefore cannot be used. This results in useful data being available at larger intermittent intervals than desired. In these situations, it is important to obtain proper estimates of the model errors. However, the model error is the error in the model results, which is caused by model deficiencies. Model errors may change over time and space, and they are correlated with various model state variables, which results in an unpredictable propagation of model errors. Therefore, model errors are complex and difficult to quantify.

TABLE 4. MODIS standard LAI products from January (1) to December (12) at six sites (m^2m^{-2}).

	Year	1	2	3	4	5	6	7	8	9	10	11	12
Lindenberg	2005	0.0	0.5	0.7	1.2	1.5	1.6	1.1	1.0	1.0	0.8	0.5	0.0
Goodwin	2005	0.3	0.3	0.5	1.6	2.0	1.8	2.2	1.0	1.5	1.2	0.6	0.3
Bondville	2006	0.4	0.3	0.3	0.5	0.5	2.3	4.9	2.4	0.9	0.5	0.2	0.3
TongYu	2004	0.0	0.0	0.0	0.2	0.2	0.3	0.6	0.9	0.8	0.5	0.0	0.0
GuanTao	2009	0.1	0.2	0.5	1.2	1.0	0.5	2.2	3.3	2.6	0.5	0.3	0.2
MiYun	2008	0.1	0.1	0.2	0.3	0.9	1.7	2.2	2.1	1.4	0.9	0.2	0.1

The following equations can be used to describe the difference between the model simulations and the observations:

$$\mathrm{MD} = T_{\mathrm{sim}} - T_{\mathrm{obs}} \quad \text{and} \tag{31}$$

$$\mathrm{RMSE} = \left[\frac{1}{N-1}\sum_{i=1}^{N}(\mathrm{MD}_i - \overline{\mathrm{MD}})^2\right]^{0.5}, \tag{32}$$

where MD is the model deviation; T_{obs} and T_{sim} are the observed and simulated soil temperatures, respectively; N is the sample number; and $\overline{\mathrm{MD}}$ is the mean value of the model deviation. As the soil temperature changes rapidly in a single day, we create tables with 48 values of each layer for the diurnal variations. Figures 3a–f depict RMSE values for different layers of the soil temperature profiles at the six experiment sites (these figures only show the RMSE values of layers 1, 3, and 5). Figure 3g shows the average RMSE values of the six experiment sites, which were used to construct the error covariance matrix in the data assimilation tests. These figures depict the diurnal variations of the soil temperature RMSE values, which attained the highest value of approximately 6 K at noon (Fig. 3g). The first layer values had obvious diurnal variations, and the values decreased gradually from the upper layers to the lower layers. Using the same method, soil moisture RMSE value tables were developed for all the six sites (Table 5). As the soil moisture changes slowly in a single day, the tables were created with one daily value. Similarly, the average RMSE values at the six experiment sites were used to construct the error covariance matrix.

To obtain the observation errors, ground-measured surface temperatures were used to validate the MODIS LST products. Ground-measured surface temperatures from the test sites were obtained from a radiometer usually mounted on a tower, which requires a correction for the emissivity effect. Moreover, they are also affected by

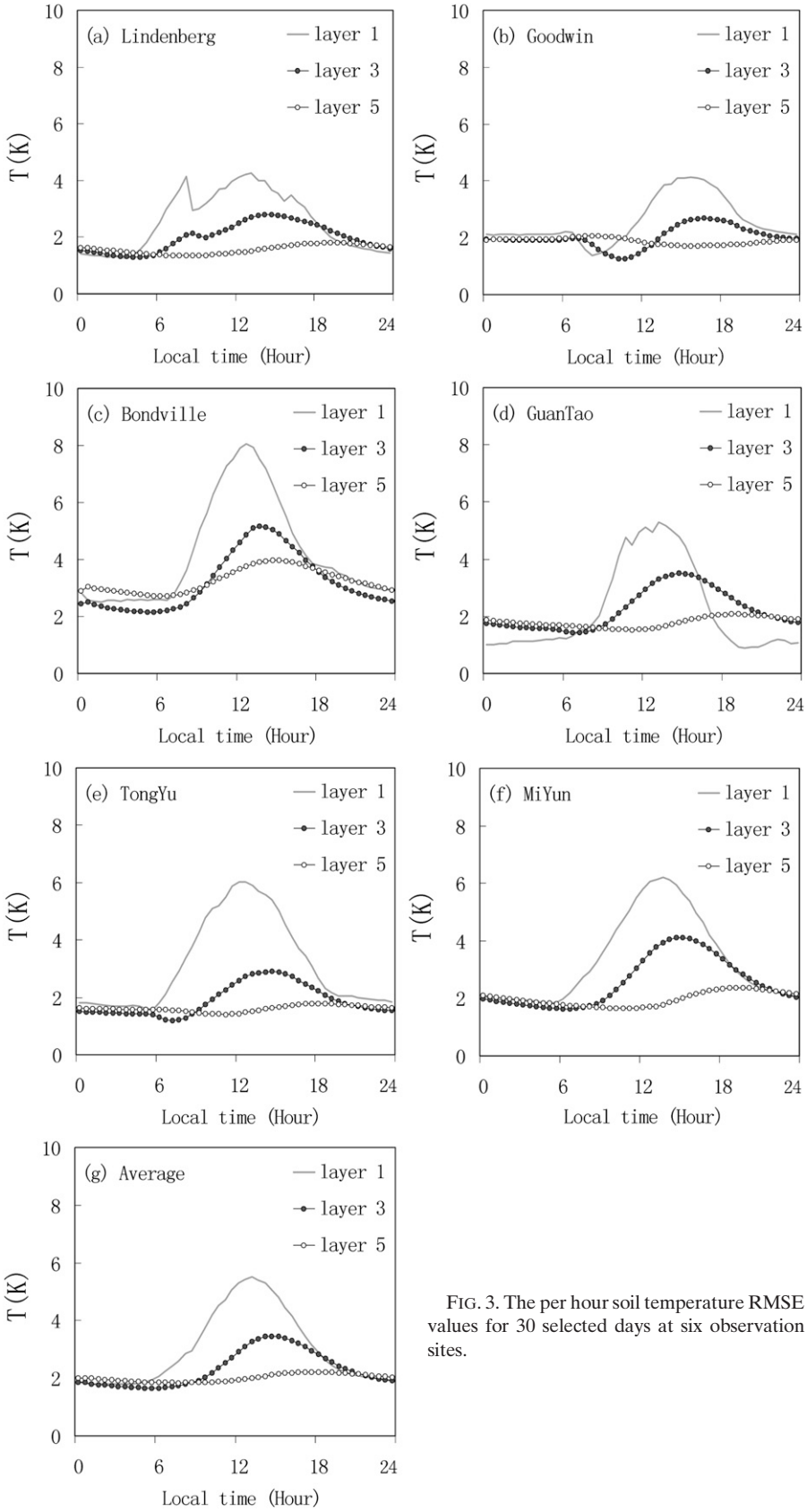

FIG. 3. The per hour soil temperature RMSE values for 30 selected days at six observation sites.

