Assimilation of remote sensing data into land-surface models: the importance of uncertainty estimation to the filter performance

Darren Ghent, Heiko Balzter and **Jörg Kaduk**
Department of Geography,
University of Leicester,
Leicester, LE1 7RH, UK.
{djg20, hb91, jk61}@le.ac.uk

***Abstract*— Land-surface models calculate the surface to atmosphere fluxes of heat, water and carbon; and are crucial elements of General Circulation Models (GCMs). Much variation however, exists in their parameterization and representation of physical processes, leading to uncertainty in how climate change influences the land surface on a regional or global scale. A key variable in the calculation of the surface energy budget is land-surface temperature (LST), which influences the partitioning of downward radiant energy into ground, sensible and latent heat fluxes. Furthermore, LST can be applied in the prediction of live fuel moisture content, a critical variable determining fire ignition and propagation; and is crucial to soil moisture – climate feedbacks.**

Reductions in the uncertainty in model predicted soil moisture and surface energy fluxes are achievable by constraining LST simulations with remote sensing data through the process of assimilation. An often used data assimilation mechanism is the Ensemble Kalman Filter (EnKF), which is a variant of the Kalman Filter sequential assimilation method, taking a Monte Carlo approach. Of key importance to the performance of the filter are the determination of both the uncertainty in the observation source and the size of the ensemble. Results presented here indicate significantly different assimilated LST can emerge as a consequence of changes made to either of these two factors.

Keywords: data assimilation; Ensemble Kalman Filter; land-surface modelling

I. INTRODUCTION

Land surface models are crucial elements of General Circulation Models (GCMs), since they determine the surface to atmosphere fluxes of heat, water and carbon; influencing atmospheric chemistry, cloud cover and precipitation. Much variation however, exists in their parameterization, and their representation of biophysical processes. These together with the heterogeneous nature of the land surface can lead to uncertainty in how climate change influences the terrestrial biosphere. Data assimilation presents an opportunity for constraining land surface models by the integration of realistic sources of observation data. In particular, remotely sensed data from Earth Observation (EO) satellites can offer a suitable source of observations at appropriate temporal resolutions covering large geographical regions. Indeed, several previous data assimilation implementations (Huang *et al.*, 2008; Quaife *et al.*, 2008) have successfully integrated EO data into model frameworks to elicit improvements in variable simulations.

Data assimilation can be described as a method of adjusting the model state at observation times with measurements of a prescribed uncertainty, to minimize the errors in the model estimation. Inherent uncertainty in the model physics can be minimized through the updating of the model state variables at each time step when observations become available. One such mechanism is the Ensemble Kalman Filter (EnKF), which is a variant of the Kalman Filter, in which a Monte Carlo approach is taken. The foundation of the methodology is in the treatment of the respective error covariance matrices of both model and remotely sensed observations. The model error covariance matrix can be determined from the distribution of the ensemble spread, whereas the observation error covariance matrix represents the ensemble of observations with normally distributed random number perturbations with zero mean and unit variance constrained by the uncertainty of the remote sensing product. Where the magnitude of the observation uncertainty is lower than the model uncertainty then the model estimates are adjusted to more closely correspond to these observations. This uncertainty in the remote sensing product is therefore critical to the performance of the assimilation filter.

II. METHODS

In this study remotely sensed LST observations were assimilated into the Joint UK Land Environment Simulator (JULES) land-surface model, over the African continent during 2007. Both the ensemble size and the observation uncertainty were adjusted to investigate the effect these factors have on the optimization potential of applying the EnKF as an assimilation mechanism for JULES. For evaluation of the experimental results an alternative EO source of LST – the moderate resolution imaging spectroradiometer (MODIS) - was considered.

A. Experimental Setup

JULES is the community version of the UK Met Office's MOSES land-surface model, which is described in detail by Cox *et al.* (1999). Here, JULES was run at a spatial resolution of 1° x 1°, with the timestep set as 1-hour. Initial conditions were set from the final state of a cycle of spin-up - which was over 200 years in duration – prior to the main run for the year 2007. The meteorological input data required for the model run was provided by 6-hourly NCEP reanalysis datasets (Kalnay *et al.*, 1996); with precipitation data calibrated from monthly TRMM precipitation data (Kummerow *et al.*, 1998). Soil parameters were derived from the ISLSCP II soil data set (Global Soil Data Task, 2000). IGBP land-cover classes, which were mapped onto JULES according to Dunderdale *et al.* (1999), provided the vegetative cover distribution.

The spinning enhanced visible and infrared imager (SEVIRI) on board the Meteosat Second Generation (MSG) geostationary satellites, which acquire an image every 15 minutes at a spatial resolution of between 3km and 5km for the African continent, was the source of remotely sensed observations for assimilation into the JULES model. The accuracy of the split-window algorithm for processing LST is reported as 1.5K for most simulations with viewing zenith angle up to 50° (Sobrino and Romaguera, 2004). However, as Trigo *et al.* (2008) report, daytime observations over desert and semi-desert regions regularly fail to meet the satellite application facility on land surface analysis (LandSAF) target for accuracy, which they set at 2.0K. It is this uncertainty which forms an integral component in the performance of the assimilation filter, and as such a range of accuracies were experimented with.

B. Data Assimilation

The EnKF, first proposed by Evensen (1994), in this study was implemented according to the method of Evensen (2003), from where a full description of the process can be found. Briefly though, at each timestep model estimates are nudged towards the observations based on the respective state and observation error covariance matrices, whereby the correction to the forecast state vector correction is determined by the Kalman gain matrix. The optimum model state value is represented as the mean of the ensemble members. The observation error covariance matrix is a measure of the ensemble spread of SEVIRI observations, with randomly generated perturbations constructed using the observation uncertainty. The model error covariance matrix on the other hand is determined from the model ensemble spread. In this study, uncertainty in the model parameterization or in the initial conditions was not considered. Perturbations to the meteorological forcing data, generated as normally distributed random numbers with zero mean and unit variance, were solely responsible for the model ensemble spread.

Here, LST observations from SEVIRI were assimilated into JULES at each time step for the entirety of 2007. Three alternative observation uncertainties were experimented with: 1.0K; 1.5K, as defined by Sobrino and Romaguera (2004); and 2.0K. The resultant LST biases and Root Mean Squared Errors (RMSEs) between the model and the MODIS LST data set were examined. Furthermore, the assimilation was investigated under six different ensemble sizes: 10; 20; 50; 100; 150; and 200. For these experiments the product uncertainty was set at 1.5K.

III. RESULTS AND DISCUSSION

When SEVIRI LST was assimilated into JULES over the year 2007, whereby the observation uncertainty was set at 2.0K, the LST RMSE with respect to MODIS LST was reduced by 40.6% from 2.93K for the standard model run to 1.74K for the assimilated estimates (Table 1). When the assimilation was repeated with an observation uncertainty of 1.5K, the LST RMSE was further reduced by 25.9% to 1.29K; with a final reduction of 17.1% to 1.07K when the observation uncertainty was set as 1.0K.

TABLE I. COMPARISON OF MEAN BIAS AND RMSE BETWEEN JULES AND MODIS LST, BOTH FOR MODELLED AND ASSIMILATED RUNS OVER AFRICA FOR 2007. THE RESULTS REPRESENT SEPARATE EXPERIMENTS IN WHICH THREE SEPARATE OBSERVATION UNCERTAINTIES ARE APPLIED.

	Observation Uncertainty (K)	*Bias (K)*	*RMSE (K)*
Modelled	-	-2.89	2.93
Assimilated	2.0	-1.65	1.74
	1.5	-1.16	1.29
	1.0	-0.91	1.07

The EnKF balances the uncertainty in both the model and observations, but is subject to sampling errors which are a function of the ensemble size (Evensen, 2003). Figure 1 illustrates the reduction in RMSE with respect to MODIS LST as the ensemble size is initially increased. The reduction in sampling errors continues as the ensemble size is further increased, but a decreasing rate. Indeed, the reduction in RMSE remained statistically significant at the 5% level as the ensemble size was increased from 10 to 20, from 20 to 50, and from 50 to 100, but ceased to be statistically significant once the ensemble size was increased from 100 to 150. A size of 100 would thus minimize the need for covariance localization methods.

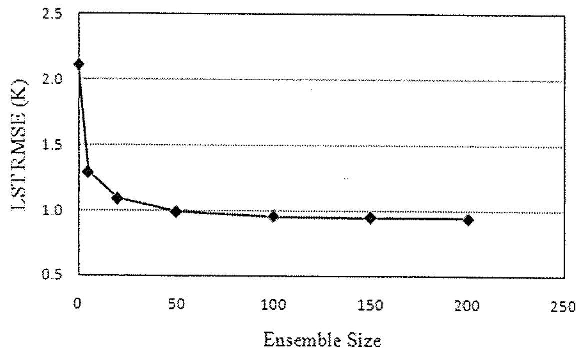

Figure 1. Comparison test for experiments with different ensemble sizes of the assimilation of SEVIRI LST into the JULES model for the entire year 2007. LST RMSEs are with respect to mean MODIS LST observations over the same period.Example of a figure caption.

