GEOPHYSICAL RESEARCH LETTERS, VOL. 35, L01404, doi:10.1029/2007GL031986, 2008

Contribution of soil moisture retrievals to land data assimilation products

R. H. Reichle,[1,2] W. T. Crow,[3] R. D. Koster,[2] H. O. Sharif,[4] and S. P. P. Mahanama[1,2]

Received 11 September 2007; revised 19 November 2007; accepted 6 December 2007; published 10 January 2008.

[1] Satellite measurements (retrievals) of surface soil moisture are subject to errors and cannot provide complete space-time coverage. Data assimilation systems merge available retrievals with information from land surface models and antecedent meteorological data, information that is spatio-temporally complete but likewise uncertain. For the design of new satellite missions it is critical to understand just how uncertain retrievals can be and still be useful. Here, we present a synthetic data assimilation experiment that determines the contribution of retrievals to the skill of land assimilation products (soil moisture and evapotranspiration) as a function of retrieval and land model skill. As expected, the skill of the assimilation products increases with the skill of the model and that of the retrievals. The skill of the soil moisture assimilation products always exceeds that of the model acting alone; even retrievals of low quality contribute information to the assimilation product, particularly if model skill is modest. **Citation:** Reichle, R. H., W. T. Crow, R. D. Koster, H. O. Sharif, and S. P. P. Mahanama (2008), Contribution of soil moisture retrievals to land data assimilation products, *Geophys. Res. Lett.*, *35*, L01404, doi:10.1029/2007GL031986.

1. Motivation

[2] A common approach to estimating soil moisture is to drive a land surface model (LSM) with observed meteorological forcing. The physical formulations within the LSM effectively integrate the forcing and produce estimates of soil moisture and associated land surface fields [*Rodell et al.*, 2003]. These model products, however, are subject to error due to errors in meteorological forcing, faulty estimates of relevant parameters, and deficient LSM formulations.

[3] Indirect measurements (retrievals) of surface soil moisture can be obtained from satellite sensors that measure microwaves emitted by the land surface [*Bindlish et al.*, 2003; *Njoku et al.*, 2003]. The data coverage, however, is incomplete in space and time, and retrieval errors arise from limitations in instrument design (including sensor hardware, antenna, and orbit parameters), difficulties in parameterizing the physical processes that relate passive microwave brightness temperature (a measure of the microwave emission) to soil moisture, and difficulties in obtaining the global distributions of the retrieval algorithm's parameters.

[4] Data assimilation systems are designed to merge the retrieval information with the spatially and temporally complete information provided by the LSM [*Drusch*, 2007; *Reichle et al.*, 2007] and produce a superior product (e.g. root zone soil moisture). The assimilation system acts to propagate the surface retrieval information into deeper soil layers, giving the retrievals an otherwise unobtainable relevance to such applications as the initialization of weather and seasonal climate forecasts.

[5] Data assimilation systems are thus an invaluable part of any satellite-based soil moisture measurement mission. 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. Knowledge of such retrieval skill requirements is directly relevant to the planning of the L-band (1.4 GHz) Soil Moisture Active-Passive (SMAP) mission recommended by the National Academy of Sciences for launch in the 2010–2013 timeframe [*Space Studies Board*, 2007]. Here, we describe a Data Assimilation-Observing System Simulation Experiment (DA-OSSE) that measures the contribution of surface soil moisture retrievals to the skill of the assimilation estimates (of surface and root zone soil moisture and evapotranspiration) as a function of the errors in the satellite retrievals and in the LSM. By including a land data assimilation system, our DA-OSSE differs fundamentally from earlier "retrieval" OSSE's [*Crow et al.*, 2001, 2005a].

2. Approach

[6] The DA-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. The skill of the retrievals, model estimates, and assimilation products 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 truth. The R measure is appropriate because land surface models generally differ in their soil moisture climatologies (in mean, variability, and higher moments), so that skill cannot usefully be measured in terms of RMS error. Knowledge of soil moisture anomalies is, in any case, more important than knowledge of absolute soil moisture for weather and climate forecast initialization.

[1] Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland, USA.
[2] Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[3] Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland, USA.
[4] Civil Engineering Department, University of Texas, San Antonio, Texas, USA.

