Advances in Land Data Assimilation at NASA

R. H. Reichle

NASA Goddard Space Flight Center,
Greenbelt, MD, USA

***With contributions from*: M. Bosilovich, W. Crow, R. Koster, S. Kumar, G. De Lannoy, Q. Liu, C. Peters-Lidard, S. Mahanama, M. Rodell, J. Santanello, and B. Zaitchik**

Abstract

Research in land surface data assimilation has grown rapidly over the last decade. We provide a brief overview of key research contributions by United States National Aeronautics and Space Administration (NASA), including the continued development and application of the Ensemble Kalman filter (EnKF) for land data assimilation and the NASA Land Information System (LIS) software infrastructure. These systems were used successfully to assimilate satellite observations of surface soil moisture, land surface temperature, snow cover, and terrestrial water storage. Additionally, synthetic experiments were conducted in preparation for the NASA Soil-Moisture-Active-Passive (SMAP) mission.

1. Motivation

Land surface conditions are intimately connected with the global climate system and have been, through different pathways, associated with predictability of atmospheric variability. Land surface models driven with observation-based meteorological forcing data (precipitation, radiation, air temperature and humidity, etc.) offer estimates of global land surface conditions (Rodell *et al.* 2004). Satellite remote sensing provides complementary information about land surface conditions, including surface soil moisture, snow water equivalent, snow cover, land surface temperature (LST), and terrestrial water storage (TWS).

Land data assimilation systems combine the modeled land surface fields with observational estimates and produce dynamically consistent, spatially complete and temporally continuous estimates of global land surface conditions based on both sources of information. The land assimilation estimates can be used, for example, within atmospheric assimilation systems and for the initialization of global short-term climate forecasts. Through such use land data assimilation systems offer a unique validation and monitoring perspective because the satellite-based land surface data are continually confronted with independent observations and model estimates. Land data assimilation systems can also be used to establish measurement requirements for future land surface satellite missions such as the SMAP mission.

2. Ensemble-based land data assimilation

Ensemble-based algorithms have emerged as a common and promising method for land assimilation (Reichle *et al*. 2002a,b). Research at NASA contributed significantly to this progress through the development of an ensemble-based land data assimilation system (Reichle *et al*. 2009) and the LIS software framework (Kumar *et al*. 2008a,b). At NASA, the EnKF has been used primarily with the

NASA Catchment land surface model (Koster *et al.* 2000). Through LIS, other land surface models can also be used.

Large differences have been identified between the temporal moments of satellite and model estimates, for example, of soil moisture (Reichle *et al*. 2004, 2007). Because the standard EnKF is only designed to address short-term "random" errors, the climatological differences need to be addressed separately within the assimilation system. This can be accomplished by scaling the satellite observations to the model's climatology so that the cumulative distribution functions (cdf) of the satellite soil moisture and the model soil moisture match (Reichle and Koster 2004). For weather and climate forecast initialization, knowledge of soil moisture anomalies is, in any case, more important than knowledge of absolute soil moisture.

Generally, data assimilation products are sensitive to input observation and model error variances. Figure 1 shows an example from a suite of experiments in which synthetic surface soil moisture observations are assimilated (Reichle *et al*. 2008b). Each assimilation experiment has a unique set of input error parameters that leads to a unique pair of scalars: the (space and time) average forecast error variance (P0) and the input observation error variance (R0) for surface soil moisture. We can thus plot two-dimensional surfaces of filter performance as a function of sqrt(P0) and sqrt(R0). Figure 1a, for example, shows one such surface with the performance measure being the RMSE of surface soil moisture estimates from the (non-adaptive) EnKF.

