

## Research Article

# Retrieval of Land Surface Model State Variables through Assimilating Screen Level Humidity and Temperature Measurements

## **Diandong Ren<sup>1</sup> and Ming Xue<sup>2</sup>**

<sup>1</sup>Department of Physics and Astronomy, Curtin University, Perth, WA 6845, Australia <sup>2</sup>School of Meteorology, University of Oklahoma, Norman, OK 73072, USA

Correspondence should be addressed to Diandong Ren; rendianyun@gmail.com

Received 9 November 2015; Accepted 28 April 2016

Academic Editor: Francois Counillon

Copyright © 2016 D. Ren and M. Xue. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study demonstrates successful variational retrieval of land surface states by assimilating screen level atmospheric measurements of specific humidity and air temperature. To this end, the land surface scheme is first validated against the Oklahoma Atmospheric Surface Layer Instrumentation System measurements with necessary refinements to the forward model implemented. The retrieval scheme involves a 1D land surface-atmosphere model, the corresponding adjoint codes, and a cost function that measures residuals between observed and modeled screen level atmospheric temperature and specific humidity. The retrieval scheme is robust when subjected to observational errors with magnitudes comparable to instrument accuracy and for initial guess errors larger than typical model forecast errors. Using varying assimilation window lengths centered on different periods of a day, the sampling strategy is assessed. The daytime observations are more informative compared to nocturnal observations. An assimilation window as narrow as four hours, if centered on local noon, contains comparable information to an expanded window covering the whole day. There exists an optimal assimilation window length resulting from the contest between degrading forecast accuracy and increasing information content. For an assimilation window less than two days, the "optimal" assimilation window length is inversely proportional to the data ingesting frequency.

#### **1. Introduction**

Variational data assimilation of energy fluxes is a core component of numerical models used to investigate convective initiation. For example, the hurricane landfall was found (personal communication with L. Leslie, 2014) not only to be sensitive to SSTs but also to be the land surface condition in the coastal region for both flat and hilly terrains. A major obstacle to improving numerical modeling of near surface variables is the sensitivity of model predictions to the accuracy of specification of key land surface parameters. The motivation for retrieving initial land surface conditions and the reason for choosing a variational approach have been justified in [1], where we also discussed the assimilation of ground surface temperature for retrieving initial land surface condition. In reality, except for the data rich areas such as Oklahoma, the only source for skin temperature is that derived from satellite remote sensing. Under cloudy conditions, there is no guarantee for high quality data availability if thermal infrared sensors are used. If microwave sensors are involved as the sources for skin temperature observations, the accuracy of the measurements remains an ongoing problem and there only will be a couple of satellite overpasses per day from one low Earth orbit platform [2]. Data collected this way may be too sparse for many study purposes. It would be desirable if routinely available screen level observations can be utilized to infer initial land surface conditions.

Here we discuss several closely related numerical modeling studies of recent years so as to put our work into the appropriate context. Indirect methods have been proposed to retrieve information on soil moisture from atmospheric information and conventional land use data. Mahfouf [3] attempted the retrieval of surface moisture from observations of screen level air temperature and relative humidity (10 m wind and 2 m temperature and humidity as SYNOPs). Using the Interactions Soil Biosphere Atmosphere (ISBA [4]; NP89 henceforth) land surface scheme, positive conclusions were obtained for an agricultural crop area under clear sky conditions. Two possible approaches were described: a variational algorithm where a cost function is minimized over an assimilation period and a sequential assimilation scheme that consists of a set of predictions and in-static correlations of soil moisture. Both methods were validated against in situ data collected during the HAPEX-MOBILHY [5] field experiment, using a one-column model to represent the interactions between surface processes and the planetary boundary layer structure. Bouttier et al. [6, 7] further argued for the feasibility of estimating both superficial and deep/bulk layer soil moisture using the time evolution of atmospheric temperature and relative humidity near the surface. They also emphasized the fact that this method requires a close relationship between the near surface atmospheric state and the soil moisture.

Calvet et al. [10] also applied a nonvariational inverse scheme for estimation of the bulk soil moisture content using surface variables (either surface soil moisture or temperature). They argued that knowing the atmospheric forcing (especially precipitation) and four to five surface soil moisture measurements over a two-week period were adequate for retrieving the bulk soil moisture by inverting ISBA scheme. Because there is a strong relationship between the deeper layer soil moisture and surface soil moisture, especially when the vegetation cover is in full growth, it is feasible to infer the bulk soil moisture. For assimilating surface temperatures, they realized that satisfactory retrieval is obtainable for only relatively dry conditions.

Based on the same forward model as used by [10], Bouyssel et al. [8] performed a series of variational surface analyses using screen level atmospheric parameters. Their tangent linear analysis identified several interesting topics that inspired the present study. Their study, however, was based solely on synthetic data (i.e., identical twin experiments).

Compared with Ren [11], the new set of experiments described here is more indirect and more demanding because screen level atmospheric measurements are assimilated to infer the initial soil model conditions [3, 8, 12]. In addition to synthetic data, the real Oklahoma Atmospheric Surface Layer Instrumentation System (OASIS) observations also will be assimilated.

Since the data set used and the land surface component of the forward model are identical to those of [11], our description focuses on describing the atmospheric planetary boundary layer scheme and the forward model verification followed by a rather general description of the variational formalism and the construction of the adjoint model. The retrieval experiments assuming data availability of screen level atmospheric parameters are performed and results are analyzed for both synthetic observations and real OASIS measurements.

#### 2. Data and the Forward Model

There are two prerequisites for a successful variational retrieval: an accurate forward model (in terms of physical mechanisms, numerical schemes, and coding) and high quality observational data and a cost function sensitive to the control variables. If the uncertainty in a retrieved variable contributes little to the model-data misfit, as measured by the cost function, that variable simply cannot be effectively retrieved. Thus, effort must first be applied to verify the forward model and rectify any problems related to soil temperature/moisture estimation. During the forward model calibration and backward model inversion, the high quality, high-frequency OASIS measurements are used as the ground truth and as input for retrieval experiments. The second requirement as applied to the retrieval of soil moisture and temperature from atmospheric forecasting errors is that the impact of soil moisture/temperature on near surface observations dominates the impacts of other error sources [6, 7].

2.1. Forward Model Description. The forward model is based on the Advanced Regional Prediction System (ARPS [13–15]). All of our experiments, including forward prediction and 4D-Var retrieval, will be conducted in 1D column mode. This is necessary because, otherwise, a full 3D adjoint code would be required for the atmospheric component of model, and a full blown 4D-Var data assimilation system is involved. This is a major undertaking that is beyond the scope of this study primarily because the associated computational cost would preclude routine implementation.

Almost all land surface models used in atmospheric prediction systems are column based; namely, they do not include horizontal transport of heat or moisture. This is also true of the land surface model in the ARPS that we use. For our study, the vertical boundary layer mixing process is of foremost importance in the atmospheric model, particularly, for the periods when weather is inactive so that horizontal advection is less important. In fact, for the experiments that use observational data (real data experiments), the quiescent atmospheric periods are intentionally chosen, so that our 1D assumption is valid.

The core of the forward model involves solving the coupled energy and water budget equations of the land surface and of the overlying atmospheric boundary layer (Figure 1). Because of the 1D assumption, the hosting system (ARPS) is trimmed to keep a minimum number of relevant components. In all prognostic equations of the atmospheric model, the advection terms are neglected and so are horizontal mixing terms. For momentum equations, the Coriolis terms are retained, and the horizontal pressure gradient terms are expressed in terms of geostrophic winds. For the potential temperature equation and the water vapor equation, only the vertical mixing term is kept. Within the atmospheric boundary layer, the turbulent eddy coefficients for momentum, heat, and moisture are diagnostically prescribed using the PBL model of Hong and Pan [16]. Since the land surface component has been discussed in [11], our discussion focuses mainly on the atmospheric component.



FIGURE 1: Diagram illustrating the LSM-PBL (reduced ARPS) model structure and vertical grid stencil. TOA is top of atmosphere, S signifies scalar quantity,  $z_p$  is physical height, and all other symbols are as defined in the text. This figure is partially adapted from Bouyssel et al. [8].

The PBL is the layer of atmosphere that directly interacts with the land surface. For the retrieval of soil state variables using surface atmospheric observations, the soil component will be coupled with the atmospheric model and the PBL mixing is the most important process. The governing equation reads

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial \overline{u'w'}}{\partial z} + f\left(\overline{v} - v_g\right), \qquad (1a)$$

$$\frac{\partial \overline{\nu}}{\partial t} = \frac{\partial \overline{\nu' w'}}{\partial z} - f\left(\overline{u} - u_g\right), \qquad (1b)$$

$$\frac{\partial \overline{\theta}}{\partial t} = \frac{\partial \overline{\theta' w'}}{\partial z} - \overline{R_{\text{Radiative}}}, \quad (1c)$$

$$\frac{\partial \overline{q}}{\partial t} = \frac{\partial \overline{q'w'}}{\partial z} - \overline{C_{\text{Condensation}}},$$
 (1d)

where *t* is time, *z* is altitude, *u* is the zonal velocity component (eastward positive), v is the meridional velocity component (northward positive), w is vertical velocity, and  $\theta$  is virtual potential temperature, which is defined as  $\theta = \theta_d (1 + \theta_d)$ 0.608q); here  $\theta_d$  is the potential temperature for dry air, and q is the mixing ratio of water vapor. Here  $u_q$  is the eastward component of the geostrophic wind,  $v_q$  is the northward component of the geostrophic wind, f is the Coriolis parameter,  $R_{\text{Radiative}}$  is the radiative cooling/heating rate, and  $\overline{C_{\text{Condensation}}}$  is the sink term due to condensation (varnishes for clear boundary layer). Here the overbar "-" denotes the grid resolvable or mean quantity, while the superscription prime "/" indicates subgrid quantity or perturbation. The vertical turbulent flux of a quantity (q,  $\theta$ , or momentum) is expressed as the covariance between its perturbation and the vertical velocity perturbation. Specifically, u'w' and v'w' are components of the Reynolds stress tensor, respectively, in the directions east and north; and  $\theta' w'$  and q' w' are, respectively, the components of the turbulent heat and moisture fluxes.

The Reynolds stress term is usually parameterized as proportional to the vertical gradient of the mean flow to close the system. The key to a turbulent closure is the determination of the mixing coefficient K which is usually parameterized using PBL parameterization scheme inside the convective boundary layer and is estimated using a subgrid scale turbulence scheme above the convective boundary layer. In this study, for simplicity, K is a diagnostically obtained profile following the work of Hong and Pan [16]. The nonlocal PBL parameterization scheme implemented into ARPS is described in the Appendix.