0094-8276/08/2007GL031986$05.00

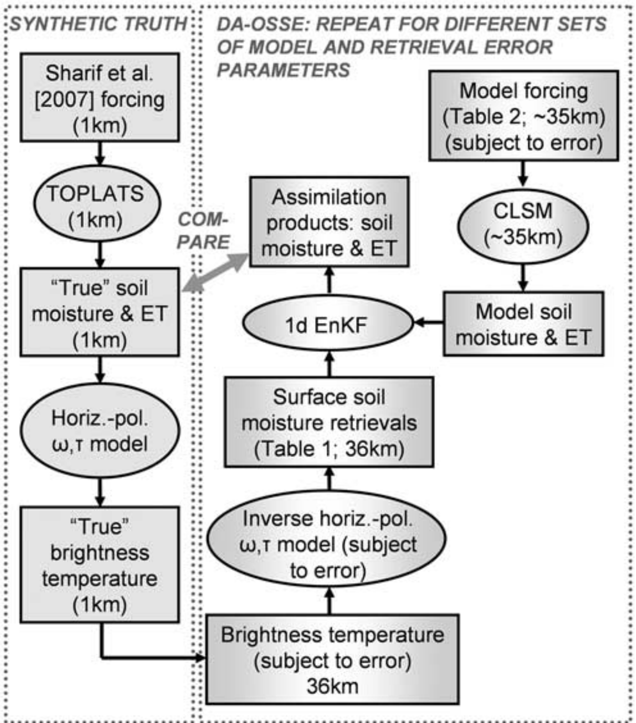

Figure 1. Flow diagram of the DA-OSSE.

[7] Figure 1 shows a flow diagram of the experiments. "True" soil moisture fields and passive microwave brightness temperatures (L-band) are from a high-resolution (1 km), long-term (1981–2000) integration of the TOPLATS land surface model [*Peters-Lidard et al.*, 1997] over the Red-Arkansas river basin in the United States, using high-quality meteorological forcing data [*Sharif et al.*, 2007] and a horizontal-polarization radiative transfer model with parameterized vegetation single scattering albedo (ω) and opacity (τ) [*Crow et al.*, 2005a]. The Red-Arkansas domain was chosen because it exhibits a range of land cover conditions that favor or prohibit soil moisture retrieval from space. From the "true" brightness temperature fields, we simulate N_R = 12 different retrieval datasets of surface soil moisture at a typical satellite footprint scale (36 km) and temporal resolution (at most once a day) according to the expected availability of retrievals from an L-band satellite radiometer. Aggregation errors in brightness temperature were modeled by adding synthetic noise at the footprint-scale. Prior to running the retrieval model (the inverse radiative transfer model), unbiased Gaussian noise terms were added to the footprint-scale canopy vegetation water content (VWC), VWC-to-canopy opacity conversion parameters, and near-surface soil temperature. Starting from the design error parameters of the Hydros soil moisture mission concept [*Entekhabi et al.*, 2004; *Crow et al.*, 2005a, 2005b], the noise variances were successively changed to create a set of soil moisture retrieval products with variable accuracies. Anomaly time series correlation coefficients R for the synthetic retrieval datasets range from 0.26 to 0.91 (Table 1). These and all subsequent R-values for soil moisture are computed from daily, catchment-scale (see below) time series for the 19-year experiment period (after subtracting the monthly climatology) and then area-averaged over the entire domain.

[8] Next, we construct N_M = 8 distinct modeling scenarios by integrating the NASA Catchment land surface model (CLSM) [*Koster et al.*, 2000] with various forcing datasets over the Red-Arkansas domain, divided into 308 catchments with a median linear scale of 35 km [*Reichle et al.*, 2008]. Specifically, three different base forcing datasets are used here (Table 2): The first is the high-resolution "truth" forcing interpolated to catchment space, and the second and third are derived from two different global reanalysis datasets that have been bias-corrected with additional observations [*Dirmeyer and Tan*, 2001; *Sheffield et al.*, 2006]. Three of the eight modeling scenarios are constructed by forcing CLSM with the three base forcing datasets; the rest are constructed by forcing CLSM with time-shifted (degraded) versions of these datasets.

[9] The R-values for the model estimates without data assimilation (Table 2) range from 0 to about 0.8 for surface and root zone soil moisture and from 0 to 0.65 for (monthly) evapotranspiration (ET). Since only monthly values of ET were saved in the *Sharif et al.* [2007] truth integration, we use monthly anomalies to assess the skill of ET estimates. Structural differences between TOPLATS and CLSM and scale differences imply that CLSM estimates are not perfect even when forced with the "truth" meteorological dataset. For example, in the TOPLATS "truth" the surface and root zone soil moisture refer to the top 5 cm and the top 40 cm of the soil column, respectively, whereas the corresponding CLSM surface and root zone values are for the top 2 cm and the top 100 cm, respectively.

[10] Because the "true" TOPLATS climatology of surface soil moisture differs from that of CLSM, each retrieval

Table 1. Skill of Synthetic Retrieval Datasets[a]

Retrievals	R_{sf}
R1	0.91
R2	0.86
R3	0.78
R4	0.70
R5	0.63
R6	0.57
R7	0.52
R8	0.48
R9	0.42
R10	0.35
R11	0.30
R12	0.26

[a] R_{sf} is the anomaly time series correlation coefficient of surface soil moisture with respect to truth data.