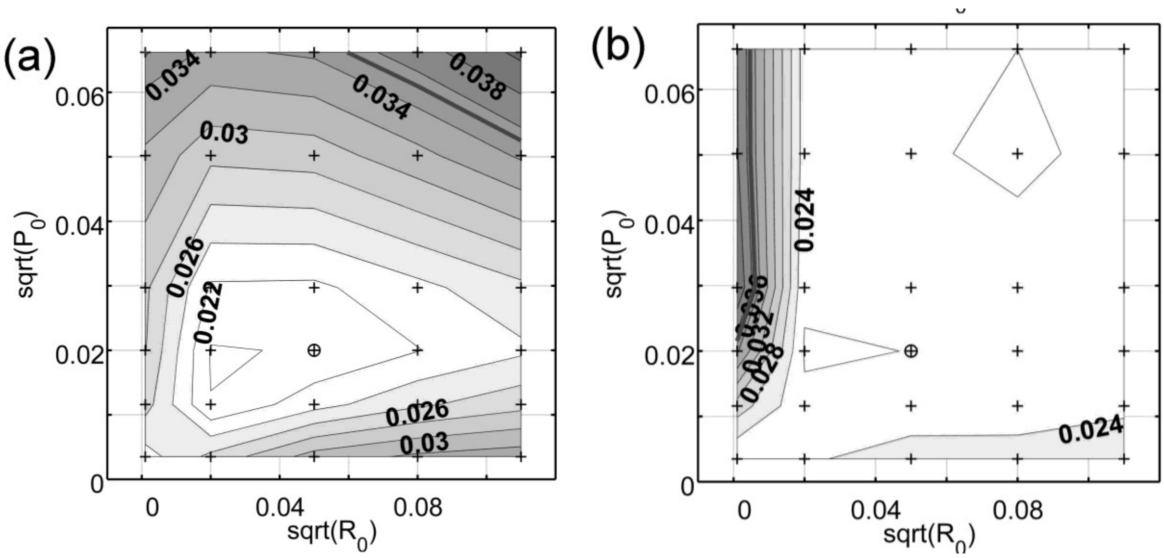

Figure 1. RMSE in m^3m^{-3} of surface soil moisture for (a) standard and (b) adaptive EnKF as function of input error parameters: (ordinate) forecast and (abscissa) observation error std-dev. Each plus sign indicates the result of a 19-year assimilation integration over the Red-Arkansas domain. Circled plus signs indicate experiments that use true input error parameters for assimilation. Thick gray lines indicate RMSE of open loop integration.

Figure 1a illustrates that the estimation error in surface soil moisture is smallest near the experiment that uses the true model and observation error inputs. The minimum estimation error is around 0.02 m^3m^{-3}, down from the open loop (no assimilation) value of 0.035 m^3m^{-3}. The estimation error increases as the input error parameters deviate from their true values. Figure 1a also indicates where the estimation error surface intersects the open loop error. For grossly overestimated model and observation error variances, the assimilation estimates of surface soil moisture are in fact worse than the open loop estimates. Ultimately, the success of the assimilation (measured through independent validation) suggests whether the selected input error parameters are acceptable.

3. Adaptive filtering

Adaptive filtering methods can assist with the estimation of the filter's input error parameters. The central idea behind adaptive filtering methods is that internal diagnostics of the assimilation system (such as the statistics of the "observation-minus-forecast" residuals and the assimilation increments) should be consistent with the values that are expected from input parameters provided to the data assimilation system. Following Desroziers *et al*. (2005), Reichle *et al*. (2008b) developed an adaptive algorithm for land assimilation that permits the separate estimation of model and observation error parameters. An example of the benefits of the adaptive module is given in Figure 1b (Reichle *et al*. 2008b). The adaptive estimation of input error parameters leads to improved estimates of surface soil moisture regardless of initial error estimates, except for the case of severe underestimation of the input observation error variance. The poor performance in this special case is due to technicalities in the implementation of the adaptive module and can easily be avoided in applications.

4. Observing system design

For the design of new satellite missions it is critical to understand just how uncertain satellite retrievals can be and still be useful. Consider, for example, that a mission assimilation product will have some target accuracy requirement. For a given level of model skill, a specific level of retrieval skill would be needed to bring the merged product to the target accuracy. The required skill level for the retrievals would undoubtedly increase with a decrease in the skill of the raw model product. Quantitative knowledge of such retrieval skill requirements, for example, is directly relevant to the planning of the L-band (1.4 GHz) SMAP mission.

Reichle *et al*. (2008a) designed an Observing System Simulation Experiment (OSSE) that determines the contribution of surface soil moisture retrievals to the skill of land assimilation products (soil moisture and evapotranspiration) as a function of retrieval and land model skill. The OSSE consists of a suite of synthetic data assimilation experiments based on integrations of two distinct land models, one representing "truth", and the other representing our flawed ability to model the true processes. Skill is measured in terms of the correlation coefficient R between the time series of the various estimates (expressed as anomalies relative to their seasonal climatologies) and the assumed (synthetic) truth.