For our 1D model run, for simplicity, the geostrophic winds for atmospheric levels above the PBL are interpolated from the two subsequent soundings about 12 hours apart and the winds are linearly interpolated to the corresponding model vertical levels. For the layers within the PBL, a constant geostrophic wind profile is assumed, with the wind speed values at the top of the boundary layer. The radiation heating/ cooling processes and the microphysics are retained as in ARPS.

Implementation of this PBL scheme in the ARPS framework is not straightforward because the ARPS equations are first written in a Cartesian coordinate projected onto a plane tangent to or intercepting the Earth's surface. A coordinate transformation into a curvilinear coordinate system is then performed to put the governing equations into an equally spaced computational domain. The special curvilinear coordinate system that ARPS applies is

$$\xi = x,$$
  

$$\eta = y,$$
(2)  

$$\zeta = \zeta (x, y, z).$$

Equation (2) represents a transformation that maps a domain with a vertically stretched grid and an irregular lower boundary to a regular rectangular domain with equal grid spacing in each direction. We call the latter the computational domain. The dynamic equations are discretized in the computational space. Our column run voids the horizontal coordinate transformation and makes the map projection transformation irrelevant or alternatively makes the map projection factor equal to unity. With the definition of  $J_3 = \partial z/\partial \zeta$ ,  $\rho^* = J_3\overline{\rho}$ ,  $u^* = \rho^* u$ , and  $v^* = \rho^* v$ , the governing equations for atmospheric components become

$$\frac{\partial u^*}{\partial t} = f\rho^* \left( v - v_g \right) + \frac{\partial}{\partial \zeta} \left( \overline{\rho} K_{mv} \frac{1}{J_3} \frac{\partial u}{\partial \zeta} \right), \qquad (3a)$$

$$\frac{\partial v^*}{\partial t} = -f\rho^* \left( u - u_g \right) + \frac{\partial}{\partial \zeta} \left( \overline{\rho} K_{mv} \frac{1}{J_3} \frac{\partial v}{\partial \zeta} \right), \quad (3b)$$

$$\frac{\partial \rho^* \theta}{\partial t} = f_r + \frac{\partial}{\partial \zeta} \left( \overline{\rho} K_{H\nu} \frac{1}{J_3} \frac{\partial \theta}{\partial \zeta} \right), \tag{3c}$$

$$\frac{\partial \rho^* q}{\partial t} = \frac{\partial}{\partial \zeta} \left( \overline{\rho} K_{H\nu} \frac{1}{J_3} \frac{\partial q}{\partial \zeta} \right), \tag{3d}$$

$$\frac{\partial p}{\partial \zeta} = \rho^* g \left[ \frac{\theta}{\overline{\theta}} - \frac{p}{\overline{\rho} C_s^2} + \frac{q}{\varepsilon \overline{q}} \right], \tag{3e}$$

where  $C_s$  is the speed of sound,  $\varepsilon = R_d/R_v \approx 0.622$  is the ratio of the gas constants for dry air and water vapor, and  $f_r$  is the radiative forcing.

In ARPS, wind components and the state variables are defined as the sums of the base-state variables (those with an overbar in (3a)-(3e)) and the deviations from the base state (with primes removed). The base state is typically constructed using an external sounding and is assumed to be horizontally homogeneous, time invariant, and hydrostatically balanced.

 $J_3$  is fixed once the elevation and vertical grid setting is set. For all the following numerical experiments, 80 grid points are used in the vertical direction. The vertical grid is stretched from 4 m at the bottom to 396 m at the top level, according to the hyperbolic tangent function given by Eq. 7.3.6 in [13]. The vertical dimension of the simulated domain is ~32 km, deep enough for assuming zero pressure perturbation at the top of atmosphere. This is actually the upper level boundary condition for numerically solving the hydrostatic perturbation pressure equation (3e).

For a description of the fully fledged 3D ARPS grid stencil setting, the reader is referred to [14, 15]. For our discussion, we are concerned only about its vertical grid setting (see left side of Figure 1). It is a reduced Arakawa C-grid [17], where all prognostic scalar variables are defined at the center of the grid box, while the normal velocity components are defined on their respective box surfaces. Other derived variables are evaluated at locations that minimize spatial averaging in the difference operations. According to the variable arrangement relative to the physical boundary, the second level of scalar variables is the first level above ground surface. The vertical gradient of a quantity is evaluated using the first and second layer and is centered on the land surface, where the vertical turbulent heat flux, *H3*, is defined.

Since our column does not support gravity waves, the pressure perturbation equation is further reduced to the hydrostatic equilibrium form. The separation of acoustically active and inactive modes as appears in the split-explicit time integration scheme of 3D ARPS is unnecessary. The horizontal processes (i.e., the Coriolis' force and the pressure gradient force) can be treated explicitly without stability constraints. The leap-frog scheme is used for the time integration. However, for the often large vertical mixing coefficients used by the PBL scheme, vertical turbulent mixing often results in a linear stability constraint on the time step size when treated explicitly, especially when the vertical resolution is high. To overcome this potentially severe restriction on the time step size, the implicit Crank-Nicolson type scheme is used for the vertical mixing so that the time integration is absolutely stable for the mixing terms [14].

In ARPS, the atmospheric and land surface components use different conventions. For example, the equation for potential temperature (see (3c)) is not in the energy form (the heat capacity factor is not shown). Hence, the sensible heat flux output from the land surface model is not of the same quantity as the potential temperature flux at the lower boundary. Similarly, the equation for specific humidity q (see (3d)) is written in mass flux form (vapor mass per unit time per unit area), whereas the land surface model provides latent heat flux (energy per unit time per unit area) at the lower boundary. Due to these specific features of the atmospheric model, proper adjustments must be made before surface fluxes provided by land surface model may be used by the atmospheric component of the model. In addition, the formulation of the land surface model uses differences in temperature, while the eddy turbulent mixing term is expressed in potential temperature form; a transform factor  $(p_0/p)^{0.2856(1-0.23q_v)}$  should also be taken into account, so that

$$\begin{aligned} -\overline{\rho}K_{H\nu}\frac{1}{J_{3}}\frac{\partial\theta}{\partial\zeta}\Big|_{1} &= \overline{\rho}\left(\overline{w'\theta_{\nu}'}\right)_{\rm sfc} = \overline{\rho}C_{H}\left|\overrightarrow{V_{1}}\right|\left(\theta_{\rm sfc}-\theta_{1}\right) \\ &= \frac{H}{C_{p}}\left(\frac{1000}{p}\right)^{0.2856(1-0.23q)}, \\ -\overline{\rho}K_{H\nu}\frac{1}{J_{3}}\frac{\partial q}{\partial\zeta}\Big|_{1} &= \overline{\rho}\left(\overline{w'q'}\right)_{\rm sfc} \\ &= \overline{\rho}H_{u}C_{H}\left|\overrightarrow{V_{1}}\right|\left(q_{\nu,\rm sfc}-q_{\nu,1}\right) = \frac{LE}{L_{\nu}}. \end{aligned}$$
(4)

Here  $C_H$  is surface layer exchange coefficient,  $\overline{V_1}$  is the horizontal velocity vector for the lowest model layer,  $L_{\nu}$  is the latent heat of evaporation, and subscripts "sfc" and "1" identify surface and the first atmospheric model layer quantities, respectively.

The kinematic surface fluxes are given by

$$\overline{\rho} \left( \overline{w'u'} \right)_{\rm sfc} = -\overline{\rho} C_M \left| \overline{V_1} \right| \overline{u_1},$$

$$\overline{\rho} \left( \overline{w'v'} \right)_{\rm sfc} = -\overline{\rho} C_M \left| \overline{V_1} \right| \overline{v_1},$$
(5)

where  $u_1$  and  $v_1$ , respectively, are zonal and meridional components of the vector wind  $\overrightarrow{V_1}$ . Note also that all fluxes are defined as positive when directed upward in the atmospheric model, whereas they are defined positive when contributing to the ground's substrate.

Advances in Meteorology

Symbol	Meaning	Value
$\Delta z_{\min}$	Minimum vertical grid interval	4 m
κ	Karman constant	0.4
Nz	Number of atmosphere layers (ARPS physical layers = $Nz - 3$ )	80
В	A coefficient of proportionality, which is used in describing the nonlocal enhancement effects from large-scale eddies to total flux	7.8
$\Delta t_{\rm big}$	Large time step for model integration	60 s
$\Delta t_{\rm rad}$	Interval for updating radiation process	1800 s
$\Delta t_{\rm sfc}$	Time step for land surface processes	60 s
$\lambda_{0}$	The asymptotic length scale for mixing length (free atmosphere)	30 m
1	Flux mixing depth	200 m
$Km_{\min}/Km_{\max}$	Lower/upper bounding for turbulent eddy coefficients	0.01/1000.0
$Pr_{min}/Pr_{max}$	Lower/upper turbulent Prandtl number	0.5/4.0

TABLE 1: Atmospheric model parameter setting.

In ARPS, the parameterization scheme for estimating  $C_H$  and  $C_M$  together with the original TKE scheme for estimating K in the PBL layer and the force-restore type of soil scheme was tested as in the framework of ARPS against several field experiment data sets. We follow the ARPS estimation of  $C_H$  and  $C_M$  rather than using those in the original Medium Range Forecast (MRF [18], Appendix) code because we believe the consistency of the original soil scheme with the surface layer parameterization is more important than the consistency with K profile in PBL layer.

ARPS diagnoses the PBL depth h based on the virtual potential temperature profile. The top of the PBL is assumed to be at the level where the environmental virtual potential temperature exceeds that of the first level above ground. If the atmosphere is stable right above ground, the PBL depth is set to the thickness of the layer below the first scalar point above ground. In this study, the iteratively obtained h in MRF is used instead, which includes further adjustment to the height of the inversion by thermal and nonlocal processes.

2.2. Verification of Forward Model. Because the forward model is used as a strong constraint, its accuracy affects the accuracy of the retrieval. This section contributes to verifying the forward coupled land surface atmospheric model. For the land surface quantities (i.e., soil temperature, moisture, *LE*, and *H*) the model output is checked against corresponding OASIS measurements. Atmospheric profiles of potential temperature and specific humidity are compared with soundings that are available every 12 hours. Our verification experiments are performed for a variety of situations and the desirable results are obtained. In the following, we present results from one clear sky dry period, August 6–8, 2000. Model forecasts continuously for two days, with a time step of 1 minute for both atmospheric and soil components. The atmospheric model parameters used in this study are listed in Table 1.