Table 2. Model Scenarios[a]

	Model Scenario							
	M1	M2	M3	M4	M5	M6	M7	M8
Base forcing dataset	F1	F2	F1	F3	F3	F1	F3	F1
Forcing shift [days]	0	0	7	0	7	28	28	365
R_{sf} [dimensionless]	0.76	0.63	0.50	0.41	0.37	0.28	0.22	−0.01
R_{rz} [dimensionless]	0.78	0.55	0.64	0.46	0.44	0.41	0.31	0.01
R_{ET} [dimensionless]	0.65	0.38	0.58	0.37	0.30	0.19	0.15	0.02

[a] Forcing datasets are from (F1) [*Sharif et al.*, 2007], (F2) [*Dirmeyer and Tan*, 2001], and (F3) [*Sheffield et al.*, 2006]. Anomaly time series correlation coefficient R with respect to truth data indicates model skill (without data assimilation) for (R_{sf}) surface soil moisture, (R_{rz}) root zone soil moisture, and (R_{ET}) monthly ET.

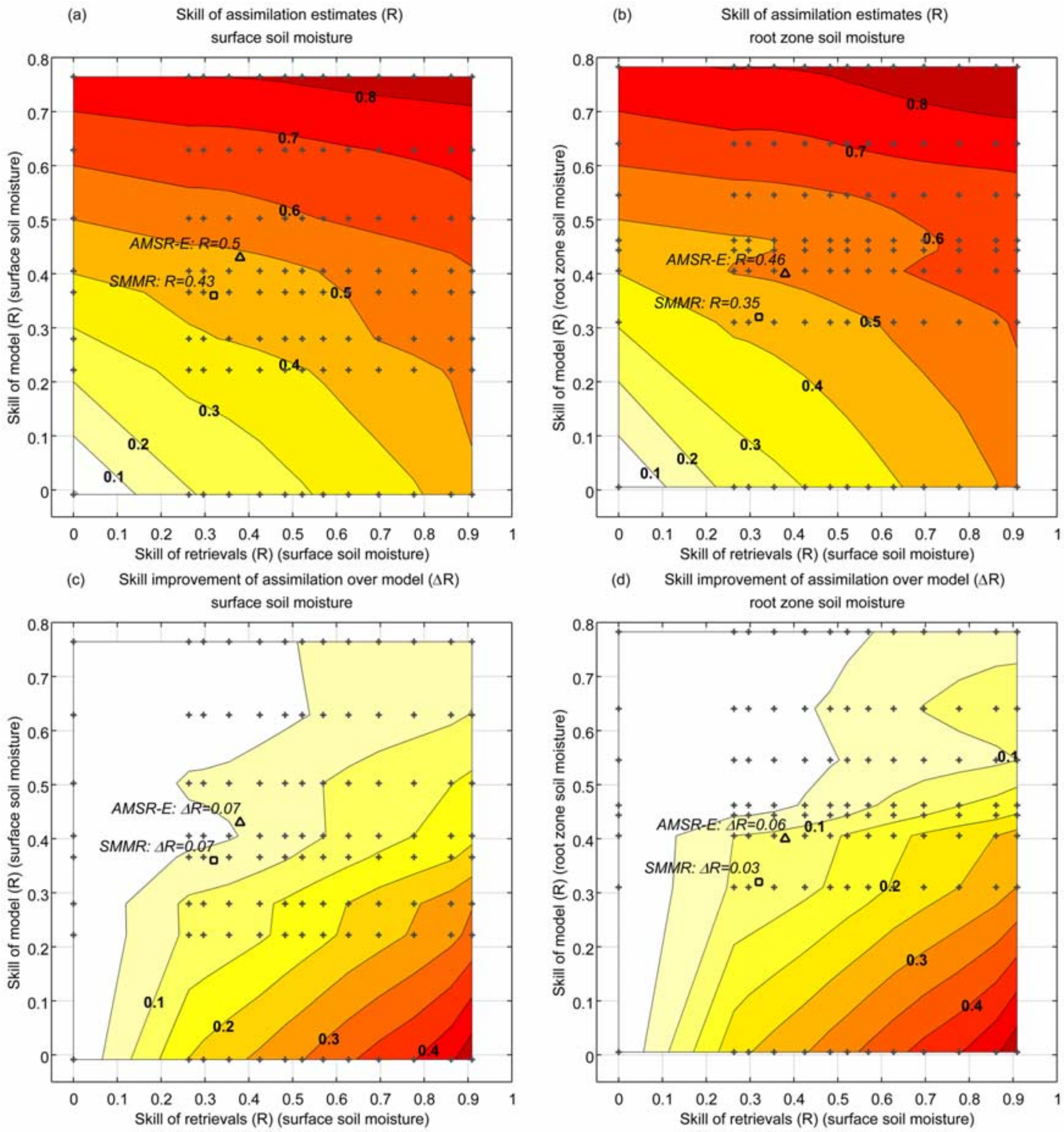

Figure 2. (a, b) Skill (R) and (c, d) skill improvement (ΔR) of assimilation product for surface (Figures 2a and 2c) and (Figures 2b and 2d) root zone soil moisture as a function of the (ordinate) model and (abscissa) retrieval skill. Skill improvement is defined as skill of assimilation product minus skill of model estimates. 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.