Each assimilation experiment is a unique combination of a retrieval dataset (with a certain level of skill, measured in terms of R) and a model scenario (with its own level of skill). We can thus plot two-dimensional surfaces of skill in the data assimilation products as a function of retrieval and model skill. Figure 2a, for example, shows the two dimensional surface corresponding to the surface soil moisture product. As expected, the skill of the assimilation product generally increases with the skill of the model and the skill of the retrievals, for both surface (Figure 2a) and root zone (Figure 2b) soil moisture estimates. Except for very low model skill, the contour lines are more closely aligned with lines of constant model skill; that is, the skill of the assimilation product is more sensitive to model skill than to retrieval skill.

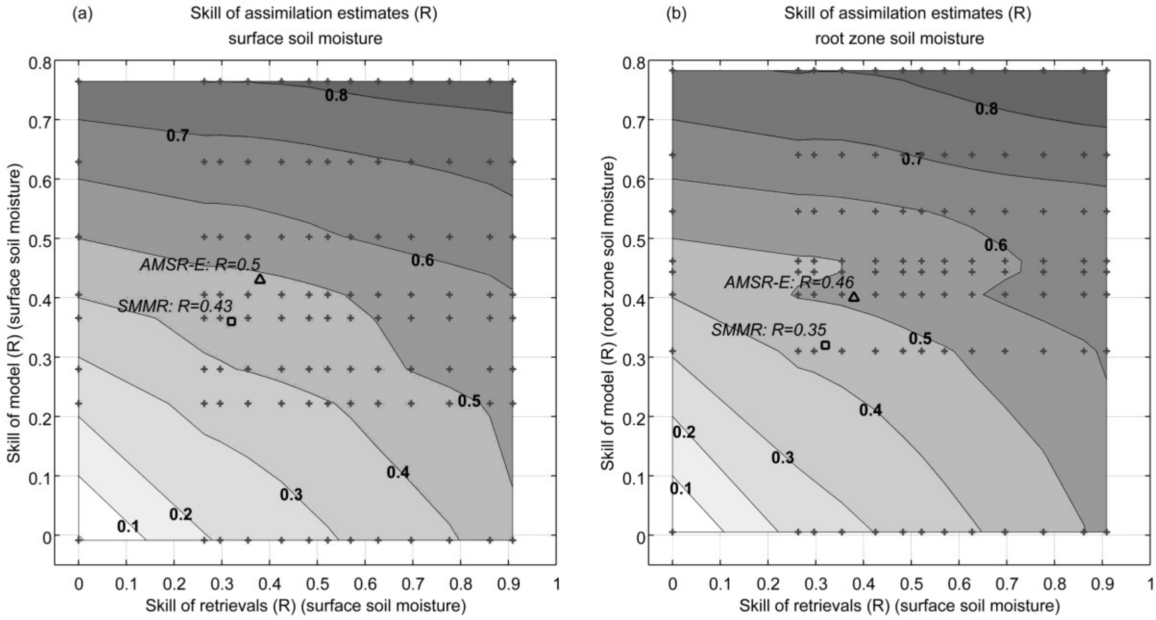

Figure 2. Skill (R) of assimilation product for (a) surface and (b) root zone soil moisture as a function of the (ordinate) model and (abscissa) retrieval skill. Each plus sign indicates the result of one 19-year assimilation integration over the entire Red-Arkansas domain. Also shown are results from Reichle *et al. (2007) for (triangle) AMSR-E and (square) SMMR.*

Figure 2 can easily be redrawn in terms of skill improvement through data assimilation (not shown). The skill of the surface and root zone soil moisture assimilation products always exceeds that of the model. As expected, the improvements in R through assimilation increase with increasing retrieval skill and decrease with increasing model skill. Perhaps most importantly, though, is that even retrievals of low quality contribute some information to the assimilation product, particularly if model skill is modest.