The atmospheric component is initialized using sounding of 12 Z on 6 August, 2000. The surface and deep soil moisture values are initialized using the measurements at 5 and 25 cm, respectively. For the selected period, the respective values for



FIGURE 2: Comparisons between modeled (solid lines), original (before the modifications from Ren and Xue [9]) (dotted lines), and observed (thick dot dashed lines) land surface and screen level atmospheric model parameters for August 6–8, 2000 period.

superficial and deep soil moisture are 0.253 and 0.278 m<sup>3</sup> m<sup>-3</sup>, respectively. The surface temperature is initialized using a measurement by an infrared instrument (295.8 K), while the deep soil temperature (298.5 K) is initialized using the preprocessed value according the procedure described in [9].

With the above parameter settings, satisfactory descriptions of the soil temperatures are obtained with our model (Figure 2). The surface temperature is predicted accurately, with maximum errors generally less than 2 K. There lacks apparent phase error and a slight warm bias is present in the prediction that is less than 1 K. The time series for deep soil temperature ( $T_{dp}$ ) indicate that revisions to the forcerestore temperature equations are very necessary. Otherwise, the deep soil temperature will drift upward and finally assumes the same daily average temperature as that of the surface temperature. Particularly satisfying are the accurate predictions of the surface and deep soil temperatures: the maximum model-data differences of 2 K and 0.5 K, respectively, fall well within the measurement error range. Without the modifications introduced to the model by Ren and Xue [9], the errors would be much larger.

The out-of-phase behavior in the simulated surface soil moisture with respect to diurnal atmospheric forcing is believed to be due to more complicated vegetation activity and the hydraulic displacement of the soil water potential [19–21]. It is beyond the capacity of the current force-restore model, which does not include the effects of hydraulic lift. Without implementing the effects of hydraulic lift, the simulated surface soil moisture shows phase error compared with the observations. However, the absolute differences between forecasts and observations are generally small, typically the magnitude of the measurement error. We emphasize the fact that the amplitude of the daily cycle is rather accurately modeled. The simulation of the deep soil moisture is accurate.

The measured and predicted specific humidity is not exactly the same from the starting time, due to the fact that the model is initialized using a 12 Z sounding and the 2 m value is interpolated from the sounding which is ~1g/kg larger than the corresponding measurement. The evolution over time occasionally deviates from the measurements. The downward wiggles (15 Z and 21 Z) corresponding to switching on and off of the stability measured by the bulk Richardson number. Using a value of bulk Richardson number < 0.5 helps reduce these wiggles. It is interesting that the measured specific humidity increases during heating period and decreases during nighttime. The out-of-phase features are connected with the surface soil moisture simulation. The absolute difference is generally less than 3 g/kg and the simulation is satisfactory on daily basis. Whether or not the out-ofphase feature affects our retrieval based on OASIS screen level observations depends critically on the sensitivity of specific humidity to soil moisture content. The retrieval scheme can perform at best to reduce the initial guess errors in initial soil moisture values within the perturbation range which can cause specific humidity bifurcating as large as 3 g/kg for this relatively dry case.

Figure 3 indicates that the simulations of *LE* and *H* are accurate, with peak value differences less than  $50 \text{ W m}^{-2}$  during daytime heating period. During nighttime, the model overestimates sensible heat flux, while it underestimates latent heat flux. However, both magnitudes are small (less than 15 W m<sup>-2</sup>) and well within the instrumental error ranges for such variable measurements.

Accurate prediction of the atmospheric state is also critical for our retrieval experiments. We compare the vertical profiles of atmospheric potential temperature and specific humidity with the corresponding soundings (Figures 4(a)



FIGURE 3: Comparisons between modeled (solid lines) and observed (dot dashed lines) latent and sensible heat fluxes for the 00 Z, August 6–8, 2000 period.

and 4(b)). 12 hours later (00 Z August 7), the modeled and observed potential temperature profiles are close to each other (modeled is slightly warmer) and the difference is generally less than 2 K. 24 hours later, the profiles are still quite similar as to the PBL height, whereas the difference can be as large as 5 K. Most of the errors in the specific humidity profile are near the surface, whereas the prediction is pretty accurate at higher levels. Although the detailed structures in the profile are not predicted after 24 hours, the general trend is quite accurate (<2 g/kg).

The time-height cross sections as plotted in Figure 5 illustrate the daily cycle of eddy diffusivities coefficients (for vapor and thermal) calculated using the nonlocal scheme as implemented in our model. It can be seen that the diffusivity coefficients gradually increase from sunrise up to mid-afternoon and then decrease relatively quickly near sunset. The location of the maximum values of  $K_m$  and  $K_h$  is about 1/3 of the overshoot height. The shapes of K's profiles experience a transition from like one-side distribution to biased two-sided distribution and to normal distribution as vertical stratifications transit from stable to neutral and to convective. Such results are in concert with those of Hong and Pan [16] (see Figure 2 therein).

A sequence of hourly plots (not shown) indicates that a mixed boundary layer is fully developed by 17 Z. Due to the vegetated ground surface, the superadiabatic layer near the surface is not apparent, whereas the negative potential temperature perturbation develops at the top of the boundary layer as a result of entrainment. The depth of the planetary boundary layer is accurately predicted. The nighttime very cold and dry spikes which appear for the potential temperature and mixing ratio simulations, respectively, are



FIGURE 4: The evolution of atmospheric potential temperature (a) and specific humidity (b).



FIGURE 5: Time-elevation cross sections of the eddy diffusivity for thermal calculated with the nonlocal scheme.

successfully eliminated by implementing the flux distribution scheme of ARPS, which distribute the estimated surface radiative cooling caused near surface air potential temperature drops during nighttime into a certain depth. The depth for swapping flux redistribution was set to 200 m, which is clearly shown in the 3 Z profiles. We are pretty sure that the development and collapse of the planetary boundary layer are accurately described by the model because, on 12 Z August 8 (the ending time of forward integration), the soundings for mixing ratio and potential temperature are rather close to our modeled profiles.

The accurate description of daytime growth and nighttime collapse of the PBL as well as the evolution of the PBL height is important not only because PBL height exhibits a strong daily cycle driven by the surface heating but also because it grows in conjunction with the entrainment of the warm (higher in potential temperature) and dry air from the overlying free atmosphere. The entrainment fluxes, which serve as the main connection between the coupled land surface PBL system and the overlying free atmosphere, warm and dry the mixed layer. Considering that the similarity theory is applied on the mixed layer property and the corresponding surface property, this process of entrainment directly affects not only the PBL energy and moisture distribution but also ultimately the surface energy and moisture budget.

The results confirm the accuracy of the forecast prediction model, which is essential for the retrieval experiments to be conducted based on this forward model and its adjoint. Also, the reader is cautioned that, due to technical difficulties in adjoint coding, some parameterizations in the PBL component as well as in land surface component will be altered. However, all such modifications are acceptable only when they do not alter the overall forecasting performance as shown in this section.

### 3. Definition of the Cost Function and Verification of the Optimization System

For the soil state variable retrieval problem using screen level atmospheric observations as the constraint, the cost function is constructed as a quadratic function of screen level atmospheric forcing variables. They are related through model dynamics to the land surface model variables. We define  $U(t_0) = (T_{sfc}^0, T_{dp}^0, w_{sfc}^0, w_{dp}^0, w_{canop}^0)^T$ , and

$$J(U) = \sum_{i=1}^{N} \left[ \left( \frac{q_{\text{air},i}^{o} - q_{\text{air},i}^{f}(U)}{\sigma_{q}^{o}} \right)^{2} + \left( \frac{T_{\text{air},i}^{o} - T_{\text{air},i}^{f}(U)}{\sigma_{T}^{o}} \right)^{2} \right] + \left( U - \overline{U} \right)^{T} \sigma_{a}^{-2} \left( U - \overline{U} \right),$$

$$(6)$$

where N is the number of observations during an assimilation period; T represents temperature (superscript T means matrix transpose), with subscripts "air" indicating the screen level of air. Here q is water vapor specific humidity, with the notational conventions as those for temperature.  $\sigma_0$ 's represent the relative confidence accredited to each observation and prediction pair, which are typically the standard deviation of the observational error.  $\overline{U}$  is the prior estimate of the initial condition vector and  $\sigma_a^2$  is the prior error covariance. The three  $\sigma$ 's weight the model-measurement misfit and prior estimate misfit, respectively, based on the uncertainty involved in each component. For example, if the measurement error was very small and the prior knowledge of the initial conditions was poor, then we would expect the estimation procedure to change greatly the initial guess or prior estimate in an effort to fit the model output closer to the measurements.

When many high quality observations are available in the assimilation period or assimilation window, the background term (i.e., the last term in (6)) is less valuable since the situation of retrieval problem tends to be overdetermined. This tends to be the case for our 1D problem because of the relatively small degree of freedom. Also, for land surface variables, we usually lack proper background analyses *a priori* (e.g., usually no *a priori* soil moisture information). In this study, the measurement standard deviation is assumed to be 1 K and that for specific humidity is 1 g/kg. Apparently, the ratio of these two  $\sigma$ 's plays a key role in the convergence of the retrieval procedure.

Suppose the vector form of the physical model (see (3a)-(3e)) is  $\partial U/\partial t = F(U, \alpha)$ , where  $\alpha$  represents model parameters. Following the treatment in [22], one obtains the adjoint model as

$$\frac{\partial \lambda}{\partial t} = -\frac{\partial F}{\partial U}\lambda - 2\frac{\partial T_{\text{air}}}{\partial U}\sigma_T^{o-2}\left(T_{\text{air},i}^o - T_{\text{air},i}^f\right) 
- 2\frac{\partial q_{\text{air}}}{\partial U}\sigma_q^{o-2}\left(q_{\text{air},i}^o - q_{\text{air},i}^f\right).$$
(7)

The adjoint model, (7), is integrated backward in time to obtain  $\lambda$ ; the gradient of the objective function with respect to the initial conditions can then be computed as

$$\frac{\partial J}{\partial U(t_0)} = -\lambda_0 + 2\left(U - \overline{U}\right)\sigma_a^{-2}.$$
(8)

The right-hand side of (7) determines how the information in  $T_{\rm air}$  and  $q_{\rm air}$  is assimilated. The second term on the right-hand side of (8) is trivial, since we have no proper prior estimate of the initial condition vector as argued previously.

