Towards improved local analysis of the soil moisture in Catchment scale with the ensemble Kalman filter and the geostatistics

Xujun Han, Xin Li

Cold and Arid Regions Environmental and Engineering Research Institute, PR China

Abstract Various patterns of soil moisture spatial correlations in different observation scales between the adjacent grids have been observed and analyzed by the geostatistics method. Incorporating this spatial correlation information into the land data assimilation system can improve the analysis performance. The ensemble Kalman filter with small-sized ensembles could result in large sampling errors in the approximation of the background error covariances in large scale assimilation applications. Several kinds of the covariance localization techniques such as the Schur product have been applied. The Schur product multiplies the ensemble approximation of the background error covariance with a distance-dependent correlation function to suppress the correlations with increasing distance. This localization tampers the distance that an observation affects. It filters out the small correlations associated with remote observations. In this paper we adopt these distance-dependent correlation functions from the geostatistics theory and aim to improve the analysis accuracy of the surface soil moisture by combining the ensemble Kalman filter and the spatial correlation model of the geostatistics. The new developed distributed hydrological model GEOtop is used in this catchment scale hydrologic data assimilation. An Observing System Simulation Experiment (OSSE) will be conducted in the Little Washita watershed during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. The results will show that geostatistics can be regarded as a complementary term of the data assimilation.

Introduction

The land data assimilation is usually treated as a one-dimensional problem; we conduct the assimilation studies under the assumption that the horizontal error correlations between the different computational grids of the land surface model can be neglected. This scheme makes the land data assimilation an effectively low dimensional filter and reduces the computational burden considerably. This assumption is consistent with the land surface model formulation which solves the energy and mass balance for each grid cell independently. Many land data assimilation researches (Walker and Houser, 2001; Reichle et al., 2002) have been undertaken under this assumption. The spatial correlations between the soil moisture variables of adjacent grids have been observed during the soil moisture variability study. Reichle and Koster (2003) found that grid-to-grid horizontal error correlations could improve the estimation of the surface state through propagating the observation information from the observed regions to the unobserved ones. However, the spatial correlations in a land surface problem change over time as the environment changes (Ryu and Famiglietti, 2006; McLaughlin et al., 2006). For example, the soil moisture tends to be highly horizontal correlated after a prolonged wet period and after a long drydown period, horizontal correlations tend to decrease (Ryu and Famiglietti, 2006; McLaughlin et al., 2006).

The ensemble Kalman filter (EnKF) has been studied in numerous sequential data assimilation applications. We also choose the EnKF as the sequential land data assimilation method. In a large scale data assimilation the ensemble size of the EnKF is much less than the state dimension due to the computational limits. The small-sized ensembles could result in large sampling errors in the approximation of the background error covariances, and produce spurious large magnitude correlations between long-range separated grids (Hamill et al., 2001); but this is not the true case in reality. So the covariance localization is incorporated into the estimating background error covariance. With the localization, one can allow the observation having a great influence on the adjacent grids, and a small influence on the far girds. The Schur product proposed by Houtekamer and Mitchell (2001) is a localization method which multiplies the ensemble approximation of the background error covariance with a distance-dependent correlation function to suppress the correlations with increasing distance. The benefits of including distance-dependant reduction of the background error covariance estimates are demonstrated with an EnKF by Hamill et al. (2001). The choice of the distance-dependant correlation functions in atmospheric data assimilation is usually the correlation function proposed by Gaspari and Cohn (1999). These correlation functions cannot fully represent various spatial correlation structures of the surface soil moisture, so we want to use the spatial correlation functions proposed in the geostatistics to replace these correlation functions. Geostatistics provides a series of correlation functions models. The researches of the surface soil moisture variability in the SGP97 Experiment showed that the isotropic spatial correlation lengths of the surface soil moisture ranged from some tens of meters to tens of kilometers or a hundred kilometers. Several scaling characteristics of the spatial correlation are also observed (Mohanty et al., 2000; Ryu and Famiglietti, 2006). Several semivariogram models are proposed to describe the patterns of the soil moisture spatial correlations in different observation scales. These findings are very useful to the data assimilation which takes the horizontal error correlations into consideration. We think that the combination of the EnKF and geostatistics will provide a better spatial representation of errors statistics and improve the analysis accuracy of the soil moisture in the catchment scale data assimilation.