We can also compare the contoured skill levels of Figure 2 with those obtained by Reichle and Koster (2005) and Reichle *et al*. (2007) through the assimilation satellite retrievals from the Scanning Multichannel Microwave Radiometer (SMMR) and the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E). From the contours of Figure 2a we expect that for retrievals with R=0.38 and model estimates with R=0.43, the surface soil moisture assimilation product would have skill of about R=0.50, which is indeed consistent with the AMSR-E result. For root zone soil moisture, the assimilation of AMSR-E surface soil moisture retrievals also yields improvements, though these improvements fall somewhat short of those suggested by Figure 2b. Possible explanations include (i) the imperfect translation of information from the surface layer to the root zone in the data assimilation system and (ii) the fact that the in situ data used for validation of the AMSR-E result are themselves far from perfect (unlike the perfectly known truth of the synthetic experiment).

5. Terrestrial water storage (TWS) assimilation

The Gravity Recovery and Climate Experiment (GRACE) satellite mission provides unprecedented observations of variations in TWS, albeit at low spatial (>105 km2) and temporal (monthly) resolutions. Depending on topographic and climatologic conditions, TWS variability may be dominated by ground water, soil moisture, surface water, and/or snow. Zaitchik *et al*. (2008) assimilated GRACE TWS data for the Mississippi River basin. Because of the temporally integrated nature of the GRACE observations, the assimilation was set up as an ensemble smoother and GRACE TWS anomalies were converted to absolute TWS values by adding the corresponding time-mean TWS from a Catchment model simulation. Ensemble perturbations were generated with a horizontal correlation scale of 2 degrees, which roughly represents error scales in global-scale precipitation fields (Reichle and Koster 2003).

The assimilation system separates the contributions of GRACE observations into individual TWS components and down-scales the GRACE observations to scales (~103 km2) typical of global land surface integrations. Assimilation products include catchment-scale groundwater, root zone soil moisture, surface heat fluxes, and runoff. The spatial resolution of the assimilation products is much higher than that of GRACE observations alone, making the results more useful for water resources and forecasting applications. Figure 3 shows that the groundwater time series from the GRACE assimilation integration resembles in situ estimates more closely than model estimates alone, with RMSE reduced by 21% (from 23.5 mm to 18.5 mm). For the four sub-basins of the Mississippi, RMSE reductions ranged from 7% to 36%.

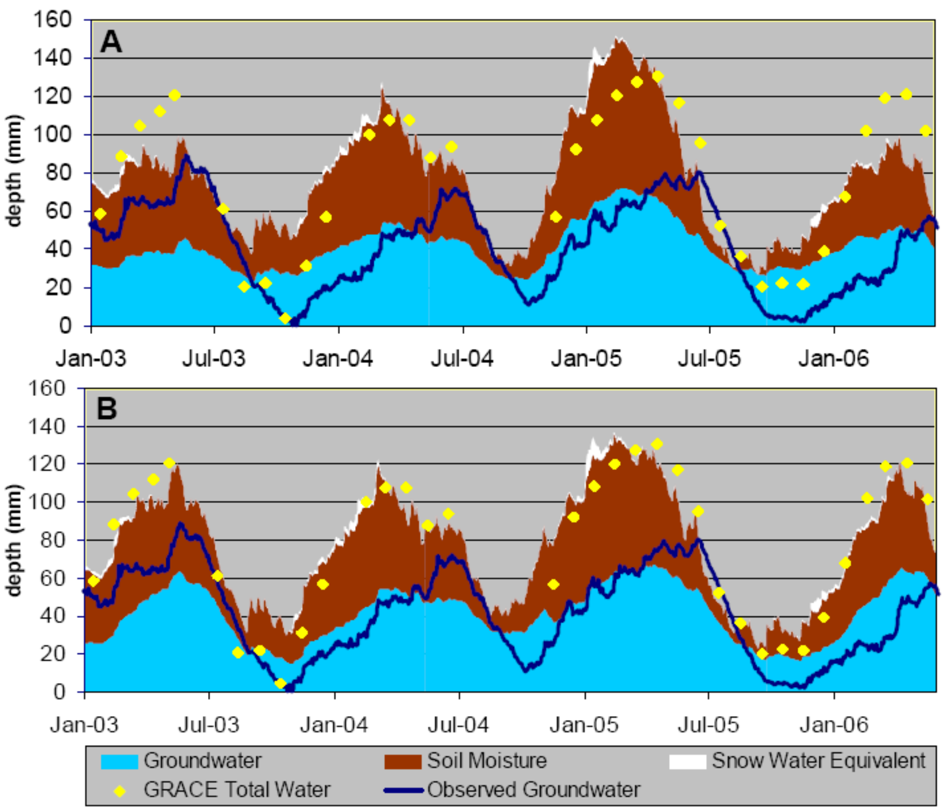

Figure 3: Groundwater, soil moisture, and snow water equivalent for the Mississippi river basin for estimates from (A) the model without assimilation and (B) the GRACE assimilation integration. Also shown are (solid line) area averaged daily in situ groundwater observations and (diamonds) monthly GRACE-derived TWS anomalies. Note that GRACE assimilation improves agreement of the groundwater estimates with in situ data. GRACE and modeled TWS are adjusted to a common mean, as are observed and modeled groundwater.

Assimilation of GRACE observations also produced improved estimates of hydrologic variability at the sub-observation scale and a small increase in correlation between runoff estimates and gauged river flow in the majority of test watersheds (not shown). The results demonstrate that – through data assimilation – coarse resolution, vertically integrated TWS anomalies from GRACE can be spatially and temporally disaggregated and attributed to different components of the snow-soil-aquifer column in a physically meaningful way.

6. Multi-model soil moisture assimilation

Root zone soil moisture controls the land-atmosphere exchange of water and energy and exhibits memory that may be useful for climate prediction at monthly scales. Assimilation of satellite-based surface soil moisture observations into a land surface model is an effective way to estimate large-scale root zone soil moisture. The propagation of surface information into deeper soil layers depends on the model-specific representation of subsurface physics that is used in the assimilation system. In a suite of EnKF experiments Kumar *et al*. (2009) used LIS to assimilate synthetic surface soil moisture observations into four different models (Catchment, Mosaic, Noah and CLM). They demonstrated that identical twin experiments significantly overestimate the information that can be obtained from the assimilation of surface soil moisture observations. The second key result indicates that the potential of surface soil moisture assimilation to improve root zone information is higher when the surface to root zone coupling is stronger. The experiments also suggest that (faced with unknown true subsurface physics) overestimating surface to root zone coupling in the assimilation system provides more robust skill improvements in the root zone compared with underestimating the coupling.

7. Land surface temperature (LST) assimilation

Satellite retrievals of LST (also referred to as “skin temperature”) are available from a variety of polar orbiting and geostationary platforms. Assimilating such LST retrievals into a land surface model (that is either driven by observed meteorological forcing data or coupled to an atmospheric model) should improve estimates of land surface conditions. Similar to surface soil moisture, however, LST data from retrievals and models typically exhibit very different climatologies for a variety of reasons. Bosilovich *et al*. (2007) developed an algorithm for LST assimilation into a global, coupled land-atmosphere data assimilation system by introducing an incremental bias correction term into the model’s surface energy budget and assuming that the LST bias is solely due to the model.

As an alternative strategy, Reichle *et al*. (2009) tested several combinations of dynamic bias estimation and a priori rescaling. The latter approach does not assume that the model is the only source of bias. They assimilated LST retrievals from the International Satellite Cloud Climatology Project (ISCCP) with the ensemble-based, off-line land data assimilation system into the Noah and Catchment (CLSM) land surface models. LST is described very differently in the two models. When compared to in situ measurements, LST estimates from Noah and CLSM without data assimilation ("open loop") are comparable and superior to that of ISCCP retrievals. Assimilation of ISCCP retrievals provides modest yet statistically significant improvements (over open loop) of 0.5-0.7 K in terms of raw RMSE and of 0.3 K in terms of anomaly RMSE. Surface turbulent flux estimates from CLSM and Noah assimilation integrations are essentially identical to open loop estimates. Noah assimilation estimates of ground heat flux, however, are significantly worse. Provided the assimilation system is properly adapted to each land model, the benefits from the assimilation of LST retrievals are comparable for both land models.