We did not discretize (7) to obtain the adjoint system. Rather, we here follow a code-to-code approach [23, 24], which is especially suitable for sophisticated models with complicated boundary conditions and on/off switches. The adjoint model (ADM) constructed this way is also strictly consistent with the forward model. We also chose to handcode the tangent linear model (TLM) and ADM because hand-coded adjoint is more humane and more efficient than machine produced one. More importantly, it helps the researcher to become more familiar with the operation of both the forward and backward systems, which comprises the prerequisite for good research. For example, in the forward scheme, the second-order Runge-Kutta method requires two calls of the subroutine tendency, which provides the righthand side of the soil equations in [1]. In the adjoint coding, the synchronizing of these two half-time steps (for soil model integration) is necessary. We solved this dilemma by interim information storage.

Also, in the adjoint coding (not shown here for reasons of clarity), several continuous alternatives for stomatal resistance  $R_s$  parameterization are available to cope with the above-mentioned differential problems. Similar problem exists for the PBL parameterization for the bulk Richardson number. In calculating the bulk Richardson number, the convective contribution to the wind speed is omitted when the virtual potential temperature difference between the air and surface is less than 0.1 K. We adopted this criterion mainly because there is a singular point in the adjoint of the PBL component resulting from the parameterization of the convective component of wind speed. The gradient becomes infinite when the PBL experiences a transition from a stable to an unstable regime. We thus do not calculate the gradient for a temperature difference less than 0.1 K between the two temperatures used for the parameterization.

In formality, adjoint model is exactly a transform of the tangent linear model. The ADM code also is implemented strictly in the reverse order of that in TLM for the main driver routine and its subroutines. Using Fortran code as an example, in each statement, sensitivities of right-hand sided variables are expressed as the common left-hand sided variable. For example, if the TLM has a statement as var0 =  $c_1 \times var1 + c_2 \times var2$ , these become two corresponding statements in the ADM: del\_var1 = del\_var1 +  $c_1 \times del_var0$  and del\_var2 = del\_var2 +  $c_2 \times del_var0$ . The adjoint variables (del\_x) accumulate the sensitivity information within the assimilation window and eventually become the gradients of the cost function with respect to the corresponding control variables (x). Backward modeling (adjoint model construction), unlike the forward modeling (forecasting model and its

tangent linear model), has no direct physical laws to guide the researcher; thus, finishing the adjoint code is merely half-way to finishing the adjoint system implementation. Thorough and objective testing of the components is indispensable. This involves the verification of the adjoint code and then the entire 4D-Var system.

Because ADM is based on TLM or, alternatively, the former is the matrix transform of the latter, the error sources of TLM are also the error sources of ADM. However, functions of TLM and ADM models are essentially different. TLM is used to approximate the evolution of a small perturbation on forward model. The time integration is forward. However, the backward integration of ADM is to unveil the sources of the perturbation. The adjoint variables corresponding to output of the forward model are actually inputs to the backward integration. Reflecting on the hand coding, for each statement, the perturbation of the right-hand terms is expressed in the form of perturbation on the left-hand side term.

To verify the correctness of the adjoint model, we use the inner product of the solutions to the direct and adjoint equations, a time invariant quantity [25] (Eqs. (4.4) and (4.8) in [25]). Consider that the magnitudes of different control variables are quite different. For example, soil moisture is less than 1.0, whereas soil temperatures are around 300. If all variables are perturbed simultaneously, the small quantities (considering only the numerical values, not the units) tend to be overshadowed by the larger quantities in the inner product. Therefore, we performed component-bycomponent verifications. Using double precision compiling, the relative difference is generally less than  $10^{-8}$ .

The function of ADM in a 4D-Var system is to estimate the gradient of the cost function with respect to the control variables. The correctness of the gradient value is verified against the Taylor expansion of cost function around specific control variables (p32 in [26]):

$$I(\alpha) = \frac{J(U(t_0) + \alpha h) - J(U(t_0))}{\alpha h^T \nabla J(U(t_0))} = 1 + o(\alpha), \quad (9)$$

where  $\alpha$  is a small real number, *h* is a random perturbation vector which can be generated by using the Fortran library function, and  $h^T$  is the transpose of *h*.

For values of  $\alpha$  which are small but not very close to the machine zero, one should expect a value of  $I(\alpha)$  approaching 1 linearly for a wide range of magnitudes of  $\alpha$ . We performed an experiment using coupled run of the land-atmosphere model and its adjoint with a cost function defined as in (6). The numerical results are shown in Figure 6. It is clearly seen that for values of  $\alpha$  between  $10^1$  and  $10^{-4}$  unit values for  $I(\alpha)$  are obtained. The curve deviates from unity for different reasons as  $\alpha \rightarrow 10^{-5}$  and  $\alpha \rightarrow 10^2$ . The former is due to truncation error, while the latter is because the small perturbation assumption is violated. The correctness of the vector product have therefore been verified. Interested readers can find a more sophisticated and efficient method in evaluating the TLMs using Taylor expansions [27].



FIGURE 6: Verifications of the correctness of the gradient calculation by ADM. This experiment is performed with the micrometeorological observations of 00 Z, 4-6 August, 2000 at Norman's OASIS super site. Function *I* as defined in (9) is plotted.

#### 4. Numerical Experiments

We use a variational method to determine the land surface model initial states for which the model forecast best fits screen level atmospheric observations of potential temperature and specific humidity within the assimilation window. This is achieved through minimizing a cost function J(U) (see (6)) defined as a quadratic difference of the screen level air temperature and specific humidity. J is an implicit function of the control variable U. U contains the soil temperature  $(T_{\rm sfc} \text{ and } T_{\rm dp})$ , the soil water contents ( $w_{\rm sfc} \text{ and } w_{\rm dp}$ ), and the canopy interception/dew  $w_{\rm canp}$ . Because  $w_{\rm canp}$  is insignificant in the selected periods, U has only four effective dimensions. For OSSE (and actual measurements), the convergence criterion for the optimization procedure is when the cost function decreases by 3 (2) orders of magnitude or the iteration number exceeds 30 (50).

Using OSSE experiments, we discuss the retrieval scheme robustness to the magnitudes of initial guess errors and to magnitudes of Gaussian noise. In practical application, the shorter the assimilation window the better the performance. However, because of model error and measurement inaccuracy, sufficient information content is required for reliable retrieval. The optimal window length is apparently a function of sampling (data ingest) frequency. Some techniques for improving convergence of the optimization will also be addressed for OASIS data assimilation. With these issues in mind, we designed a set of numerical experiments (see the following).

Retrieval experiment design for coupled model system is as follows:

OSSE. We have a 3-hour nature couple run starting from 00 Z, 6 August 2000; synthetic (simulated) observations of atmospheric potential temperature (PT) and specific humidity  $(q_v)$ were sampled every 5 minutes and used by the retrieval experiments;  $U^{\text{True}} = (295.8 \text{ K}, 298.5 \text{ K}, 0.253 \text{ m}^3 \text{m}^{-3}, 0.278 \text{ m}^3 \text{m}^{-3}, 0.0 \text{ m}^2 \text{m}^{-2})$ :

- (a) Initial guess errors exist in all control variables:  $U^{i} = (300.8 \text{ K}, 303.5 \text{ K}, 0.303 \text{ m}^{3}\text{m}^{-3}, 0.328 \text{ m}^{3}\text{m}^{-3}, 0.0 \text{ m}^{2}\text{m}^{-2})$  or  $\delta U^{T} = (5 \text{ K}, 5 \text{ K}, 0.05 \text{ m}^{3}\text{m}^{-3}, 0.05 \text{ m}^{3}\text{m}^{-3}, 0.05 \text{ m}^{3}\text{m}^{-3}, 0.05 \text{ m}^{2}\text{m}^{-2})$  (OSSEa).
- (b)  $\delta U^T$  in OSSEa was halved in addition to several other representative experiments (*OSSEb*).
- (c) There is scheme resistance to Gaussian noise in PT and q<sub>v</sub> (OSSEc):
  - (1) Gaussian noise series of zero mean and different standard deviations added to obs.
  - (2)  $T_{\rm air}$ .
  - (3)  $q_{v}$ .
  - (4) Both  $T_{air}$  and  $q_{\nu}$  (guarantee that noise series added on  $T_{air}$  and  $q_{\nu}$  are not correlated).

OASIS. OASIS observations of PT and  $q_v$  available every 5 minutes are used by the retrieval experiments:

- (a) Control run: initial guess control variable *U*<sup>*t*</sup> the same as OSSE runs (*OASISa*).
- (b) Effects of data availability (OASISb):
  - (1) data frequency: 5 min, 30 min, 1 hr, 3 hr, and 6 hr;
  - (2) data availability period: running over the daily cycle with hourly interval and with assimilation window length from 3 to 24 hrs.
- (c) Identifying the optimal assimilation window length (*OASISc*).

4.1. Observing Simulation System Experiment (OSSE) Retrieval. We simulate observations every minute of temperature and specific humidity at 2 m. The retrieval scheme may use the data less frequently. For the first retrieval experiment (OSSEa), we assume initial guess errors in all control variables (i.e.,  $\delta U^T = (5 \text{ K}, 5 \text{ K}, 0.05 \text{ m}^3 \text{m}^{-3}, 0.05 \text{ m}^3 \text{m}^{-3}, 0 \text{ m}^2 \text{m}^{-2})$ ), all significant compared to typical existing model errors for such quantities. The reference run is a coupled 3-hour nature run starting from 12 Z on 6 August 2000 with true initial states:  $T_{\text{sfc}}^0 = 295.8 \text{ K}$ ,  $T_{\text{dp}}^0 =$ 298.5 K,  $w_{\text{sfc}}^0 = 0.253 \text{ m}^3 \text{m}^{-3}$ , and  $w_{\text{dp}}^0 = 0.278 \text{ m}^3 \text{m}^{-3}$ . Synthetic observations of atmospheric potential temperature and specific humidity are sampled every five minutes and used by the retrieval experiments.

The first set of experiments is performed assuming perfect observations. The retrieval procedure is shown in Figure 7 within the assimilation window. We see that the retrieval method is able to realize a good analysis, that is, to approach closely the real initial states in just seven steps. The retrieved initial values are  $U^r = (295.6876 \text{ K}, 298.98 \text{ K}, 0.2539 \text{ m}^3 \text{m}^{-3}, 0.2739 \text{ m}^3 \text{m}^{-3})^T$ . Compared with the true values, the differences are insignificant (-0.12 K, 0.4 K, 0.0009 m<sup>3</sup>m<sup>-3</sup>, 0.005 m<sup>3</sup>m<sup>-3</sup>)<sup>T</sup> (Figure 7(a)). As a result (of the successful adjustments to the initial guess errors on the

states) the cost function decreased significantly to about  $2.54 \times 10^{-4}$  at step seven (Figure 7(b)). The trajectory fitness within the assimilation window is overall very satisfactory for all four state variables (Figure 7(c)) and surface fluxes (Figure 7(d)). The initial hot moist guess signifies a drastic overestimation to *LE* but underestimation to *H*. The retrieved curves are indiscernible to the synthetic true trajectories.

Using the retrieved initial states, we made extended forecasting of 24 hours (Figure 8). The improvements of using retrieved over initial guessed states are significant, especially the case for daytime heating period (12–24 Z), as measured by the statistics as summarized in Tables 2 and 3. MAE means maximum difference of two time series; MBE means mean bias error; and RMS means root mean biased error. For example, the RMS errors for screen level air temperature, specific humidity, *LE*, and *H* are reduced from 3.39 K to 0.15 K, from 6.22 g/kg to 0.05 g/kg, from 17.42 W m<sup>-2</sup> to 0.62 W m<sup>-2</sup>, and from 13.2 W m<sup>-2</sup> to 0.52 W m<sup>-2</sup>, respectively. The improvements resulting from using retrieved over initial guess state variables are also shown in the trajectory fitness of the state variables themselves (columns 4–7 in Tables 2 and 3).

4.1.1. Robustness to Initial Guess Magnitudes. In addition to the case discussed above, we also tried using the initial guess control variables halved (the second experiment listed in Table 4). The retrieval is a very successful one in that the retrieved states are close enough to the true states ( $\delta U$  =  $(0.4 \text{ K}, 0.5 \text{ K}, -0.002 \text{ m}^3 \text{m}^{-3}, 0.012 \text{ m}^3 \text{m}^{-3}, 0 \text{ m}^2 \text{m}^{-2})^T).$ Retrieval from using further reduced initial guess state variables (e.g., the third experiment listed in Table 4) is always successful in accordance with our tangent linearization assumption. The fourth experiment listed in Table 4 is a rather free guess of 300 K soil temperatures and 35% wetness on soil moisture. The corresponding retrieval is a successful one. Actually, the scheme is very robust for initial guess on soil temperature (270–320 K). To investigate the scheme robustness to soil moisture initial guess errors, we performed two retrieval experiments with the same soil temperature initial guess of 300 K (the last two as listed in Table 4). For the very dry initial guess case, the adjustment to deep soil moisture is insignificant, whereas for the wet initial guess, the initial guess error in surface soil moisture cannot be effectively removed. This can be explained through analyzing the structure of the cost function (note: its production is rather time consuming), especially the  $w_{\rm sfc}$ - $T_{\rm sfc}$  cross section.

Around and below wilting point, the isoligns tend to be vertical and the cost function is no longer sensitive to initial deep soil moisture value. Above saturation point, the isoligns tend to orient horizontally and the cost function becomes insensitive to initial value of surface soil moisture values. Generally speaking, for this relatively dry period, the separation of sensitivity between  $T_{\rm sfc}$  and  $w_{\rm sfc}$  is not so strong that it is easy to get both quantities successfully retrieved.

4.1.2. Robustness to Gaussian Noise. Another important issue is the noise level that the scheme can tolerate. To this end, the observations created with the coupled 1D land surface



FIGURE 7: Reference OSSE retrieval experiment for assimilating screen level atmospheric measurements. (a) The convergence of the control variables to the "true" values (dashed lines). (b) Reduction in cost function. (c) Comparison between state trajectories resulting from prior guess initial conditions and retrieved values. (d) is the surface *LE* and *H* trajectories from initial guess and retrieval as compared with synthetic truth.

atmospheric model are corrupted with realizations of a Gaussian noise with zero mean and different standard deviation magnitudes on  $T_{\rm air}$  and  $q_{\rm air}$ . Using the reference retrieval experiment configuration (OSSEa), we performed retrieval

experiments with Gaussian noise of different magnitudes (0.1, 0.5, and 2.0 K for temperature and 1, 2, and 5 g/kg for specific humidity) added to the air temperature and mixing ratio time series within the assimilation window, respectively, and

TABLE 2: Some statistics between Initial guess and synthetic observations within the forecasting period.

	$T_{\rm air}$ (K)	$q_{\nu}$ (kg/kg)	$T_{\rm sfc}$ (K)	$T_{\rm dp}$ (K)	$w_{\rm sfc} ({ m m}^3{ m m}^{-3})$	$w_{\rm dp} ({ m m}^3{ m m}^{-3})$	$H (Wm^{-2})$	$LE (Wm^{-2})$
RMSE	3.39	$6.22 \times 10^{-3}$	3.165	0.366	$5.4 \times 10^{-2}$	$4.73 \times 10^{-2}$	13.2	17.42
MBE	-3.01	$6.16 \times 10^{-3}$	-2.639	0.06	$5.4 \times 10^{-2}$	$4.74 \times 10^{-2}$	3.09	6.2
MAE	6.55	$8.16 \times 10^{-3}$	7.027	1.0	$5.4 \times 10^{-2}$	$4.75 \times 10^{-2}$	53.5	86.43

TABLE 3: Some statistics between retrieved and observations within the forecasting period.

	$T_{\rm air}$ (K)	$q_v$ (kg/kg)	$T_{\rm sfc}$ (K)	$T_{dp}$ (K)	$w_{\rm sfc} ({ m m}^3{ m m}^{-3})$	$w_{\rm dp} ({ m m}^3{ m m}^{-3})$	$H (Wm^{-2})$	$LE (Wm^{-2})$
RMSE	0.15	$5.45 \times 10^{-5}$	0.163	0.268	$7.81 \times 10^{-4}$	$4.19 \times 10^{-3}$	0.52	0.62
MBE	-0.14	$5.12 \times 10^{-5}$	0.162	-0.267	$7.15 \times 10^{-4}$	$4.19 \times 10^{-3}$	-0.21	0.21
MAE	0.204	$1.08 \times 10^{-4}$	0.19	0.297	$1.23 \times 10^{-3}$	$4.20 \times 10^{-3}$	1.83	2.54

TABLE 4: Initial states retrieval using synthetic data.

Initial guess control variable			ariable	Initial DMS arror in A		# it pooded	Retrieved initial condition			
$T_{\rm sfc}$	$T_{dp}$	$w_{ m sfc}$	$w_{\mathrm{dp}}$		$J_{\text{final}}/J_0$	# It. Inceded	$T_{\rm sfc}$	$T_{dp}$	$w_{ m sfc}$	$w_{dp}$
300.8	303.5	0.303	0.328	2.27	$2.52 \times 10^{-4}$	6	295.7	298.9	0.254	0.274
298.3	301.0	0.278	0.303	1.07	$1.71 \times 10^{-2}$	8	296.2	299.0	0.251	0.290
297.0	299.8	0.265	0.291	0.52	$4.26 \times 10^{-2}$	9	296.2	298.8	0.252	0.287
300	300	0.35	0.35	3.05	$1.04 \times 10^{-3}$	15	295.5	297.4	0.243	0.302
300	300	0.2	0.2	2.88	$8.11 \times 10^{-2}$	12	293.2	296.0	0.255	0.199
300	300	0.4	0.4	1.37	0.4	4	297.5	298.2	0.4	0.378



FIGURE 8: Comparison between state trajectories resulting from prior guess initial conditions and retrieved values for the true forecasting period (following the assimilation window).

jointly. We compared retrieval experiments with noise added to both screen level parameters and those without noise corruption. Generally, the higher the noise level, the worse the retrieved initial states and consequently trajectory fitness (as measured by the cost function). Also, the retrieval of deep soil moisture and soil temperature seems to be affected more severely by the corrupted air temperature observations, whereas the bad effects from corrupted specific humidity observations are felt by all control variables. For Gaussian noise added to the 2 m air temperature, when the standard deviation is smaller than 4 K, convergence is not apparently affected. For Gaussian noise added to specific humidity, standard deviation smaller than 3 g/kg has no apparent effects on convergence.

4.2. Retrieval Experiments with Real Data. The first real data assimilation experiment (see Figure 9) uses exactly the same initial guess as OSSEa but instead assimilate the OASIS observed screen level air potential temperature and specific humidity. Similar to the corresponding OSSE experiment (Figure 7), all control variables get significantly improved (Figure 9(a)). Primarily, due to the model errors, the decrease of the cost function is not as apparent as the corresponding OSSE. Upon convergence, there is still ~36% error as measured by the cost function (Figure 9(b)). Meanwhile the trajectories of the prognostic variables are significantly updated (Figure 9(c)) within the assimilation window due to the accurate retrieval of the control variables ( $\delta U^T = (0.8 \text{ K}, 1.1 \text{ K})$  $0.003 \text{ m}^3 \text{m}^{-3}$ ,  $0.007 \text{ m}^3 \text{m}^{-3}$ ,  $0 \text{ m}^2 \text{m}^{-2}$ )). Consequently, the improvement in surface fluxes is significant (Figure 9(d)). We further examine the improvements for the one-day long extra forecasting period that follows the assimilation window period (Figure 10). The improvements are all apparent and



FIGURE 9: Reference OASIS retrieval experiment for assimilating screen level atmospheric measurements. (a) The convergence of the control variables to the true values (dashed lines). (b) Reduction in cost function. (c) Comparison between state trajectories resulting from prior guess initial conditions, retrieved values, and OASIS observations. (d) is similar to (c) but for *LE* and *H*.

significant, especially for the daytime heating period. The RMS error for *H* is reduced from  $13.2 \text{ W m}^{-2}$  to  $<5 \text{ W m}^{-2}$ . For *LE*, RMS error is reduced from  $17.3 \text{ W m}^{-2}$  to merely  $4.7 \text{ W m}^{-2}$ . More detailed statistics are given in Tables 5 and 6.

Unlike the corresponding OSSE, the retrieval of  $q_v$  is good only on the daily or longer time scale basis. Although the mean value and the amplitude are correctly retrieved, there is still significant phase error associated with the retrieved time series. As stated above, forward model needs to include more

TABLE 5: Some statistics between retrieved and observations within the forecasting period.

	$T_{\rm air}$ (K)	$q_v$ (kg/kg)	$T_{\rm sfc}$ (K)	$T_{\rm dp}$ (K)	$w_{\rm sfc} ({ m m}^3{ m m}^{-3})$	$w_{\rm dp} \ ({\rm m}^3 {\rm m}^{-3})$	$H (\mathrm{Wm}^{-2})$	$LE (Wm^{-2})$
RMSE	0.41	$6.59  imes 10^{-4}$	0.546	0.23	$5.51 \times 10^{-3}$	$7.0 \times 10^{-3}$	4.68	4.71
MBE	0.34	$-6.5 \times 10^{-4}$	0.41	-0.22	$-5.51 \times 10^{-3}$	$-7.0 \times 10^{-3}$	-0.2	-1.3
MAE	1.06	$1.3 \times 10^{-3}$	1.62	0.42	$5.86 \times 10^{-3}$	$7.0 \times 10^{-3}$	18.7	29.6

TABLE 6: Some statistics between initial guess and OASIS observations within the forecasting period.

	$T_{\rm air}$ (K)	$q_v$ (kg/kg)	$T_{\rm sfc}$ (K)	$T_{\rm dp}$ (K)	$w_{\rm sfc} ({\rm m}^3{\rm m}^{-3})$	$w_{dp} (m^3 m^{-3})$	$H (Wm^{-2})$	$LE (Wm^{-2})$
RMSE	3.39	$6.22 \times 10^{-3}$	3.16	0.4	$5.41 \times 10^{-2}$	$4.74 \times 10^{-2}$	13.2	17.3
MBE	-3.02	$6.17 \times 10^{-3}$	-2.64	0.06	$5.42 \times 10^{-2}$	$4.74\times10^{-2}$	3.1	6.2
MAE	6.55	$8.16  imes 10^{-3}$	7.0	1.0	$5.46  imes 10^{-2}$	$4.75\times10^{-2}$	53.5	86.4



FIGURE 10: Comparison between state trajectories resulting from prior guess initial conditions and retrieved values for the true forecasting period following the assimilation window.

sophisticated vegetation processes to improve the simulation of the daily cycle.

4.2.1. Influence of the Assimilation Window Length. For OSSE experiments without noise corruption to the simulated observations, there is no apparent benefit in using longer

TABLE 7: Effects of assimilation window length on retrieval-control variables.

Assimilation window	Retrieved initial condition						
length (hours)	$\delta T_{sfc}$ (K)	$\delta T_{dp}$ (K)	$\delta w_{sfc}$ (K)	$\delta w_{dp}$ (K)			
1	3.75	3.8	0.0287	0.037			
2	3.8	3.7	0.0288	0.039			
6	0.75	0.95	0.005	0.004			
12	-2.19	-1.94	0.012	0.024			
18	-2.7	-2.4	0.014	-0.03			
24	-2.26	-1.93	0.006	0.025			

assimilation window length (>6 hours for 5 min ingesting frequency). Retrieval experiments with real OASIS observations are performed for six assimilation window lengths: 1, 2, 6, 12, 18, and 24 hours. The performance of the retrieval system, as measured by the closeness of the retrieved initial states to true states ( $U^{\text{retrieval}} - U^{\text{true}}$ ), can be seen from Table 7. For assimilation window less than 12 hours, increasing of assimilation window is beneficial. Choosing assimilation window longer than 3 hours can steadily reduce the initial guessed error by more than one order of magnitude (considered significant reduction for real data assimilation). Further increasing the assimilation window length from 6 to 12 hours is less effective in further reducing the initial guess error. More interestingly, using too long assimilation window (e.g., 18 hours) sacrifices the retrieval. This degrading trend is consistent from 12 hr-36 hr assimilation window length and becomes not practical beyond 36-hour assimilation window length. This signifies contesting between model error and the information content.

Another severe hindrance of using too long assimilation window is the computational cost associated with both forward model integration and the backward adjoint integration. For the sake of efficiency, we suggest assimilation window of about six hours for this relatively moist period and vegetated surfaces. From Table 7, for this data ingesting frequency (5 minutes), using 6-hour assimilation seems a reasonable choice. The retrieved temperature values can be within 1 degree of the true (0.75 and 0.95 K, resp.) and the retrieved soil moisture is even close (0.005 and 0.004 m<sup>3</sup> m<sup>-3</sup>, resp.).



FIGURE 11: Relationship between the accuracy in retrieval and the assimilation window length.

For the shared forecasting period of 12 Z 7 August-12 Z 8 August, we compare three assimilation window lengths: 1 hr, 6 hr, and 24 hr. The 6 hr assimilation window gives the most desirable retrieval results in this case. The RMS errors are 0.42 K for air temperature time series as compared with the OASIS observations,  $6.35 \times 10^{-4}$  kg kg<sup>-1</sup> for specific humidity, 0.83 K for skin temperature, 0.23 K for deep soil temperature,  $7.89 \times 10^{-3} \text{ m}^3 \text{ m}^{-3}$  for superficial soil moisture,  $1.08 \times 10^{-2} \text{ m}^3 \text{ m}^{-3}$  for deep soil moisture,  $11.7 \text{ W} \text{ m}^{-2}$  for sensible heat flux, and 15.6  $W m^{-2}$  for latent heat flux, respectively. The RMS errors using too short (1-hour) assimilation window are much worse, 3.29 K for air temperature,  $3.09 \times 10^{-3}$  kg kg<sup>-1</sup> for specific humidity, 5.09 K for skin temperature, 2.08 K for deep soil temperature,  $3.81\times 10^{-2}\,\text{m}^3\,\text{m}^{-3}$  for superficial soil moisture,  $3.42 \times 10^{-2} \text{ m}^3 \text{ m}^{-3}$  for deep soil moisture,  $61.1\,W\,m^{-2}$  for sensible heat flux, and  $84.2\,\tilde{W}\,m^{-2}$  for latent heat flux, and so are those using too large assimilation window length (24-hour), 0.61 K for air temperature, 9.37  $\times$  $10^{-4}$  kg kg<sup>-1</sup> for specific humidity, 1.05 K for skin temperature, 0.47 K for deep soil temperature,  $1.16 \times 10^{-2} \text{ m}^3 \text{ m}^{-3}$  for superficial soil moisture,  $2.37 \times 10^{-2} \text{ m}^3 \text{ m}^{-3}$  for deep soil moisture,  $19.1 \text{ W} \text{ m}^{-2}$  for sensible heat flux, and  $24.3 \text{ W} \text{ m}^{-2}$ for latent heat flux. Other aspects of the comparison can be seen from Figure 11. In Figure 11, we plotted the time series

of all the control variables, the latent and sensible heat fluxes, and the screen level air temperature and humidity resulting from using the final retrieved initial land surface conditions. Measured by all three statistical indices (MAE, MBE, and RMS), the 6-hour assimilation window length performs the best. For example, the mean bias error for *LE* is  $5.3 \text{ W m}^{-2}$ , significantly less than  $62.7 \text{ W m}^{-2}$  from using a 1-hour assimilation window and 20 W m<sup>-2</sup> from using a 24-hour assimilation window. The maximum difference is only  $45 \text{ W m}^{-2}$  for *LE*, whereas it can be as large as  $190 \text{ W m}^{-2}$  when using a 1-hour assimilation window length.

Using assimilation window lengths from 4 to 24 hours and initial guess control variables as in the control run, we addressed another related issue about influences of the assimilation frequency. For each assimilation window length, we tested data ingestion frequency of 5–60 minutes with increment of 5 minutes. A subset of this experiment (assimilating the observations every 10, 15, 30, and 60 minutes) is listed in Table 8. For each combination of assimilation window length and data ingesting frequency, we provided the final retrieved surface temperature and moisture values as indices for evaluating the performance of the retrieval.

As expected, for every assimilation window length, less frequent data ingesting by the adjoint integration is much more efficient. Generally speaking, the longer the assimilation window is, the more deserving to perform less frequent assimilation. For example, 18-hour assimilation window can tolerate assimilating the data every 60 minutes, whereas 4-hour assimilation can at best assimilate data grid every 10 minutes. When assimilation frequency becomes too low (<every 10 minutes in this case), the retrieved states will deviate from the true states significantly and even fail to converge. The apparent benefit for performing assimilation at lesser frequency is for efficiency of the scheme. Updating the forcing every minute makes the scheme less practicable for real application. For example, for some initial guesses, it takes 2 hours for a single column run of the retrieval scheme for a 24-hour assimilation window and ingesting the observations every 5 minutes.

For a given assimilation window length, it seems that the more frequent data ingesting yields better retrieval. However, as assimilation becomes less frequent and the assimilation is long, the uncertainty increases. For example, for 18-hour window length, using hourly data ( $T_{\rm sfc} = 297.02$  K) performs better than using 30-minute data ( $T_{\rm sfc} = 299.5$  K).

As mentioned above, as data ingesting becomes less frequent, the optimal assimilation window length tends to increase. For example, 15-minute data supports an optimal window length of around 18 hours. There is no edge for using window lengths between 18 and 24 hours since all retrieval experiments arrive close to  $T_{\rm sfc} = 300.8$  K,  $T_{\rm dp} = 300.2$  K,  $w_{\rm sfc} = 0.24$  m<sup>3</sup> m<sup>-3</sup>, and  $w_{\rm dp} = 0.30$  m<sup>3</sup> m<sup>-3</sup>.

An interesting aspect we would like to mention is that the degrading effects on retrieval from less frequent data availability are quite different for different components of the control variables. For shorter assimilation windows (e.g., 4-hour and 6-hour in Table 8), the retrieval of  $T_{\rm sfc}$  is deteriorating fast (298.65 K using 5-minute data and 300.79 K when using

	Window length (hour)								
Assim. freq. (minute)	4		6		18		24		
	$T_{\rm sfc}$	$w_{ m sfc}$	$T_{ m sfc}$	$w_{ m sfc}$	$T_{ m sfc}$	$w_{ m sfc}$	$T_{ m sfc}$	$w_{\rm sfc}$	
5	298.65	0.2633	296.54	0.256	293.1	0.267	293.5	0.259	
10	300.79	0.3025	294.76	0.2602	298.05	0.245	290.08	0.2835	
15			301.48	0.3025	300.76	0.239	300.78	0.2426	
30					299.50	0.251	299.53	0.220	
60					297.02	0.260			

TABLE 8: Comparisons of surface soil moisture and temperature retrieved by using different assimilation window lengths and data ingesting frequencies.

