Assimilation of surface temperature in a land-surface model

Praveen Kumar, University of Illinois at Urbana-Champaign
Amy L. Kaleita, University of Illinois at Urbana-Champaign
Assimilation of surface temperature in a land-surface model

PRAVEEN KUMAR & AMY L. KALEITA
Environmental Hydrology and Hydraulic Engineering, Department of Civil and Environmental Engineering, University of Illinois, Urbana, Illinois 61801, USA
e-mail: kumarl@uiuc.edu

Abstract An extended Kalman filter (EKF) scheme, for the assimilation of near-surface temperature observations in a hydrological model, is developed. The formulation is based on the modification of the diffusion equation of heat flux into the ground. Both model and measurement uncertainties are incorporated. It is found that, in addition to the model error, the accuracy in specification of the initial error covariance has an important bearing on the performance of the assimilation scheme. An inadequate specification can result in decreased rather than enhanced model performance. Study suggests that an "equilibrium" error covariance, arrived at by allowing the model to run with an arbitrary value for a long enough time such that the resulting perturbations subside, can capture the basic correlation structure between the different layers. An arbitrarily scaled equilibrium error covariance, to capture the large initial uncertainty, provides significantly improved performance.

Key words AVHRR; data assimilation; error covariance; land-surface model; SGP97; surface temperature

INTRODUCTION

Among the several model variables that are involved in hydrological and climatological studies, accuracy of surface temperature plays a crucial role in determining the predictive capability of the model. It influences the partitioning of incoming radiant energy into ground, sensible and latent heat fluxes. Outgoing long-wave flux is also a function of the surface temperature.

Often reliable observations of near-surface temperature can be obtained from various satellite instruments, such as AVHRR (Advanced Very High Resolution Radiometer) using the split-window technique (Price, 1984; Sugita & Brutsaert, 1993). These observations can be assimilated into the land-surface models to improve their predictive capability. In this paper we discuss the application of an extended Kalman filter algorithm for the assimilation of near-surface temperature measurements in a land-surface model (LSM) (Bonan, 1996).

DATA DESCRIPTION

The model is driven using data collected during the Southern Great Plains 1997 Experiment (SGP97) (Jackson et al., 1999). SGP97 was a field experiment carried out in the sub-humid environment of Oklahoma (USA) from 18 June to 17 July 1997 (Julian days 169–198). Meteorological and soil temperature measurements were obtained from US Agricultural Research Service (ARS) Micronet stations in the Little
Washita River watershed. In particular, observations from site 151 at the western edge of the Little Washita basin are used in this study.

RESULTS

Starting with an arbitrary initial condition the LSM model is driven from 10 May 1997 (Julian Day 130) at 5-min time steps to provide a spin-up period. The results (LSM) of the last 24 days of the run, for which observations are available, are shown in Fig. 1 along with corresponding actual observations (obs). As seen from this figure, the unmodified LSM gives predictions with an exaggerated amplitude in the top layer when compared to the observed data. While the model and the observations are in relatively close agreement for the nighttime phase of the diurnal cycle, the model estimates are considerably higher than observations during the daytime. Furthermore, the LSM overpredicts the soil temperature in lower layers compared to observations (note that observations corresponding to model layers 4, 5 and 6 are not available).

In order to validate the extended Kalman filter algorithm, studies using synthetic observations are performed. The top layer output values of soil temperature from the unmodified LSM run are corrupted with 0.5% white noise. These are then used as top layer “observations” for the EKF mode. The purpose of using these synthetic observations is to see whether or not the assimilated values in the other five layers converge to those in the unmodified LSM run.

Two situations are considered. First, an arbitrary initial error covariance is specified in the diagonal elements of the matrix and zero elsewhere. The simulation results are shown in Fig. 2(a). One can see that while the top layer behaves the same as in the unmodified run with a slight equilibration period in which some degree of fluctuation occurs in the predictions, the deeper into the column one goes, the greater the magnitude of the initial fluctuations and the longer the equilibration time. The reason for the initial fluctuations in assimilated temperatures is inadequate specification of the error covariance. The diagonal matrix would indicate that errors in one layer are correlated only with errors in that same layer from one time step to the next. Because the temperature profile propagates through the layer with time, it is reasonable to expect that the correlation structure is in fact more complicated. The fluctuations occur while this correlation structure is being numerically established during the assimilation period.