TABLE 5. Model error reference table for soil water contents at six sites (m^3m^{-3}).

Layer	Lindenberg	Goodwin	Bondville	GuanTao	TongYu	MiYun	Avg
1	0.032	0.030	0.045	0.043	0.036	0.033	0.036
2	0.031	0.024	0.035	0.041	0.035	0.031	0.033
3	0.029	0.028	0.040	0.029	0.035	0.032	0.033
4	0.028	0.036	0.028	0.047	0.034	0.030	0.034
5	0.027	0.038	0.012	0.045	0.036	0.030	0.032
6	0.020	0.046	0.007	0.040	0.037	0.030	0.030
7	0.007	0.037	0.006	0.070	0.037	0.030	0.032
8	0.004	0.036	0.006	0.032	0.037	0.027	0.024
9	0.004	0.036	0.006	0.032	0.037	0.027	0.024
10	0.004	0.036	0.006	0.032	0.037	0.027	0.024

water vapor when ground instruments are mounted at high altitudes. Therefore, the incoming longwave radiation must be considered. Based on the thermal radiative transfer theory, the outgoing longwave radiation at the surface level depends on the land surface temperature, emissivity, and incoming longwave radiation (Liang 2004; Wang et al. 2008). Ground-measured surface temperatures can be obtained as follows:

$$T_{\text{sfc}} = \left\{[F_u - (1-\varepsilon)F_d]/\varepsilon\sigma\right\}^{0.25}, \tag{33}$$

where F_u is the surface outgoing longwave radiation (W m^{-2}), F_d is the surface incoming longwave radiation (W m^{-2}), ε is the emissivity of experiment sites (–), and σ is the Stefan–Boltzmann constant (5.67×10^{-8} W m^{-2} K^{-4}).

The MODIS LST products were compared with the ground-measured surface temperatures (Fig. 4). As shown in Fig. 4, the MODIS LST products and ground-measured surface temperature follow the same trend, and the correlation coefficients (R^2) for the Lindenberg, Goodwin, Bondville, TongYu, GuanTao, and MiYun sites were 0.936, 0.881, 0.944, 0.809, 0.962, and 0.936, respectively, during the daytime and 0.977, 0.795, 0.876, 0.925, 0.903, and 0.827, respectively, during the nighttime. In addition, the MODIS LST products were lower than the ground-measured surface temperatures, which was particularly evident at the Goodwin and TongYu sites during the daytime. The RMSE between the MODIS LST products and the ground-measured surface temperature was about 2–3 K, and it was lower in the daytime than at night at most of the sites.

However, the deviation between the MODIS LST products and the ground-measured surface temperatures was caused by not only the quality of the MODIS LST products but also different temporal scales (MODIS LST is an instantaneous value, and the ground-measured temperature is the mean value over 30 min) and spatial scales (MODIS LST is a mean value over about 1 km × 1 km, and the ground-measured temperature is over tens of square meters).

In addition, Fig. 4 indicates that it is feasible to upscale the ground-measured surface temperature for the MODIS LST pixels by using a simple linearity equation at the stations. The MODIS LST products were corrected based on the linearity equations between the MODIS LST products and ground-measured surface temperatures. The RMSE values were identified as the observation errors of the data assimilation system for daytime and nighttime and are listed in Table 6.

b. Assimilation of MODIS LST products

In this section, the results of the two data assimilation strategies were validated and compared. The two optimization algorithms (EnKF and SCE-UA) were combined with the two strategies. Therefore, four different combinations of two strategies with two optimization algorithms were tested at the six experiment sites separately. Flowcharts of these two strategies are shown in Figs. 1 and 2. The MODIS LST products were assimilated into the CLM with the four combinations. The RMSE was used to assess the assimilation results.

The simulation and assimilation results are shown in Tables 7 and 8 and Figs. 5 and 6 for the test sites (taken continuously for about 30 days at each site). Taking into account the quality of the MODIS LST products and validation data, data for the Julian day were selected from 227 to 256 at Lindenberg in 2005, from 251 to 280 at Goodwin in 2005, from 181 to 210 at Bondville in 2006, from 211 to 240 at TongYu in 2004, from 123 to 152 at GuanTao in 2009, and from 192 to 221 at MiYun in 2008.

Table 7 summarizes the RMSE values of the simulation and assimilation results compared with the observations at all sites. These results indicate that the estimates for the surface temperature and water and heat fluxes were improved by the two strategies and two optimization algorithms. Overall, strategy 2 was superior to strategy 1, and the EnKF algorithm performed slightly better than the SCE-UA algorithm for the same strategy. Strategy 2 combined with the EnKF algorithm performed best for water and heat fluxes, and the reductions in the RMSE

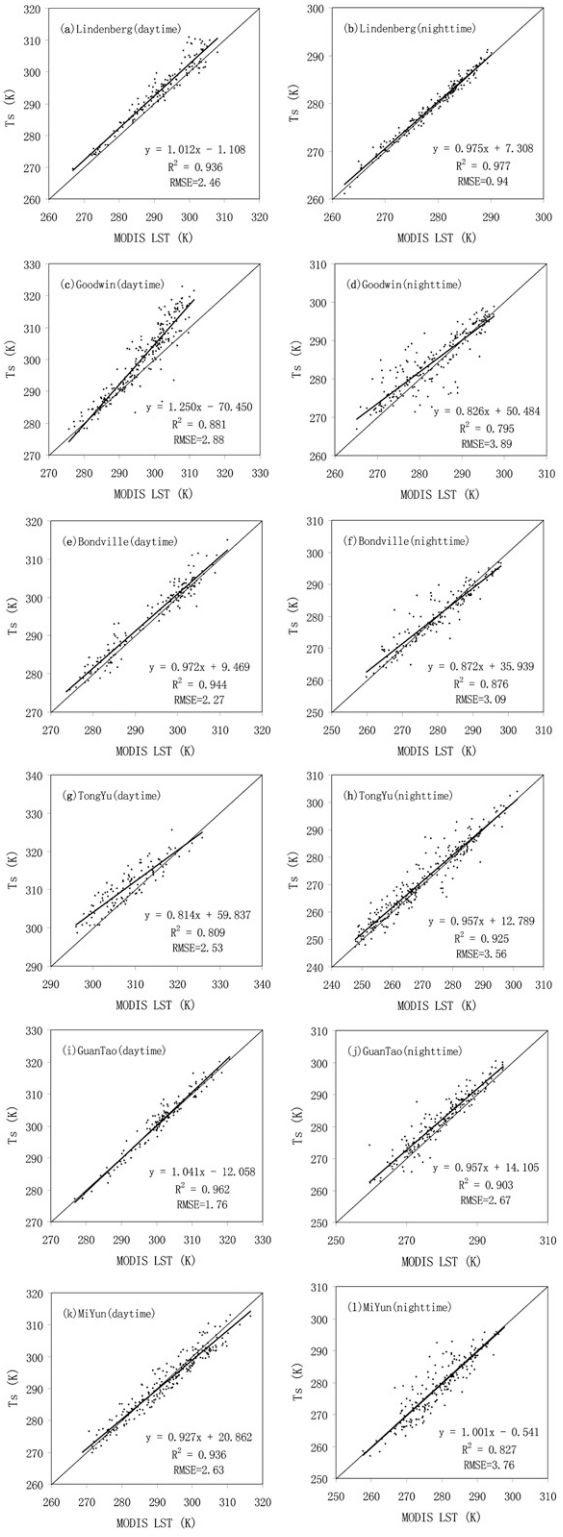

FIG. 4. Comparisons between MODIS LST products and ground-measured surface temperatures (T_{sfc}) at six experiment sites. (RMSE and R are the root-mean-square error and the correlation coefficient between the MODIS LST and the ground-measured surface temperatures, respectively.)

were 24.0 and 15.2 W m^{-2} for sensible and latent heat fluxes, respectively. Table 7 also summarizes the RMSE of the simulation and assimilation results compared with the observations for grasslands (two sites), croplands (three sites), and orchards (one site).

Table 8 details the results at the six observation sites. The simulation and assimilation results for the Lindenberg (grassland) and GuanTao (cropland) sites are shown in Figs. 5 and 6, respectively. In strategy 1, the most significant reduction in RMSE was 3.7 K for the surface temperature using the SCE-UA algorithm at the GuanTao site. For sensible heat flux, the most significant reduction was 24.4 W m^{-2} with the EnKF algorithm at the Lindenberg site. For latent heat flux, the most significant reduction was 24.5 W m^{-2} with the EnKF algorithm at the Goodwin site. In strategy 2, the EnKF algorithm achieved the most significant reductions in RMSE values. For surface temperature, the most significant reduction in the RMSE value was 4.0 K at the GuanTao site. The sensible heat flux saw the most significant reduction in its RMSE value at 48.1 W m^{-2} at the GuanTao site, and the latent heat flux was 27.8 W m^{-2} at the Goodwin site. Since latent heat flux is significantly affected by soil moisture, strategy 2 is superior to strategy 1, especially due to the improvements in the latent heat flux estimates. Figure 6 shows an example where the assimilation results of the latent heat flux exhibited no improvements during several days at the GuanTao site, while strategy 2 demonstrated significant improvements.

To test the significance of reductions in RMSE values through data assimilation, a paired t test was conducted, and the results are shown in Table 9. The smaller P values in the paired t test demonstrate the more obvious significance of reductions. Table 9 shows that strategy 2 achieved more significant reductions than strategy 1, and the EnKF algorithm showed more significance than the SCE-UA when using the same strategy.

Figures 7a and 7b show the results for the MiYun site on Julian day 217. The data assimilation curve refers to strategy 2 with the EnKF algorithm. Although the data assimilation scheme improved the surface temperature predictions, the assimilation results for the sensible heat flux may not perform better than the simulation. Equations (1), (8), and (9) indicate that sensible heat flux is determined not only by the surface–air temperature differences but also by aerodynamic resistance. To explain this phenomenon, the surface–air temperature differences of the CLM and observations on Julian day 217 at the MiYun site are plotted in Fig. 7c, and aerodynamic resistances are plotted in Fig. 7d. The aerodynamic resistance is calculated according to the following equation:

TABLE 6. Regression equations and RMSEs between MODIS LST and ground-measured surface temperatures at six sites.

	Daytime		Nighttime	
	Regression equation	RMSE (K)	Regression equation	RMSE (K)
Lindenberg	$Y = 1.012X - 1.108$	2.46	$Y = 0.975X + 7.308$	0.94
Goodwin	$Y = 1.250X - 70.450$	2.88	$Y = 0.826X + 50.484$	3.89
Bondville	$Y = 0.972X + 9.469$	2.27	$Y = 0.872X + 35.939$	3.09
TongYu	$Y = 0.814X + 59.837$	2.53	$Y = 0.957X + 12.789$	3.56
GuanTao	$Y = 1.041X - 12.058$	1.76	$Y = 0.957X + 14.105$	2.67
MiYun	$Y = 0.927X + 20.862$	2.63	$Y = 1.001X - 0.541$	3.76

$$r_{ah} = \rho_{atm} C_p (T_{sfc} - \theta_{atm})/H, \quad (34)$$

where H is the sensible heat flux (W m^{-2}) and T_{sfc} is the surface temperature (K). In Eq. (34), T_{sfc} and H are model outputs used to calculate the aerodynamic resistances of the CLM. Both T_{sfc} and H are ground measurements (from a radiometer and EC system) used to calculate the aerodynamic resistance observations. Figure 7d shows that CLM overestimated the resistance at this site, and oscillations at times close to sunset and sunrise were caused by near-zero sensible heat fluxes. The deviation between the simulated and observed aerodynamic resistance also existed at other sites such as Goodwin and TongYu (not shown). In the CLM, aerodynamic resistance was dominated by meteorological data, surface conditions, and the parameterization scheme; these need to be calibrated in the future.