8. Snow data assimilation

Snow cover over land has a significant impact on the surface radiation budget, turbulent energy fluxes to the atmosphere, and local hydrological fluxes. For this reason, inaccuracies in the representation of snow-covered area (SCA) within a land model can lead to substantial errors in both offline and coupled simulations. Data assimilation algorithms have the potential to address this problem. However, the assimilation of SCA observations is complicated because SCA indicates only the presence or absence of snow and because assimilated SCA observations can introduce inconsistencies with atmospheric forcing data, leading to nonphysical artifacts in the local water balance. Zaitchik and Rodell (2009) present a novel assimilation algorithm that introduces Moderate Resolution Imaging Spectroradiometer (MODIS) SCA observations to the Noah land model in global, uncoupled simulations. The algorithm uses observations from up to 72 h ahead of the model simulation to correct against emerging errors in the simulation of snow cover while preserving the local hydrologic balance. This is accomplished by using future snow observations to adjust air temperature and, when necessary, precipitation within the land model. In global, offline integrations, this new assimilation algorithm provided improved simulation of SCA and snow water equivalent relative to open loop integrations and integrations that used an earlier SCA assimilation algorithm. These improvements, in turn, influenced the simulation of surface water and energy fluxes during the snow season and, in some regions, on into the following spring.

In another study, De Lannoy *et al*. (2009) tested four EnKF-based methods to assimilate coarse-scale (25 km) snow water equivalent (SWE) observations (typical of passive microwave satellite retrievals) into fine-scale (1 km) land model simulations. Synthetic coarse-scale observations were assimilated directly using an observation operator for mapping between the coarse and fine scales or, alternatively, after disaggregation (re-gridding) to the fine-scale model resolution prior to data assimilation. In either case observations were assimilated either simultaneously or independently for each location. Results indicate that assimilating disaggregated fine-scale observations independently is less efficient than assimilating a collection of neighboring disaggregated observations. Direct assimilation of coarse-scale observations is superior to a priori disaggregation. Independent assimilation of individual coarse-scale observations can bring the overall mean analyzed field close to the truth, but does not necessarily improve estimates of the fine-scale structure. There is a clear benefit to simultaneously assimilating multiple coarse-scale observations even as the entire domain is observed, indicating that underlying spatial error correlations can be exploited to improve SWE estimates. The latter method avoids artificial transitions at the coarse observation pixel boundaries and can reduce the RMSE by 60% when compared to the open loop in this study.

9. Land Information System (LIS)

Another important development is the gradual implementation of the ensemble-based data assimilation modules into LIS (Kumar *et al*. 2008a,b), a land surface modeling framework that integrates various community land surface models, ground and satellite-based observations, and data management tools within an architecture that allows interoperability of land surface models and parameters, surface meteorological forcing inputs, and observational data. The high performance infrastructure in LIS provides adequate support to conduct assimilation experiments of high computational granularity. Integration of the assimilation modules into LIS therefore permits their use with a variety of land surface models and makes the NASA contributions to land data assimilation development accessible to the research community.

In fact, LIS has been used in some of the land assimilation research mentioned above (for example, (Kumar *et al*. 2009) and (De Lannoy *et al*. 2009)). LIS has also been coupled to the Weather Research and Forecasting Model (WRF). This system was used by Santanello *et al*. (2009) to develop a framework for diagnosing land-atmosphere interactions by determining the diurnal evolution of temperature and moisture in the soil and the planetary boundary layer. The LIS-WRF system and the land-atmosphere coupling diagnostics framework will enable future land assimilation experiments in the coupled land-atmosphere system.

10. Conclusions and future directions

Much has been accomplished with the development and application of ensemble-based land assimilation over the past few years at NASA and elsewhere. The general ensemble-based framework of the system has been established and demonstrated with the assimilation of satellite-based land surface observations. The results presented here, however, all reflect the impact of uni-variate assimilation of land surface observations. Errors in the coupled land-atmosphere system are difficult to pin down because they may be related to any number of causes, including errors in precipitation, cloud biases, or errors in land surface parameters. The multi-variate assimilation of land surface observations should lead to more consistent and improved estimates of the land surface water and energy budget. Future developments will thus include the implementation of the land assimilation as an integral component within the next-generation NASA global atmospheric data assimilation system.

Acknowledgments

We gratefully acknowledge support from NASA grant NNG05GB61G and are indebted to the many collaborators inside and outside NASA who contributed to our system.

References

Bosilovich M, Radakovich J, da Silva A, Todling R, Verter F (2007) Skin temperature analysis and bias correction in a coupled land-atmosphere data assimilation system. *J Meteo Soc Japan* 85A:205–228.