Model Description

For our catchment scale data assimilation we use the GEOtop model (Rigon et al., 2006). GEOtop is a distributed hydrological model with coupled water and energy balance, which simulates the complete hydrological balance in a catchment by combining the main features of the land surface models and the distributed rainfall-runoff models. The model accommodates very complex topography and, besides the water balance, unlike most other hydrological models, integrates all the terms in the surface energy balance equation. It is designed for remote sensing data because it provides the surface temperature and an accurate treatment of the radiation and the upper soil moisture. The grid resolution ranges from 2m to 500m, and the number of the soil layers is arbitrary with a quite thin surface layer that is appropriate to data assimilation applications.

Data Sets

Our Observing system simulation experiment (OSSE) is conducted over the Little Washita watershed, a critical research field in the SGP97 experiment, which is located in south central Oklahoma and covers an area of 611km^2 (Fig.1). The SGP97 is a collaborative soil moisture mapping experiment in the central plains of Oklahoma from 18th June to 17th July, 1997. The Little Washita watershed is one of the largest and best-instrumented watersheds in the U.S. Soils include a wide range of textures with large regions of both coarse and fine textures. Rangeland and pasture with significant areas of winter wheat and other crops dominate land use (Mohanty et al., 2000). There are 42 meteorological stations distributed across the watershed at a 5km spacing (Fig.1); these provide the precipitation data used in the simulation. For meteorological data other than precipitation, the three Oklahoma Mesonet stations: Acme (ACME), Apache (APAC), and Ninnekah (NINN), located inside or very near the watershed are used. The soil and vegetation parameters are assigned with the values according to Rigon et al. (2006). It serves as a good foundation for the following assimilation application. The watershed is discretized into a grid resolution of 200m and five soil layers with the first one of 5cm. The soil depth is given by field measurements and ranges from 0.5m to 1.5m.

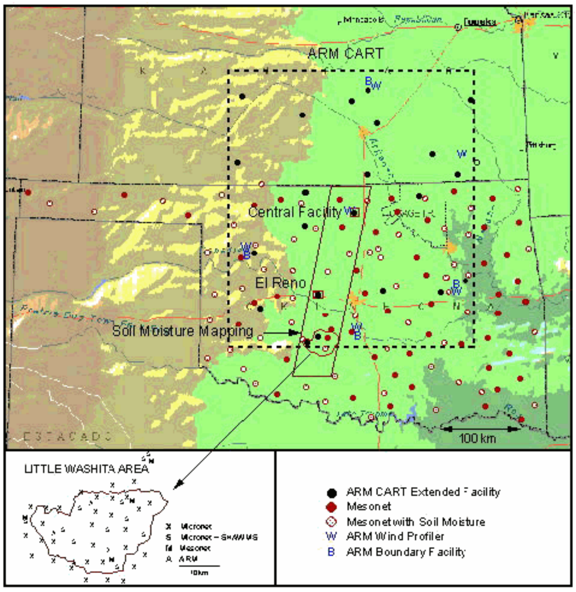

Fig.1 The layout of the SGP97 experiment site (the figure is obtained from http://hydrolab.arsusda.gov/sgp97)

Methodology

We will use the square root implementation of the EnKF analysis scheme proposed by Evensen (2004). It is shown that the sampling errors introduced by observation perturbations can be reduced by the square root analysis algorithm. The covariance localization technique is incorporated into the estimating background error covariance. The isotropic covariance models (Chilès and Delfiner, 1999) including the spherical model, the Gaussian model and the exponential model are used.