Figure 5d also indicates that strategy 2 with the SCE-UA algorithm is sensitive to the quality of the MODIS LST products. As the MODIS LST products are often contaminated by clouds, clear-sky products (QC = 0) are not easily available. Moreover, because of difference in the temporal and spatial scales of the ground measurements and the MODIS LST products, the MODIS LST products often have large errors (Fig. 4), which leads to limited improvements of the assimilation results. On Julian days 233, 234, 242, and 247, the assimilation results agreed well with the observations, as the differences between the MODIS LST products and the ground measurements were within 1 K. However, on Julian day 248, the assimilation result for the sensible heat flux was smaller than the observation, and the latent heat flux was larger than the observation; the MODIS LST was smaller than the ground-measured temperature. Finally, on Julian days 231 and 232, the assimilation results were similar to the simulation results due to the MODIS LST being badly contaminated by clouds.

Therefore, the correct results may sometimes not be obtained with the assimilation of only one dataset, and the joint assimilation of the MODIS LST products and soil moisture observations simultaneously may yield better predictions. Thus, a simple experiment was conducted to jointly assimilate both the MODIS LST products and ground-measured surface soil moisture observations using strategy 2 with the SCE-UA. The lower and upper bounds of the soil moisture variations can be defined as

$$X_{lower} = \max(X_{min}, X_{obs} - \Delta X) \quad \text{and} \quad (35)$$

$$X_{upper} = \min(X_{max}, X_{obs} + \Delta X), \quad (36)$$

where X_{obs} is the soil moisture observation.

The Lindenberg site was selected as a case study site, and the assimilation results are shown in Fig. 8 (in this

TABLE 7. Summary of the simulation and assimilation results.

	T_{sfc} (K)					H (W m^{-2})					LE (W m^{-2})				
	Sim	1st_E	1st_S	2nd_E	2nd_S	Sim	1st_E	1st_S	2nd_E	2nd_S	Sim	1st_E	1st_S	2nd_E	2nd_S
Total	4.0	3.2	3.1	2.7	3.1	48.4	35.7	41.0	24.4	31.6	62.7	54.3	57.2	47.5	53.1
Grass	2.7	2.3	2.2	2.5	2.5	37.2	22.9	28.5	20.8	30.1	67.7	44.4	55.3	41.6	48.6
Crop	4.7	3.5	3.2	3.1	3.5	55.6	45.4	48.6	27.1	34.8	57.0	56.9	53.9	45.4	51.3
Orchard	4.5	4.2	4.3	2.1	3.2	49.0	32.1	43.6	23.8	24.9	69.6	66.6	70.7	65.5	67.3

T_{sfc}, H, and LE = surface temperature, and sensible and latent heat fluxes, respectively.
Sim = simulation results.
1st_E = assimilation results of strategy 1 with the EnKF method.
1st_S = assimilation results of strategy 1 with the SCE-UA method.
2nd_E = assimilation results of strategy 2 with the EnKF method.
2nd_S = assimilation results of strategy 2 with the SCE-UA method.

TABLE 8. RMSEs of the simulation and assimilation results at six sites.

		Grass		Cropland			Orchard
		Lindenberg	Goodwin	Bondville	TongYu	GuanTao	MiYun
T_{sfc} (K)	Sim	2.0	3.4	4.0	3.7	6.3	4.5
	1st_E	1.5	3.0	3.4	3.5	3.7	4.2
	1st_S	1.3	3.1	3.5	3.5	2.6	4.3
	2nd_E	2.6	2.3	3.7	3.2	2.3	2.1
	2nd_S	1.6	3.3	3.4	3.5	3.6	3.2
H (W m^{-2})	Sim	39.4	35.0	57.3	32.8	76.8	49.0
	1st_E	15.0	30.9	35.8	29.6	70.9	32.1
	1st_S	22.6	34.3	38.1	31.9	75.7	43.6
	2nd_E	13.7	27.8	29.1	23.4	28.7	23.8
	2nd_S	25.5	34.6	34.6	28.7	41.0	24.9
LE (W m^{-2})	Sim	56.2	79.1	53.3	37.3	80.5	69.6
	1st_E	34.1	54.6	35.1	41.4	94.2	66.6
	1st_S	54.8	55.8	42.6	38.1	80.9	70.7
	2nd_E	30.9	52.2	38.7	31.1	66.5	65.5
	2nd_S	45.9	51.3	37.8	40.8	75.4	67.3

figure, assimilation_a means assimilation results without soil moisture observations, while assimilation_b means assimilation results with soil moisture observations). Figure 8 shows that the assimilation_b curves of the water and heat fluxes were closer to the observations than the simulation and assimilation_a. Since the ground-measured soil moisture was assimilated, the RMSE value decreased from 25.5 to 21.4 W m^{-2} for the sensible heat flux and decreased from 45.9 to 43.9 W m^{-2} for the latent heat flux.

5. Summary and discussion

In this study, four different data assimilation scheme combinations derived from two strategies and two optimization algorithms (EnKF and SCE-UA) were developed based on the Land Surface Model (CLM) to improve the estimates of water and heat fluxes. The MODIS LST products were assimilated into the CLM. The MODIS LAI products were used to update the LAI in the CLM, which can accurately describe vegetation variations.

Overall, the two strategies improved the estimates of surface temperature and water and heat fluxes by adjusting the soil temperature and moisture, respectively. Strategy 2 was superior to strategy 1 when using the same optimization algorithm. The EnKF algorithm performed slightly better than the SCE-UA algorithm when using the same strategy. Strategy 2 combined with the EnKF algorithm performed best for water and heat fluxes, and the reductions in RMSE values were 24.0 and 15.2 W m^{-2} for the sensible and latent heat fluxes, respectively. For the results at the six experiment sites, the combination of strategy 2 with the EnKF algorithm achieved the most significant reductions in RMSE values. For surface temperature, the most significant reduction in the RMSE value was 4.0 K at the GuanTao site. The sensible heat flux had the most significant reduction in its RMSE value of 48.1 W m^{-2} at the GuanTao site, and latent heat flux reduction was 27.8 W m^{-2} at the Goodwin site. The paired t test indicates that strategy 2 shows more significance of reductions than strategy 1, and the EnKF algorithm shows more significance than SCE-UA using the same strategy. Since latent heat flux is significantly affected by soil moisture, strategy 2 is superior to strategy 1, especially for improving latent heat flux estimates. Figure 6 shows an example where the assimilation results of the latent heat flux produced no improvements during several days at the GuanTao site, while strategy 2 demonstrated significant improvements.