Data assimilation over land with the EnKF is often based on covariance localization methods as implanted in Hamill *et al.* (2001) and Houtekamer and Mitchell (2001). Specifically, the concept here is to sub-divide the global analysis into smaller sub-domains which are independently analyzed based only on local observations. The sub-domains are simultaneously updated in parallel with sampled cross-covariances between geographically disparate points being explicitly set to zero, due to the possibility of creating spurious cross-correlations since the uncertainties in sampling cross-covariances using finite ensembles sizes are a source of uncertainty. For small ensemble sizes, Reichle and Koster (2003) have found covariance localization to increase the accuracy of estimations. The magnitude of spurious cross-correlations is however dampened as the ensemble size increases, and indeed as the ensemble size tends to infinity covariance localization is not a requirement. Indeed this need for localization manifests itself in a significant improvement in EnKF results as the ensemble size increases.

IV. Conclusions

The EnKF is a flexible and practical data assimilation method, which is relatively straightforward to implement. Indeed, evidence from numerous studies suggests that significant optimization of land-surface schemes can be achieved through data assimilation. However, in order to achieve appropriate improvements one must consider the important issue of uncertainties. Firstly, with regards to this study, a suitable ensemble size should be chosen which both minimizes computational load, while simultaneously ensuring the uncertainties resulting from cross-covariances are reduced where localization techniques are not applied. Secondly, a most accurate representation of the uncertainty in the observation measurements is critical to instill confidence in the improved simulations of land-surface models. Furthermore, there remains additional scenarios that merit investigation in order to be satisfied with the data assimilation enhancements to the JULES model; for example, experimentation with perturbed initial conditions and model parameters; and validation with in situ measurements.

Acknowledgements

This work has been funded by the NERC Climate and Land Surface Systems Interactions Centre (CLASSIC). The authors would like to thank Dr. Elizabeth Good at the UK Met Office, Exeter for the valuable input pertaining to the SEVIRI LST data retrieval and processing.

References

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R. and Smith, J. (1999). The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics.* 15(3), 183-203.

Dunderdale, M., Muller, J. P., and Cox, P. M. (1999). Sensitivity of the Hadley Centre climate model to different earth observation and cartographically derived land surface data-sets. *The contribution of polder and new generation spaceborne sensors to global change studies*, Meribel, France, (pp. 1–6).

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research-Oceans.* 99(c5), 10143-10162.

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics.* 53, 343-367.

Global Soil Data Task. (2000). Global gridded surfaces of selected soil characteristics (IGBPDIS). *International Geosphere-Biosphere Programme - Data and Information Services.* Available online [http://www.daac.ornl.gov/] from the ORNL Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an Ensemble Kalman Filter. *Monthly Weather Review.* 129(11), 2776–2790.

Houtekamer, P. L., and Mitchell, H. L. (2001). A sequential Ensemble Kalman Filter for atmospheric data assimilation. *Monthly Weather Review.* 129(1), 123–137.

Huang, C. L., Li, X., and Lu., L. (2008). Retrieving soil temperature profile by assimilating MODIS LST products with Ensemble Kalman Filter. *Remote Sensing of Environment.* 112(4), 1320-1336.

Kalnay, E., et al. (1996). The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society.* 77(3), 437-471.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J. and Simpson, J. (1998). The Tropical Rainfall Measuring Mission (TRMM) sensor package, *Journal of Atmospheric and Oceanic Technology.* 15(3), 809-817.

Quaife, T., et al. (2008). Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment.* 112(4), 1347-1364.

Reichle, R. H. and Koster, R. D. (2003). Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *Journal of Hydrometeorology.* 4(6), 1229-1242.

Sobrino, J. A., and Romaguera, M. (2004). Land surface temperature retrieval from MSG1-SEVIRI data. *Remote Sensing of Environment.* 92, 247-254.

Trigo, I. F., Peres, L. F., DaCarnara, C. C., and Freitas, S. C. (2008), Thermal land surface emissivity retrieved from SEVIRI/meteosat. *IEEE Transactions on Geoscience and Remote Sensing.* 46(2), 307-315.