The initial condition generated is normally distributed random field with Gaussian or exponential correlation functions in space. The main sources of the model uncertainty in the hydrological model include the inaccurate soil and vegetations properties, the time-dependant forcing data, the uncertainties in the model physics and the uncertain initial conditions. The advantage of the EnKF data assimilation technique is that it allows the flexible representation of these input uncertainties. In our data assimilation applications the random perturbations of each uncertain input are provided to the GEOtop to generate random ensemble replicates. The random perturbations are generated within a physically reasonable range by the random number generator. Then these perturbations are added to the nominal input values to obtain sets of physically realistic replicates ensembles. The generation methods are different, depending on the variables. The experiment starts with one model integration which servers as the true solution. The synthetic observations of the soil moisture in the 5cm surface layer are derived from the true solutions by adding the random observation noise. The spatially uncorrelated observation errors depend on the vegetation distribution; the individual observation at each location is assimilated sequentially.

As the satellite scans the surface in swaths, it cannot observe the whole catchment simultaneously. Moreover, the soil moisture cannot be retrieved accurately from the brightness temperature in the dense vegetation region. In order to make the experiment more realistic, the data should be selected according to the observation quality. We drop the data in the dense vegetation areas, for which the soil moisture information is difficult to retrieve. Thus the quality control procedure reduces the actual data volume that will be assimilated into the model, and makes all the catchment grids unobserved simultaneously. For higher analysis accuracy the observation information can be propagated to the unobserved locations. The spatial correlation structures of the soil moisture are different in different observation scales. The length scale of the horizontal error correlations and the semivariogram model in the covariance localization are set according to the results of Mohanty et al. (2000) and Ryu and Famiglietti (2006); the error correlations beyond the spatial correlation length are neglected. Dry conditions lead to small scale horizontal correlation. During wet periods the horizontal correlation scale is significantly large. The dynamic (or adaptive) localization methods will be more appreciate.

Summary

Geostatistics can be a complementary term of the data assimilation. It can provide a better description of the error statistics for the EnKF. We aim to combine the EnKF with the spatial correlation theory of geostatistics. That will improve the analysis accuracy of the surface soil moisture. The seasonal prediction depends on the initial soil moisture condition. It is also useful for Prediction in Ungauged Basins (PUB) initiative.

Acknowledgements

The authors would like to thank the SGP97 team for their invaluable data. This research is supported by the Knowledge Innovation Program of the Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences (grant number: 2003102), and CAS International Partnership Project "The Basic Research for Water Issues of Inland River Basin in Arid Region" (CXTD-Z2005-2).

References

Chilès, J, and P Delfiner, 1999: Geostatistics: Modeling Spatial Uncertainty, *Wiley Series in Probability and Statistics*, New York.

Evensen, G, 2004: Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynamics*, **54**, 539—560.

Gaspari, G, and S Cohn, 1999: Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, **125**, 723—757.

Hamill, T, J Whitaker, and C Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, **129**, 2776—2790.

Houtekamer, P, and H Mitchell, 2001: A sequential Ensemble Kalman Filter for atmospheric data assimilation, *Monthly Weather Review*, **129**, 123—137.

McLaughlin, D, Y Zhou, D Entekhabi, and V Chatdarong, 2006: Computational issues for large-scale land surface data assimilation problems, *Journal of Hydrometeorology*, **7**, 494—510.

Mohanty, B, J Famiglietti, T Skaggs, 2000: Evolution of soil moisture spatial structure in a mixed vegetation pixel during the Southern Great Plains 1997 (SGP1997) Hydrology Experiment, *Water Resource Research*, 2000, **36**, 3675—3686.

Reichle, R, J Walker, R Koster, and P Houser, 2002: Extended versus ensemble filtering for land data assimilation, *Journal of Hydrometeorology*, **3**, 728—740.

Reichle, R, and R Koster, 2003: Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation, *Journal of Hydrometeorology*, **4**, 1229—1242.

Rigon, R, G Bertoldi, and T Over, 2006: GEOtop: A Distributed Hydrological Model with Coupled Water and Energy Budgets, *Journal of Hydrometeorology*, **7**, 371—388.

Ryu, D, and J Famiglietti, 2006: Multi-scale spatial correlation and scaling behavior of surface soil moisture, *Geophysical Research Letters*, **33**, L08404.

Walker, J, and P Houser, 2001: A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *Journal of Geophysical Research*, **106**, 11761—11774.