Two different optimization algorithms were used to assimilate The MODIS LST products, i.e., the EnKF and SCE-UA. The EnKF algorithm performed slightly better than the SCE-UA algorithm when using the same strategy. Since the SCE-UA algorithm is relaunched when the observations are available, it is computationally inefficient compared to the EnKF. In this study, only point-scale tests were conducted, and it is computationally unaffordable for the current computer system to apply this algorithm on a large scale.

The CLM is designed to handle a variety of data sources, and most of the model parameters were calculated with the preprocessing program in the CLM. Since the datasets are available in a relatively coarse database (30 s), there may be biases in water and heat flux simulations at the point scale. The relationship between the surface temperature and water and heat fluxes is dependent on a range of variables (e.g., soil–vegetation parameters and meteorological conditions). Uncertainties in any one of these factors will result in suboptimal assimilation results. Figure 7 shows an example where the

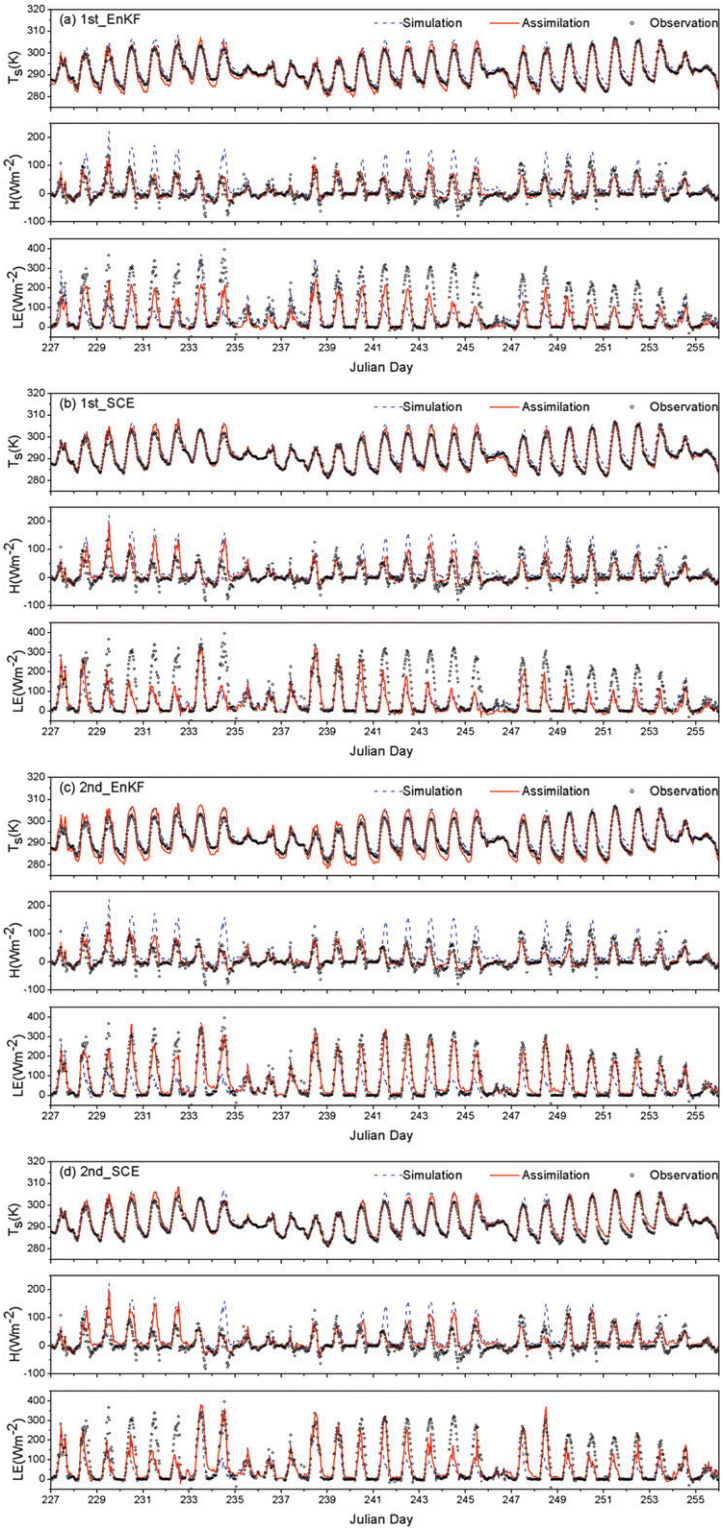

FIG. 5. Simulation and assimilation results at the Lindenberg site from Julian days 227 to 255, 2005: (a) 1st_EnKF means strategy 1 with the EnKF algorithm, (b) 1st_SCE means strategy 1 with the SCE-UA algorithm, (c) 2nd_EnKF means strategy 2 with the EnKF algorithm, and (d) 2nd_SCE means strategy 2 with the SCE-UA algorithm; (top to bottom) T_{sfc}, H, and LE represent the surface temperature, sensible heat flux, and latent heat flux, respectively.

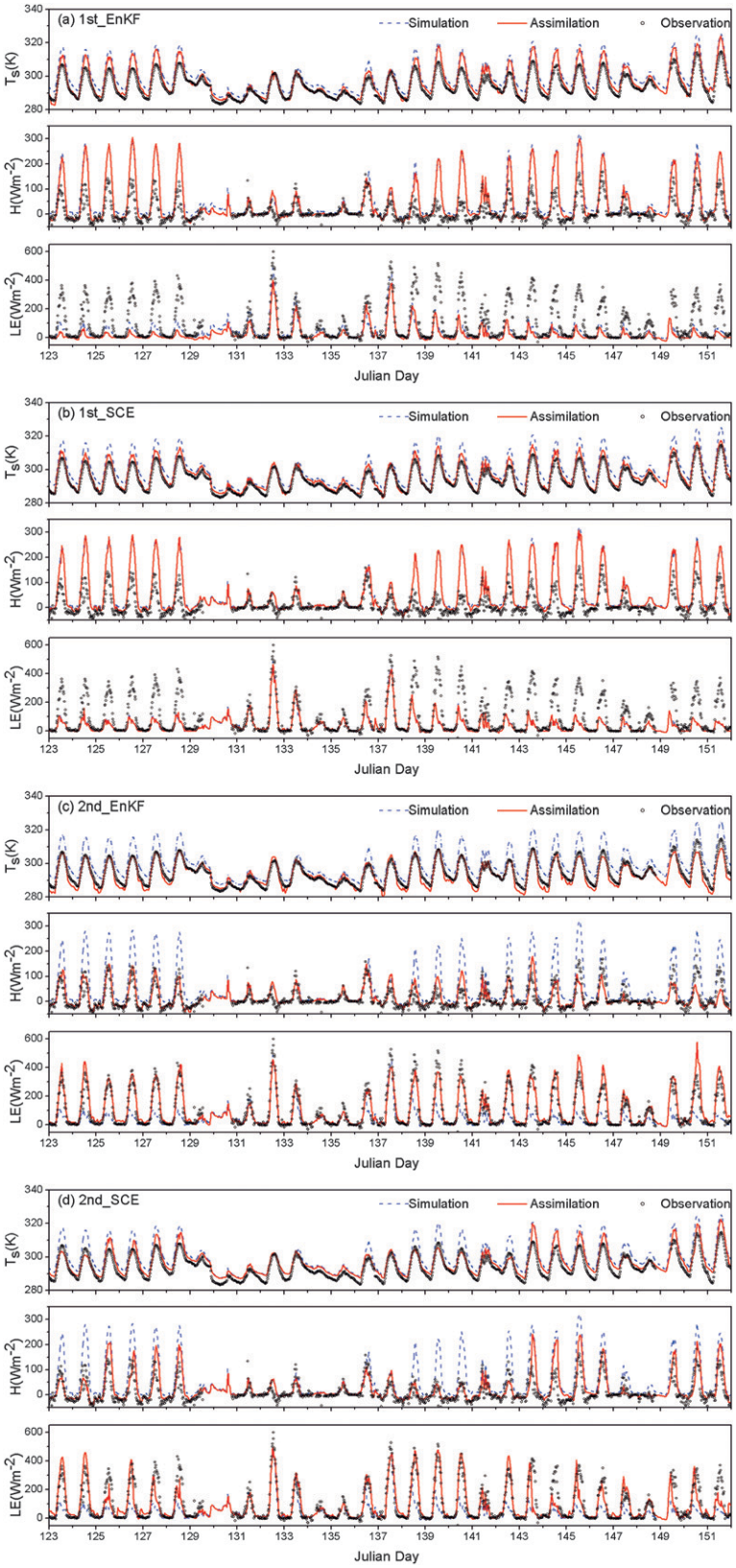

FIG. 6. As in Fig. 5, but at the GuanTao site from Julian days 123 to 151, 2009.

TABLE 9. The P values of a paired t test between the simulation and assimilation RMSEs.

	1st_E	1st_S	2nd_E	2nd_S
T_{sfc} (K)	0.09	0.16	0.12	0.08
H (W m^{-2})	0.02	0.08	0.01	0.03
LE (W m^{-2})	0.24	0.23	0.01	0.08

assimilation results for sensible heat flux were worse than simulation results, as the aerodynamic resistances were overestimated. Table 8 also indicates that the most significant reductions in RMSE values commonly took place at the same site with the largest simulation RMSE values, which may be caused by model biases. De Lannoy et al. (2007) indicated that if the model has not been calibrated will lead to bias in the model results, and EnKF has a limit to tackle the problem of model biases for the above problems, the models need to be calibrated with in situ measurements since water and heat fluxes are affected by not only soil moisture/temperature but also model parameters. An autocalibration data assimilation scheme should be developed; the model can then be calibrated automatically using data assimilation techniques. Furthermore, comparisons between different parameterization schemes should be conducted in the future, as different schemes may give distinct model outputs.

In addition, the performance of strategy 2 is sensitive to the quality of the MODIS LST products (Fig. 6d). As the MODIS LST products are often contaminated by clouds, clear-sky products (QC = 0) are not readily available. Moreover, because of different temporal and spatial scales, the MODIS LST products have large errors (Fig. 4), which may result in limited improvements. Better assimilation results were obtained by jointly assimilating the MODIS LST and soil moisture observations (Fig. 8). For these problems, LST retrieval algorithms need to be