De Lannoy, GJM., Reichle RH, Houser PR, Arsenault KR, Verhoest NEC, Pauwels VRN (2009), Satellite-Scale Snow Water Equivalent Assimilation into a High-Resolution Land Surface Model, *J Hydrometeorol,* in press, doi:10.1175/2009JHM1192.1.

Desroziers G, Berre L, Chapnik B, Poli P (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quart J R Meteo Soc* 131(613):3385-3396 Part C, doi:10.1256/qj.05.108.

Koster RD, Suarez MJ, Ducharne A, Stieglitz M, Kumar P (2000) A catchment-based approach to modeling land surface processes in a general circulation model, 1: Model structure. *J Geophys Res* 105(20):24809-24822.

Kumar SV, Reichle RH, Peters-Lidard CD, Koster RD, Zhan X, Crow WT, Eylander JB, Houser PR (2008a) A Land Surface Data Assimilation Framework using the Land Information System: Description and Applications. *Adv Water Resour*, doi:10.1016/j.advwatres.2008.01.013, in press.

Kumar S, Peters-Lidard C, Tian Y, Reichle R, Geiger J, Alonge C, Eylander J, Houser P (2008b) An integrated hydrologic modeling and data assimilation framework, *IEEE Computer*, 41:52-59, doi:10.1109/MC.2008.511.

Kumar SV, Reichle RH, Koster RD, Crow WT, Peters-Lidard CD (2009) Role of subsurface physics in the assimilation of surface soil moisture observations, *J Hydrometeorol*, in press, doi:10.1175/2009JHM1134.1.

Reichle RH, Koster RD (2003) Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J Hydrometeorol* 4(6):1229-1242.

Reichle RH, Koster RD (2004) Bias reduction in short records of satellite soil moisture. *Geophys Res Lett* 31: L19501, doi:10.1029/2004GL020938.

Reichle RH, Koster RD (2005) Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophys Res Lett* 32(2):L02404, doi:10.1029/2004GL021700.

Reichle RH, McLaughlin D, Entekhabi D (2002a) Hydrologic data assimilation with the Ensemble Kalman filter. *Mon Weather Rev* 130(1):103-114.

Reichle RH, Walker JP, Koster RD, Houser PR (2002b) Extended versus Ensemble Kalman filtering for land data assimilation. *J Hydrometeorol* 3(6):728-740.

Reichle RH, Koster RD, Dong J, Berg AA (2004) Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J Hydrometeorol* 5:430-442.

Reichle RH, Koster RD, Liu P, Mahanama SPP, Njoku EG, Owe M (2007) Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J Geophys Res* 112:D09108, doi:10.1029/2006JD008033.

Reichle RH, Crow WT, Koster RD, Sharif H, Mahanama SPP (2008a) The contribution of soil moisture retrievals to land data assimilation products. *Geophysical Research Letters*, 35, L01404, doi:10.1029/2007GL031986.

Reichle RH, Crow WT, Keppenne CL (2008b) An adaptive ensemble Kalman filter for soil moisture data assimilation. *Wat Resour Res*, doi:10.1029/2007WR006357, in press.

Reichle RH, Kumar SV, Mahanama S, Koster RD, Liu Q, Bosilovich MG (2009) Assimilation of satellite-derived land skin temperature retrievals into land surface models. *J Hydrometeorol,* in preparation.

Rodell M, Houser PR (2004) Updating a land surface model with MODIS-derived snow cover. *J Hydrometeorol,* 5:1064-1075.

Rodell M, Houser PR, Jambor U, Gottschalck J, Mitchell K, Meng C-J, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Toll DL (2004) The Global Land Data Assimilation System. *Bull Amer Meteorol Soc* 85:381-394, doi:10.1175/BAMS-85-3-381.

Santanello JA, Peters-Lidard CD, Kumar SV, Alonge C, Tao WK (2009) A Modeling and Observational Framework for Diagnosing Local Land–Atmosphere Coupling on Diurnal Time Scales. *J Hydrometeorol,* 10:577–599.

Zaitchik BF, Rodell M (2009) Forward-Looking Assimilation of MODIS-Derived Snow-Covered Area into a Land Surface Model. *J Hydrometeorol*, 10:130–148.

Zaitchik BF, Rodell M, Reichle RH (2008) Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *J Hydrometeorol*, 9:535-548.