10-minute data for 4-hour window length). However, for longer assimilation window lengths (e.g., 24-hour), the data frequency has no major effects on the retrieval of surface soil moisture value. This can be explained by analyzing the sensitivities of screen level air potential temperature and humidity time series to variations in initial  $T_{\rm sfc}$  and  $w_{\rm sfc}$ (Figure 12). In Figures 12(a) and 12(b), initial value of  $w_{\rm sfc}$ is varied down to  $0.1526 \text{ m}^3 \text{ m}^{-3}$  and up to  $0.3526 \text{ m}^3 \text{ m}^{-3}$ , with increment of  $0.05 \text{ m}^3 \text{ m}^{-3}$ . Figure 12(a) indicates clearly that the responses from screen level potential temperature are significant at all phases of a daily cycle, whereas the response from specific humidity to initial value of  $w_{\rm sfc}$  is more salient during nighttime after 18 Z and is insignificant at the beginning 6 hours. To see the sensitivity of screen level atmosphere to initial value of  $T_{\rm sfc}$ , we provided Figures 12(c) and 12(d), where initial value of  $T_{\rm sfc}$  is varied down to 285.79 K and up to 305.79 K, with increment of 5 K. In contrast to the sensitivities to  $w_{\rm sfc}$ , all sensitivities are contrasted in the first couple hours for variations in initial value of  $T_{\rm sfc}$ . This explains why decreasing the assimilation frequency sacrifices the retrieval of initial  $T_{\rm sfc}$ , especially when window length is short (for 4-hour window length, 298.65 K versus 300.79 K in Table 8). This also partly explains why the retrieval of initial  $w_{\rm sfc}$  is always pretty accurate for longer assimilation windows.

We also find that there is little benefit in using assimilation window longer than 24 hours. Except the apparent reason of model inaccuracy, nonlinearity may also play a role in scarifying the usefulness of the adjoint technique, which is based on tangent linearization as mentioned above.

4.2.2. Sampling Strategy. The general setting is identical to the OASIS control run. For the following set of retrieval experiments, we, however, assume data availability only at certain periods of a daily cycle. We tested assimilation window length from 1 hour to 12 hours, and for each window length we tested centering it at different times of a day. Results from applying different window lengths unanimously point to the daytime period being more informative for the state retrieval purpose. This finding is in accordance with [2]. We also find that using a window length longer than 4 hours and centered at local noon can be as informative as using data within all the daily cycle for retrieval of the land surface states. Certainly, this critical window length of 4 hours is connected with the relatively high data ingesting frequency (5 minutes each). We thus tested less frequent data ingesting and enlarged assimilation window length (and the overdetermination requirement is always satisfied) and find that this above assertation still holds qualitatively; that is, there exists a window length centered at local noon significantly shorter than 24 hours and is nearly equally informative as the whole daily cycle period.

We now understand better why the retrieval significantly improved when using assimilation window longer than three hours for our experiments testing effects from assimilation window length. The assimilation windows all start at 12 Z, 6 August 2000. It is a heating up stage of the daily cycle. Except that longer assimilation window length contains more observations, using assimilation of 1 hour is also closer to the nighttime period than using, say, 4-hour assimilation window.

#### 5. Summary and Conclusions

Modern numerical weather prediction requires the initial conditions of the land surface/soil-vegetation model. The lack of routinely available measurements of land surface variables motivated the development of schemes that infer land surface information from routinely measured variables.

Based on 4D-Var strategy, a retrieval system for the initial state of a soil-vegetation model is developed which assimilates operationally available screen level atmospheric measurements. Coupled soil-vegetation and the atmospheric models and their adjoint have to be constructed. To avoid the need for developing a full 4D-Var system of a 3D coupled system, a 1D column simplifying assumption is made. Radiation and vertical mixing are the main processes involved. The vertical mixing is described using the nonlocal MRF PBL scheme, and soil-vegetation processes are described by a two-layer force-restore model implemented in ARPS.

The sensitivities of the measured variables to the retrieved (control) variables are first studied via forward simulations. Such simulations are verified against OASIS measurements. An important correction is made to the deep layer soil temperature prediction equation to enable the successful retrieval of the deep layer temperature and improve that of others.

Unlike previous studies [2, 8], both OSSE and real data retrieval experiments were performed, assuming two different types (skin temperature or screen level atmospheric variables) of measurements, for a dry period during year 2000



FIGURE 12: Responses of screen level potential temperature (a and c) and specific humidity (c and d) to variations of initial surface moisture (a and b) and surface temperature (c and d). The statistics are printed in the plots for each corresponding curve (relative to the reference run).

at the Norman OK OASIS site. The retrieval system is able to recover the correct values of soil state variables in most cases. The robustness of the systems is assessed by testing different magnitudes of initial guess errors and different measurement errors. The impact of the assimilation window length and data availability on the retrieval products is also examined. Compared with [28] and several other oversimplified treatments of land surface processes [2, 29], our explicit treatment of vegetation processes is critical to the successful real data assimilation for deep soil moisture content for partially vegetated surfaces. In contrast, Xu and Zhou [29] discussed a linear regression method for retrieving bulk soil moisture contents from soil temperature measurements, based solely on the soil heat capacity's dependency on soil moisture contents. Results from [2] are also limited to the bare ground case.

Our retrieval system does not perform effectively for very wet circumstances. Ren [1] performed a series of retrieval experiments for such conditions. The most important finding by Ren [1] was that the measurement (cost function) was shown to be insensitive to the value of deep soil moisture for very wet conditions. The marked sensitivity contrast between  $T_{\rm sfc}$  and  $w_{\rm sfc}$  prevents all control variables from being simultaneously retrieved successfully without proper preconditioning.

In this study, by restricting the application of our data assimilation system to only weather situations when impact of soil moisture on near surface atmospheric conditions is dominant, for example, to situations with clear sky, low wind speed, and strong radiative input, it is certain that the analyzed soil moisture/temperature corrections are not caused by an attempt to compensate for model deficiencies in parameterization. For more sophisticated systems, the analysis will be more difficult to obtain. Significant efforts will be put into that direction in the future.

In reality, the atmospheric measurements are available less frequently (usually 6 hours for SYNOPs) than required by our retrieval scheme. Further improvement to the land surface scheme is required to reduce the phase error in the simulation of the specific humidity daily cycle. Through forward model improvements, we believe the robustness of the retrieval scheme when supplied with real observations will be improved and can work more effectively with less frequent observations.

The adjoint of a much more sophisticated land surface scheme, the Community Land Model (CLM 3.5), is now available [30]. The present study, performed a decade ago, still has significant value in guiding the operational use of adjoint-based variational schemes such as the variational data assimilation system of CLM, specifically in the specification of the assimilation window (its temporal location and length) and the ingest frequency for observational data. Although computationally very efficient, the adjoint-based variational data assimilation procedure involves coding of the adjoint model, which must strictly correspond to the forward model. Subsequent changes made to forward model require essential modification to the adjoint model. This is a technical disadvantage and explains the prevalence of other forms of data assimilation and optimal parameterization, especially the Ensemble Kalman Filter [31-37]. For these recent studies, our results on information redundancy and the observational error covariance provide useful guidance for such operational applications.

#### Appendix

For any quantity *c* (can be u, v,  $\theta$ , or *q*), the nonlocal approach for vertical mixing, based on [18], can be expressed as

$$\frac{\partial \overline{c}}{\partial t} = \frac{\partial}{\partial z} \left[ K_c \left( \frac{\partial \overline{c}}{\partial z} - \gamma_c \right) \right]. \tag{A.1}$$

This equation is a form of closure of the diffusion (first term) term in (1a)–(1d). Since local mixing uses  $K_c(\partial \bar{c}/\partial z)$  to estimate vertical turbulent fluxes, under conditions that the mean quantity is the same for the two selected level, zero flux will result. Chances are actual bulk fluxes are not zero (may flip signs between the two levels). Equation (A.1) is proposed also to deal with this shortcoming with local schemes. This also suggested that parameter  $\gamma_c$  should be a bulk property of the total convective boundary layer not that of a quantity which represents a subdomain of it. As in [16, 18, 38, 39], the diffusivity coefficient for momentum is formulated as

$$K_m = \kappa w_{s,m} z \left( 1 - \frac{z}{h} \right)^2, \qquad (A.2)$$

where  $\kappa$  is Karman constant,  $w_{s,m}$  is a characteristic turbulent velocity scale for scalar transport, and *h* is the depth of the PBL. The mixed layer velocity scale is represented as

$$w_{s,m} = \frac{u_*}{\phi_m}.$$
 (A.3)

Following formulation of [40] (Eqs. (A1.9) & (10)), here  $\phi_m$  is evaluated at the top of the surface layer z = Z, with the stability criteria expressed as the covariance of virtual potential temperature perturbation and vertical velocity evaluated at the surface. For example,  $(\overline{w'\theta'_{\nu}})_{\text{sfc}} \leq 0$  signifies unstable condition. The PBL depth *h* in (A.2) is iteratively obtained through

$$h = R_{B,cr} \frac{\overline{\theta_{\nu,1}} |\overline{\mu}(h)|^2}{g\left(\overline{\theta_{\nu}}(h) - \overline{\theta_{\nu s}}\right)},$$
(A.4)

where  $R_{B,cr}$  is critical Richardson number (set to 0.3 in this study);  $\overline{\theta}_{\nu,1}$  is the virtual potential temperature at the lowest model level and should be taken consistent with  $\Delta z$  in the following definition of bulk Richardson number:

$$R_B = g\Delta \overline{\overline{\theta_v}} \Delta z \left\{ \overline{\overline{\theta_v}} \left[ (\Delta \overline{u})^2 + (\Delta \overline{v})^2 \right] \right\}^{-1}.$$
 (A.5)

Here  $\theta_{vs}$  is the appropriate temperature near the surface (about 10 m, or near the top of the surface layer as in [38]) defined as

$$\overline{\theta_{vs}} = \overline{\theta_{v,1}} + B \frac{\overline{(w'\theta_v)}_{sfc}}{w_{s,m}}.$$
(A.6)

The right-hand side of (A.6) is the scaled virtual temperature excess near the surface, usually bounded (3 K in [16]) in case surface wind is very weak. Here *B* is a coefficient of proportionality (taken as 7.8 in [16]), which is used in describing the nonlocal enhancement effects from large-scale eddies to total flux. Consider

$$\gamma_c = B \frac{\overline{w'\theta_v'}}{w_{s,m}h}.$$
 (A.7)

Similarly, the dimensionless vertical temperature gradient given by [40] is

$$\phi_{h} = \begin{cases} \left(1 - 15\frac{Z}{L}\right)^{-0.5}, \quad \left(\overline{w'\theta_{\nu}'}\right)_{\rm sfc} \le 0\\ \left(1 + 5\frac{Z}{L}\right), \quad \left(\overline{w'\theta_{\nu}'}\right)_{\rm sfc} > 0. \end{cases}$$
(A.8)

The eddy diffusivity for temperature  $(K_h)$  and moisture  $(K_q)$  is estimated from the relationship of the turbulent Prandtl number:

$$P_r = \frac{\phi_h}{\phi_m} + \kappa B \frac{Z}{h} = \frac{K_m}{K_h}.$$
 (A.9)

Following [16], for the free atmosphere, a local-*K* approach [41] is utilized:  $K = l^2 f_{stb} |\partial \overline{u} / \partial z|$ , where the mixing length  $l = \kappa z \lambda_0 / (\kappa z + \lambda_0)$ ; here  $\lambda_0$  is the asymptotic length scale, and  $f_{stb}$  is the stability function. The stability function  $f_{stb}(\text{Ri}_g)$  is represented in terms of the local gradient Richardson number  $\text{Ri}_g = (g/\overline{\theta_v})(\partial \overline{\theta_v}/\partial z)[(\partial \overline{u}/\partial z)^2 + (\partial \overline{v}/\partial z)^2]^{-1}$  at a given level. Here,  $\theta_v$  is the virtual potential temperature. Computed  $\text{Ri}_g$  is usually numerically bounded to prevent unrealistically unstable regimes (-100 in [16]).