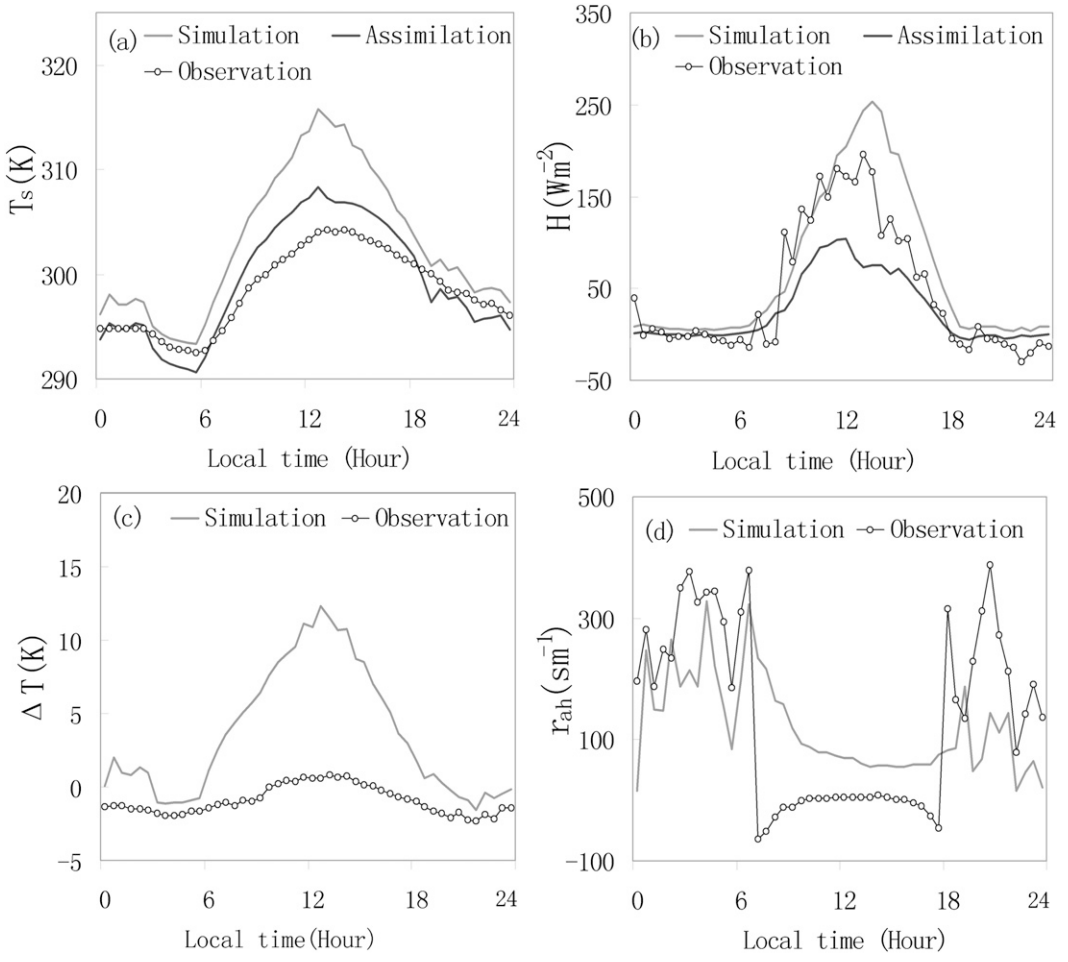

FIG. 7. Diurnal variations of aerodynamic resistance at the MiYun site on Julian day 217, 2008: (a) surface temperature, (b) sensible heat flux, (c) surface–air temperature differences, and (d) aerodynamic resistance. The assimilation curve refers to strategy 2 with the EnKF algorithm.

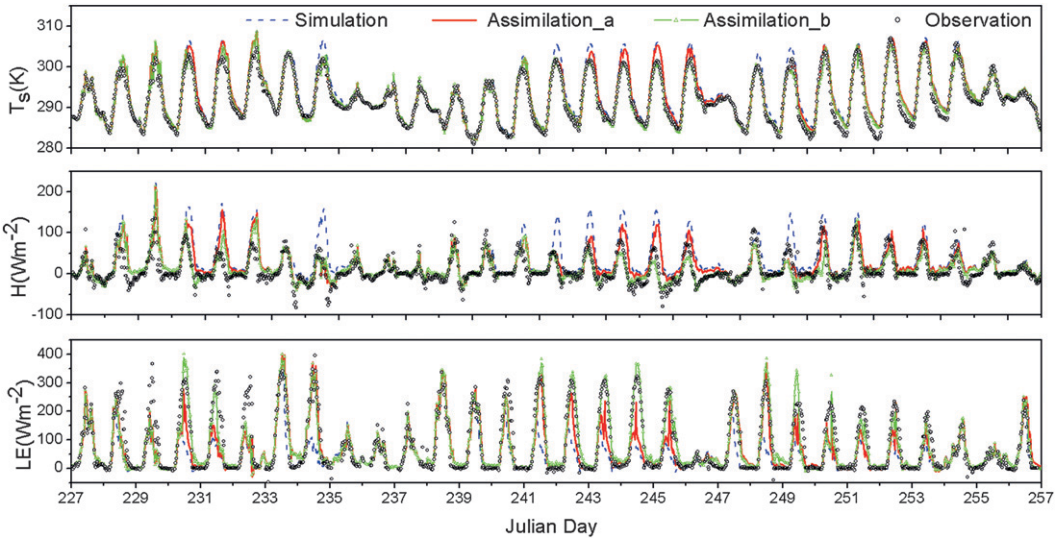

FIG. 8. Results of joint assimilation of the MODIS LST products and ground-measured soil moisture observations at the Lindenberg site from Julian days 227 to 255, 2005. (The assimilation curves refer to strategy 2 with the SCE-UA algorithm; (top to bottom) T_{sfc}, H, and LE represent the surface temperature, sensible heat flux, and latent heat flux, respectively; Assimilation_a means only the assimilation of MODIS LST products; and Assimilation_b means the joint assimilation of MODIS LST products and ground-measured soil moisture observations.)

improved, especially for complex land surfaces. Moreover, multisatellite data assimilation systems should be developed to jointly assimilate remote sensing surface temperature [MODIS, Advanced Very High Resolution Radiometer (AVHRR), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), etc.] and surface soil moisture retrievals [*Soil Moisture and Ocean Salinity Satellite* (*SMOS*), Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E), etc.], which may overcome the defects of one dataset. Meanwhile, studies on upscaling and downscaling should be enhanced due to multisatellite data at different spatial scales. On the other hand, assimilating water and heat flux observations into land surface models directly can improve the study of water and energy balances. In this paper, the relationship equations between the surface temperature from the CLM and MODIS LST products are defined as the observation operator. The MODIS LAI products were used in this observation operator, which is helpful in the development of a regional data assimilation system, but its applicability needs further validation.

Finally, there were some measurement and representative errors within the observation instrument, which may influence the validation results. For example, the surface temperature simulation took the mean temperature over a few square kilometers, while ground-measured temperatures may represent only tens of square meters. Moreover, the EC system suffered from an energy imbalance problem, which was observed in our experiments. Therefore, studies on the energy imbalance problem should be increased, which would also help improve the data assimilation algorithm.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (40971194 and 30911130504), the Special Research Foundation of Public Benefit Industry (GYHY200706046), the National Basic Research Program of China (2007CB714401), and the European Commission as part of the CEOP–AEGIS project (Call FP7-ENV-2007-1, Grant 212921). We thank three reviewers for their valuable comments that greatly improved the presentation of this paper.

REFERENCES

Anderson, M. C., J. M. Norman, W. P. Kustas, F. Li, J. H. Prueger, and J. R. Mecikalski, 2005: Effects of vegetation clumping on two-source model estimates of surface energy fluxes from an agricultural landscape during SMACEX. *J. Hydrometeor.,* **6,** 892–909.

Baldocchi, D., E. Falge, L. H. Gu, and R. Olson, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Amer. Meteor. Soc.,* **82,** 2415–2434.

Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes, and A. A. M. Holtslag, 1998: A remote sensing Surface Energy Balance Algorithm for Land (SEBAL). 1. Formulation. *J. Hydrol.,* **212–213,** 198–212.

Blanken, P. D., and Coauthors, 1998: Turbulence flux measurements above and below the overstory of a boreal aspen forest. *Bound.-Layer Meteor.,* **89,** 109–140.

Boni, G., D. Entekhabi, and F. Castelli, 2001: Land data assimilation with satellite measurements for the estimation of surface energy balance components and surface control on evaporation. *Water Resour. Res.,* **37,** 1713–1722.

Caparrini, F., F. Castelli, and D. Entekhabi, 2004: Estimation of surface turbulent fluxes through assimilation of radiometric surface temperature sequences. *J. Hydrometeor.,* **5,** 145–159.

Crow, W. T., and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.,* **26,** 137–149.

——, W. P. Kustas, and J. H. Prueger, 2008: Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model. *Remote Sens. Environ.,* **112,** 1268–1281.

Dai, Y. J., and Coauthors, 2001: Common Land Model (CLM): Technical documentation and user's guide, 69 pp. [Available online at http://climate.eas.gatech.edu/dai/clmdoc.pdf.]

——, and Coauthors, 2003: The Common Land Model. *Bull. Amer. Meteor. Soc.,* **84,** 1013–1023.

De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest, 2007: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resour. Res.,* **43,** W09410, doi:10.1029/2006WR005449.

Dickinson, R. E., P. J. Kennedy, A. Henderson-Sellers, and M. Wilson, 1986: Biosphere–Atmosphere Transfer Scheme (BATS) version 1E as coupled to the NCAR Community Climate Model. NCAR Tech. Rep. TN-275+STR, 72 pp.

Duan, Q. Y., V. K. Gupta, and S. Sorooshian, 1993: Shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory Appl.,* **76,** 501–521.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99,** 10 143–10 162.

——, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.,* **125,** 1342–1354.

Galantowiez, J. F., D. Entekhabi, and E. G. Njoku, 1999: Test of sequential data assimilation for retrieving profile soil moisture and temperature from observed l-band radio brightness. *IEEE Trans. Geosci. Remote Sens.,* **37,** 1860–1870.

Huang, C. L., X. Li, and L. Lu, 2008: Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sens. Environ.,* **112,** 1320–1336.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Elsevier, 376 pp.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.,* **82,** 35–45.

Kumar, P., and A. L. Kaleita, 2003: Assimilation of near-surface temperature using extended Kalman filter. *Adv. Water Resour.,* **26,** 79–93.

Liang, S., 2004: *Quantitative Remote Sensing of Land Surfaces*. John Wiley and Sons, 534 pp.

——, and J. Qin, 2008: Data assimilation methods for land surface variable estimation. *Advances in Land Remote Sensing: System, Modeling, Inversion and Applications,* S. Liang, Ed., Springer, 319–339.

Liu, S. M., G. Hu, L. Lu, and D. F. Mao, 2007: Estimation of regional evapotranspiration by TM/ETM+ data over heterogeneous surfaces. *Photogramm. Eng. Remote Sens.,* **73,** 1169–1178.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne, 2002: Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment. *Water Resour. Res.,* **38,** 1229, doi:10.1029/2001WR001114.

Mauder, M., C. Liebethal, M. Gckede, J. P. Leps, F. Beyrich, and T. Foken, 2006: Processing and quality control of flux data during LITFASS-2003. *Bound.-Layer Meteor.,* **121,** 67–88.

Oncley, S. P., T. Token, R. Vogt, W. Kohsiek, H. A. R. De Bruin, C. Bernhofer, C. Andreas, and E. V. Gorsel, 2007: The energy balance experiment EBEX-2000. Part I: Overview and energy balance. *Bound.-Layer Meteor.,* **123,** 1–28.

Pipunic, R. C., J. P. Walker, and A. Western, 2008: Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. *Remote Sens. Environ.,* **112,** 1295–1305.

Qin, J., S. Liang, K. Yang, T. Koike, R. Liu, and I. Kaihotsu, 2009: Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal. *J. Geophys. Res.,* **114,** D15103, doi:10.1029/2008JD011358.

Reichle, H. R., D. B. Mclaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 103–114.

Sellers, P. J., D. A. Randall, G. J. Collatz, C. B. Field, D. A. Dazlich, C. Zhang, G. D. Collelo, and L. Bounoua, 1996: A revised land surface parameterization (SiB2) for atmosphere GCMs. Part I: Model formulation. *J. Climate,* **9,** 676–705.

Su, Z., 2002: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci.,* **6,** 85–99.

Valor, E., and V. Caselles, 1996: Mapping land surface emissivity from NDVI: Application to European, African, and South American areas. *Remote Sens. Environ.,* **57,** 167–184.

Wan, Z., and J. Dozier, 1996: A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.,* **34,** 892–905.

——, Y. Zhang, Q. Zhang, and Z. L. Li, 2002: Validation of the land surface temperature products retrieved from *Terra* Moderate Resolution Imaging Spectroradiometer data. *Remote Sens. Environ.,* **83,** 163–180.

Wang, W. H., S. L. Liang, and M. Tilden, 2008: Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sens. Environ.,* **112,** 623–635.

Webb, E. K., G. I. Pearman, and R. Leuning, 1980: Correction of flux measurements for density effects due to heat and water vapor transfer. *Quart. J. Roy. Meteor. Soc.,* **106,** 85–100.

Yang, K., and J. Wang, 2008: A temperature prediction-correction method for estimating surface soil heat flux from soil temperature and moisture data. *Sci. China Ser.,* **51D,** 721–729.

——, W. Takahiro, and K. Toshio, 2007: An auto-calibration system to assimilate AMSR-E data into a land surface model for estimating soil moisture and surface energy budge. *J. Meteor. Soc. Japan,* **85,** 229–242.