In order to test this proposition, in the second simulation we specify the equilibrium error covariance from the previous run, arbitrarily scaled by a factor of 100, as the initial guess. The equilibrium values are obtained by using the values of the error covariance after 60 days of model run when all the fluctuations had died out. The use of the equilibrium error covariance enables us to numerically capture the correlation structure between the different layers while the scaling factor of 100 indicates a large initial uncertainty. The results are displayed in Fig. 2(b). We see that the use of the “equilibrium” error covariance does not produce any initial fluctuations in the assimilated temperatures. This scaled equilibrium error covariance is used as the initial guess for the assimilation of the observed temperature as described next.

Next, the EKF assimilation is run using actual top-layer observations. An estimation error of 0.5% (roughly 1.5 K) is specified. First, observations at half-hour
Assimilation of surface temperature in a land-surface model

Fig. 1 Validation using the SGP97 dataset of model and assimilated temperatures at (a) half-hour and (b) 24-h (right) time steps.

increments are used. The results, shown in Fig. 1(a), show noticeable improvement in the assimilated temperatures in the second and the third layer. However, the difference in the top layer between the unmodified LSM and EKF assimilation is nearly undetectable. This may at first seem unexpected, since the observational information is being injected into the top layer. However, the fact that the top layer shows minimal response to the assimilation is most likely a function of the way the LSM is set up. In this model, the behaviour of the top layer is tightly constrained by the surface energy balance component of the model, which does not seem to be significantly impacted by
the assimilation. Despite the minimal response in the top layer, there is considerable change in the response of the lower layers when using the EKF scheme.

Second, the observed data are pared down to a 24-h time step in order to simulate the best-case temporal scale of a polar orbiting satellite such as AVHRR. Results for the 1-day data assimilation are shown in Fig. 1(b). Here we again observe that the temperatures in the lower layers converge to the observations, albeit more slowly than the half-hour assimilation scenario. However, the top layer response seems to be still constrained by the surface energy balance computations.
SUMMARY

In order to improve the soil-temperature profile predictions in land-surface models, an assimilation scheme using the extended Kalman filter is developed. This formulation is based on the discretized diffusion equation of heat transfer through the soil column. The scheme is designed to incorporate the knowledge of the uncertainties in the model and the measurement. Model uncertainty is estimated by quantifying the model drift from observations when the model is initialized using the observed values. The model error plays an important role in the evolution of the error covariance when there are missing observations.

In addition to the model error, the initial error covariance has a significant influence on the performance of the assimilation scheme. It is shown that an inaccurate initial value for the error covariance can actually diminish the predictive capabilities of the model. When an arbitrary diagonal matrix is used as the initial error covariance in this study, the results show unreasonable amounts of fluctuations before equilibrating. This is clearly an undesirable effect. However, when the model is permitted to run for a certain amount of time and the resulting equilibrium error covariance, scaled to account for larger initial error, is used as the initial condition for a restarted run, the fluctuations are almost nonexistent. This is a consequence of the fact that the correlation structure between different layers is more complicated than can be described by a simple diagonal matrix.

In the case where measurements are available every 30 min, the lower layers respond much more rapidly to the infusion of observed data than in the case where the measurements occur every day. For the case of 24-h observations, the lower layers respond considerably more slowly. The success of the assimilation in constraining the model temperatures and preventing a drift has significant implications for both short-term and seasonal-to-interannual predictions.

While 24-h observations are the best-case scenario from polar orbiting satellites, due to factors such as cloud contamination, a longer interval between observations is probably more likely. If actual AVHRR surface temperature estimates were to be used as observations, it is likely that the frequency of good observations would be less than one every 24 h. Alternatively, data from a geostationary satellite such as GOES, which provides more-or-less continuous coverage, could be used. The trade-off, however, is that the spatial resolution of geostationary satellites currently in wide use is lower than that of AVHRR.

Acknowledgements Support for this project has been provided by the NSF Graduate Student Fellowship and NASA Grants NAGW-5247 and NAG5-7170.

REFERENCES