#### **Competing Interests**

The authors declare that they have no competing interests.

#### Acknowledgments

This work was supported by DOT-FAA Grant no. NA17RJ1227-01 and NSF Grants nos. ATM9909007 and ATM0129892. The authors thank Dr. J. Brotzge for making available the OASIS data set and for very helpful discussions on the related issues. The lead author's Ph.D. thesis under Professor M. Xue forms the basis of this research. Professor Mervyn Lynch (Curtin University) proofread the entire paper.

#### References

- D. Ren, 4DVAR retrieval of prognostic land surface model variables using advanced regional prediction system [Ph.D. thesis], School of Meteorology, University of Oklahoma, Norman, Okla, USA, 2004.
- [2] G. Boni, F. Castelli, and D. Entekhabi, "Sampling strategies and assimilation of ground temperature for the estimation of surface energy balance components," *IEEE Transactions on Geoscience* and Remote Sensing, vol. 39, no. 1, pp. 165–172, 2001.
- [3] J.-F. Mahfouf, "Analysis of soil moisture from near-surface parameters: a feasibility study," *Journal of Applied Meteorology*, vol. 30, no. 11, pp. 1534–1547, 1991.
- [4] J. Noilhan and S. Planton, "A simple parameterization of land surface processes for meteorological models," *Monthly Weather Review*, vol. 117, no. 3, pp. 536–549, 1989.
- [5] J. C. Andre, J. P. Goutorbe, and A. Perreir, "HAPEX MOBILHY: a hydrologic atmospheric experiment for the study of water budget and evaporation flux at the climatic scale," *Bulletin of the American Meteorological Society*, vol. 67, pp. 138–144, 1986.

- [7] F. Bouttier, J.-F. Mahfouf, and J. Noilhan, "Sequential assimilation of soil moisture from atmospheric low-level parameters. Part II: implementation in a mesoscale model," *Journal of Applied Meteorology*, vol. 32, no. 8, pp. 1352–1364, 1993.
- [8] F. Bouyssel, V. Casse, and J. Pailleux, "Variational surface analysis from screen level atmospheric parameters," in *HIRLAM* 4 Workshop on Variational Analysis in Limited Area Models, Toulouse, France, February 1998.
- [9] D. Ren and M. Xue, "An improved force-restore model for landsurface modeling," *Journal of Applied Meteorology*, vol. 43, pp. 1768–1782, 2004.
- [10] J.-C. Calvet, J. Noilhan, and P. Bessemoulin, "Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: a feasibility study based on field measurements," *Journal of Applied Meteorology*, vol. 37, no. 4, pp. 371–386, 1998.
- [11] D. Ren, "Adjoint retrieval of prognostic land surface model variables for an NWP model: assimilation of ground surface temperature," *Central European Journal of Geosciences*, vol. 2, no. 2, pp. 83–102, 2010.
- [12] U. Callies, A. Rhodin, and D. P. Eppel, "A case study on variational soil moisture analysis from atmospheric observations," *Journal of Hydrology*, vol. 212-213, no. 1–4, pp. 95–108, 1998.
- [13] M. Xue, K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster, ARPS Version 4.0 User's Guide, Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Okla, USA, 1995.
- [14] 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," *Meteorology and Atmospheric Physics*, vol. 75, no. 3-4, pp. 161–193, 2000.
- [15] 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," *Meteorology and Atmospheric Physics*, vol. 76, no. 3-4, pp. 143–165, 2001.
- [16] S.-Y. Hong and H.-L. Pan, "Nonlocal boundary layer vertical diffusion in a medium-range forecast model," *Monthly Weather Review*, vol. 124, no. 10, pp. 2322–2339, 1996.
- [17] A. Arakawa and V. Lamb, "Computational design of the basic dynamical processes of the UCLA general circulation model," in *Methods in Computational Physics*, J. Chang, Ed., pp. 174–246, Academic Press, New York, NY, USA, 1977.
- [18] I. B. Troen and L. Mahrt, "A simple model of the atmospheric boundary layer; sensitivity to surface evaporation," *Boundary-Layer Meteorology*, vol. 37, no. 1-2, pp. 129–148, 1986.
- [19] M. M. Caldwell, T. E. Dawson, and J. H. Richards, "Hydraulic lift: consequences of water efflux from the roots of plants," *Oecologia*, vol. 113, no. 2, pp. 151–161, 1998.
- [20] C. M. Ishikawa and C. S. Bledsoe, "Seasonal and diurnal patterns of soil water potential in the rhizosphere of blue oaks: evidence for hydraulic lift," *Oecologia*, vol. 125, no. 4, pp. 459– 465, 2000.
- [21] D. Ren, M. Xue, and A. Henderson-Sellers, "Incorporating hydraulic lift into a land surface model and its effects on surface soil moisture prediction," *Journal of Hydrometeorology*, vol. 5, no. 6, pp. 1181–1191, 2004.

- [22] D. Ren, "A modified Richards equation, its adjoint, and a new perspective on land data assimilation," *Meteorology and Atmospheric Physics*, vol. 92, no. 1-2, pp. 25–32, 2006.
- [23] R. Giering and T. Kaminski, "Recipes for adjoint code construction," ACM Transactions on Mathematical Software, vol. 24, no. 4, pp. 437–474, 1998.
- [24] X. Zou, F. Vandenberghe, M. Pondera, and Y. H. Kuo, "Introduction to adjoint techniques and the MM5 adjoint modeling system," Tech. Rep. NCAR/TN-435-STR, 1997.
- [25] F.-X. LeDimet and O. Talagrand, "Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects," *Tellus*, vol. 38, no. 2, pp. 97–110, 1986.
- [26] Z. Wang, Variational data assimilation with 2-D shallow water equations and 3-D FSU global spectral models [Ph.D. thesis], College of arts and sciences, The Florida State University, Tallahassee, Fla, USA, 1993.
- [27] M. E. Gharamti and I. Hoteit, "Complex step-based low-rank extended Kalman filtering for state-parameter estimation in subsurface transport models," *Journal of Hydrology*, vol. 509, pp. 588–600, 2014.
- [28] D. Entekhabi, H. Nakamura, and E. Njaku, "Retrieval of soil moisture profile by combined remote-sensing and modeling," in *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*, B. J. Choudhury, Y. H. Kerr, E. G. Njoku, and P. Pampaloni, Eds., pp. 485–498, 1995.
- [29] Q. Xu and B. Zhou, "Retrieving soil water contents from soil temperature measurements by using linear regression," *Advances in Atmospheric Sciences*, vol. 20, no. 6, pp. 849–858, 2003.
- [30] C. M. Hoppe, H. Elbern, and J. Schwinger, "A variational data assimilation system for soil-atmosphere flux estimates for the Community Land Model (CLM3.5)," *Geoscientific Model Devel*opment, vol. 7, no. 3, pp. 1025–1036, 2014.
- [31] J. L. Anderson, "An adaptive covariance inflation error correction algorithm for ensemble filters," *Tellus Series A: Dynamic Meteorology and Oceanography*, vol. 59, no. 2, pp. 210–224, 2007.
- [32] H. J. Hendricks Franssen and W. Kinzelbach, "Real-time groundwater flow modeling with the Ensemble Kalman Filter: joint estimation of states and parameters and the filter inbreeding problem," *Water Resources Research*, vol. 44, no. 9, Article ID W09408, 2008.
- [33] C. M. Dechant and H. Moradkhani, "Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting," *Water Resources Research*, vol. 48, no. 4, Article ID W04518, 2012.
- [34] H. Moradkhani, C. M. Dechant, and S. Sorooshian, "Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method," *Water Resources Research*, vol. 48, no. 12, Article ID W12520, 2012.
- [35] J. Qin, S. Liang, K. Yang, I. Kaihotsu, R. Liu, and T. Koike, "Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal," *Journal of Geophysical Research Atmospheres*, vol. 114, no. 15, 2009.
- [36] Y. Shi, K. J. Davis, F. Zhang, C. J. Duffy, and X. Yu, "Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: a synthetic experiment," *Water Resources Research*, vol. 50, no. 1, pp. 706–724, 2014.
- [37] J. Rasmussen, H. Madsen, K. H. Jensen, and J. C. Refsgaard, "Data assimilation in integrated hydrological modeling using

ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance," *Hydrology and Earth System Sciences*, vol. 19, no. 7, pp. 2999–3013, 2015.

- [38] A. Holtslag and B. A. Boville, "Local versus nonlocal boundarylayer diffusion in a global climate model," *Journal of Climate*, vol. 6, no. 10, pp. 1825–1842, 1993.
- [39] A. Holtslag and C. Moeng, "Eddy diffusivity and countergradient transport in the convective atmospheric boundary layer," *Journal of the Atmospheric Sciences*, vol. 48, no. 14, pp. 1690–1698, 1991.
- [40] A. J. Dyer, "A review of flux-profile relationships," *Boundary-Layer Meteorology*, vol. 7, no. 3, pp. 363–372, 1974.
- [41] J.-F. Louis, "A parametric model of vertical eddy fluxes in the atmosphere," *Boundary-Layer Meteorology*, vol. 17, no. 2, pp. 187–202, 1979.









The Scientific World Journal







Journal of Earthquakes



Submit your manuscripts at http://www.hindawi.com





Advances in Meteorology

International Journal of Mineralogy



Journal of Climatology



Journal of Geological Research





International Journal of Atmospheric Sciences



Advances in Oceanography



Applied & Environmental Soil Science



International Journal of Oceanography



Journal of Computational Environmental Sciences