dataset is first scaled to the soil moisture climatology of each model scenario for bias removal [*Reichle and Koster*, 2004]. Next, each set of scaled retrievals is assimilated into each model scenario with a "one-dimensional" ensemble Kalman filter (EnKF) using 12 ensemble members [*Reichle et al.*, 2007]. Only soil moisture is updated in the EnKF, but ET is also impacted because it depends on soil moisture. Poorly specified model and observation error parameters negatively affect the quality of the assimilation products. Each of the $N_R \cdot N_M$ = 96 assimilation experiments must achieve near-optimal performance; otherwise the information contributed by the retrievals cannot be compared across experiments. We therefore use an adaptive EnKF and dynamically estimate the model and observation error parameters [*Reichle et al.*, 2008]. The sources and structure of the modeling uncertainties as well as initial estimates of the input error parameters are based on our experience with retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) [*Reichle et al.*, 2007, Table 1].

3. Results

[11] Each of the 96 assimilation experiments 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 shows the two dimensional surface (linearly interpolated from the 96

data points) corresponding to the surface soil moisture product. Filling the contour plot is a computational challenge. Each of the 96 crosses in the plot indicates the performance of a 19-year assimilation integration over the entire Red-Arkansas domain (and can be mapped back to the specific retrieval/model combination using Tables 1 and 2). 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.

[12] Figure 2 also shows skill improvement through data assimilation, defined as the skill of the assimilation product minus the skill of the model estimates (without assimilation). Specifically, Figures 2c and 2d show, for a given level of accuracy in the stand-alone model product, how much information can be added to the soil moisture products through assimilation of satellite retrievals of surface soil moisture with a given uncertainty. Note that 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.

[13] We can compare previously published skill levels with the results of Figure 2. For 23 locations across the contiguous United States having in situ observations appropriate for validation, *Reichle et al.* [2007, Table 2] report, for surface soil moisture, average R values of 0.38, 0.43, and 0.50 for AMSR-E retrievals, CLSM estimates, and their assimilation product, respectively. From the contours of Figure 2a we expect that for retrievals with R = 0.38 and a model with R = 0.43, the assimilation product would have skill of about R = 0.50, which is indeed consistent with the AMSR-E result (indicated with a triangle in Figure 2a). For root zone soil moisture, *Reichle et al.* [2007] show that 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 (1) the imperfect translation of information from the surface layer to the root zone in the data assimilation system and (2) 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 presented here). Figure 2 also includes the *Reichle et al.* [2007] results for assimilating retrievals from the historic Scanning Multichannel Microwave Radiometer (SMMR), which are similarly consistent with the contours. Note that R values for SMMR results are based on monthly mean data, and that the validating in situ data for the AMSR-E and SMMR results are not within the geographical domain of our synthetic experiment.

[14] Note that the skill of the surface soil moisture assimilation product (Figure 2a) does not always match or exceed the skill of the retrievals, in particular for poor model skill (along the abscissa), where the retrieval skill exceeds the assimilation skill by up to 0.4 (in terms of R). There are several reasons for this apparent mismatch. First, the retrieval skill can only be computed over times and locations for which retrievals are available, whereas the assimilation (and model) skill is computed over the entire experiment period and domain, because applications would presumably use the spatio-temporally complete product. Second, the truth and retrieval data are instantaneous data, whereas the assimilation products are daily averages (due to data storage constraints). Third, the R calculation penalizes the assimilation product because retrieval skill is computed from unscaled retrieval and truth data (which share a consistent climatology), whereas the assimilation product is evaluated without (non-linearly) scaling it back to the truth climatology. When controlling for the impact of these three factors, the skill of the surface soil moisture assimilation product along the abscissa is within 0.15 of the corresponding retrieval skill (not shown). The remaining difference may have a number of sources: (1) the assimilation system does not optimize R itself, (2) nonlinearities pervade the system, (3) the adaptive tuning of model error parameters and the scaling algorithm are imperfect, (4) differences exist in the layer depths for the assumed "truth" and CLSM, and (5) the ensemble size (12 members) used in the EnKF is small.

[15] Finally, Figure 3 shows the skill improvement (relative to the raw model product) for monthly mean ET estimates from the data assimilation system. As expected, the assimilation of surface soil moisture retrievals contributes the most when retrievals are skillful and the model is poor. