Hydrologic Data Assimilation with the Ensemble Kalman Filter

ROLF H. REICHLE,* DENNIS B. MCLAUGHLIN, AND DARA ENTEKHABI

Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 28 December 2000, in final form 22 March 2001)

ABSTRACT

Soil moisture controls the partitioning of moisture and energy fluxes at the land surface and is a key variable in weather and climate prediction. The performance of the ensemble Kalman filter (EnKF) for soil moisture estimation is assessed by assimilating L-band (1.4 GHz) microwave radiobrightness observations into a land surface model. An optimal smoother (a dynamic variational method) is used as a benchmark for evaluating the filter's performance. In a series of synthetic experiments the effect of ensemble size and non-Gaussian forecast errors on the estimation accuracy of the EnKF is investigated. With a state vector dimension of 4608 and a relatively small ensemble size of 30 (or 100; or 500), the actual errors in surface soil moisture at the final update time are reduced by 55% (or 70%; or 80%) from the value obtained without assimilation (as compared to 84% for the optimal smoother). For robust error variance estimates, an ensemble of at least 500 members is needed. The dynamic evolution of the estimation error variances is dominated by wetting and drying events with high variances during drydown and low variances when the soil is either very wet or very dry. Furthermore, the ensemble distribution of soil moisture is typically symmetric except under very dry or wet conditions when the effects of the nonlinearities in the model become significant. As a result, the actual errors are consistently larger than ensemble-derived forecast and analysis error variances. This suggests that the update is suboptimal. However, the degree of suboptimality is relatively small and results presented here indicate that the EnKF is a flexible and robust data assimilation option that gives satisfactory estimates even for moderate ensemble sizes.

1. Introduction

Near-surface soil moisture is a key variable in the atmospheric and hydrologic models that are used to predict weather and climate. Since soil moisture controls the partitioning of moisture and energy fluxes at the land surface, it has an important influence on the hydrologic cycle over timescales ranging from hourly to interannual. Land surface fluxes in turn affect the evolution of vertical buoyancy in the atmospheric column and also affect the baroclinicity that develops in the horizontal plane (Pan et al. 1995; Paegle et al. 1996). The formation and growth of clouds as well as the evolution of precipitating weather systems over land are affected by surface fluxes and surface soil moisture (Shaw et al. 1997). In fact, the timescale of soil moisture anomalies is at least on the order of several days, which is the forecast-lead horizon of operational weather forecasts.

At seasonal to interannual timescales, predictability of climatic variables such as precipitation is dependent on the land surface boundary conditions of the climate system. Koster et al. (2000) show that over the United States and other large continental regions soil moisture rivals sea surface temperature in explaining the variance in seasonal precipitation anomalies. Increasingly soil moisture and the memory associated with it are recognized to have important roles in the feedback mechanisms that intensify and prolong climate anomalies.

Despite the importance of soil moisture in weather and climate prediction there are currently no operational networks of in situ sensors that provide data suitable for these applications. Since such networks are logistically infeasible and prohibitively expensive, the focus has turned to remote sensing techniques that provide additional information about the land surface at large scales. In particular the L-band (1.4 GHz) microwave brightness temperature of the land surface is correlated with surface soil moisture because of the sharp contrast between the dielectric constants of water and soil minerals (Njoku and Entekhabi 1995).

Interpretation of remotely sensed passive and active microwave measurements is complicated by the effects of canopy microwave optical thickness, surface microroughness, and physical temperature. Remote sensing measurements are only one of many data sources that provide valuable information about soil moisture. Precipitation, soil texture, topography, land use, and a variety of meteorological variables influence the spatial distribution and temporal evolution of soil moisture. We can gain additional information from a coupled model of the soil–vegetation–atmosphere system that relates the measured variables to one another and to soil moisture. Yet uncertainties in the forcing data, the heterogeneity of the land surface at various scales, and the nonlinear nature of land–atmosphere interactions limit our ability to accurately model and predict the state of the land surface and the associated fluxes.

Modern data assimilation theory provides methods for optimally merging the information from uncertain remote sensing observations and uncertain land model predictions (Errico 1999; Errico et al. 2000). Among the prior work on large-scale soil moisture assimilation are studies by Bouttier et al. (1993) and Rhodin et al. (1999) who assimilate low-level air temperature and relative humidity to estimate soil moisture. This approach

aims at improving numerical weather prediction and treats soil moisture as a tuning parameter. Houser et al. (1998) focus on the four-dimensional assimilation of in situ observations and soil moisture retrievals. The weak-constraint variational method of Reichle et al. (2001b) yields near-optimal estimates of the land surface states from direct assimilation of microwave observations. Reichle et al. (2001a) prove the concept of optimal downscaling for the case where soil moisture estimates are required at scales smaller than the scale of the microwave observations. They also show that soil moisture can be satisfactorily estimated even if quantitative precipitation estimates are not available.

In this paper we examine the feasibility of using the ensemble Kalman filter (EnKF) for soil moisture data assimilation. The EnKF is an attractive option for land surface applications because (i) its sequential structure is convenient for processing remotely sensed measurements in real time, (ii) it provides information on the accuracy of its estimates, (iii) it is relatively easy to implement even if the land surface model and measurement equations include thresholds and other nonlinearities, and (iv) it is able to account for a wide range of possible model errors. On the other hand, the EnKF relies on a number of assumptions and approximations that may compromise its performance in certain situations.

The EnKF and variants have been successfully applied to meteorological and oceanographic problems of moderate complexity in small- to medium-sized domains (Evensen and van Leeuwen 1996; Houtekamer and Mitchell 1998; Lermusiaux 1999; Madsen and Canizares 1999; Keppenne 2000). Hamill et al. (2000) provide an excellent discussion of the state of the art of ensemble forecasting and assimilation methods in the meteorological and oceanographic context. The models of geophysical flow used in most of these studies are chaotic in nature and typically have dominant modes that can grow rapidly within a certain subspace. Most such models also have an attractor and sample only a small subdomain of their phase space (Anderson and Anderson 1999). This greatly increases the potential to successfully apply ensemble filtering methods. By contrast, typical land surface models are dissipative in nature. Perturbations in the initial conditions tend to die out after a certain time rather than amplify. Consequently, the soil moisture ensemble filtering problem has certain distinctive aspects that merit closer investigation.

This paper evaluates the performance and computational burden of the EnKF for a synthetic experiment. As a benchmark for the EnKF we use a variational method that solves the optimal smoothing problem. We begin in section 2 with a brief review of the EnKF. The benchmark variational method is discussed in the cited references, including (Reichle et al. 2001b). In section 3 we briefly describe the land model and the setup of the synthetic experiments we use to investigate design issues. In section 4 we discuss the results of these experiments and compare the EnKF with the variational method. We conclude in section 5 with a summary of major findings.

5. Summary and conclusions

In this paper, we discuss the application of the ensemble Kalman filter to hydrologic data assimilation and in particular to the estimation of soil moisture from L-band microwave brightness temperature observations. We also compare the performance of the EnKF to an optimal smoother (weak-constraint variational algorithm). Both methods are applied to the same problem and use identical state and measurement equations, error statistics, and synthetically generated measurements. We conclude that with relatively few ensemble members the EnKF yields reasonable soil moisture estimates. For a state vector dimension of 4608 and a relatively small ensemble size of 30 (or 100; or 500), the actual errors in surface soil moisture at the final update time decrease by 55% (or 70%; or 80%) from the value obtained without assimilation (as compared to 84% for the optimal smoother).

The EnKF significantly underestimates the forecast error variances for 100 ensemble members. However, the error variance estimates derived by the filter are reasonably good when the ensemble size is increased to 500 members. Our results indicate that the forecast error variances vary strongly with time and space. This implies that it is very important to account for dynamic error covariance propagation. Assimilation schemes that use static forecast error covariances (e.g., statistical interpolation) are unlikely to produce the desired near-optimal estimates.

More research is required to better understand the EnKF and its variants. In particular, better understanding is needed of the role of nonlinearities and related asymmetries in the conditional forecast probability density function. We have found that nonlinearities in the model and measurement processes contribute to differences in the filtering and smoothing estimates even at the final update. In our application the state (soil moisture) is bounded above and below and its distribution cannot always be well approximated by a Gaussian pdf. For very wet or dry conditions, in particular, the soil moisture pdf exhibits considerable skewness. It is likely that the variational smoother is superior to the EnKF when dealing with nonlinear and non-Gaussian effects. However, it is important to recognize that the variational approach is designed to estimate the mode of the conditional forecast density while the EnKF is designed to estimate the mean. So even if both approaches work as intended their estimates at the end of the smoothing interval can be expected to differ when the density is asymmetric.

In a practical application of the EnKF, it will probably be necessary to model the forecast covariance rather than to compute it in an exact sense. In the EnKF, covariance modeling could include smoothing out the ensemble-derived covariances before the update or applying the update to subregions of the computational domain. Ultimately, a hybrid filter that combines empirical forecast error covariances with dynamic error propagation via the EnKF (Hamill and Snyder 2000) may be the best approach.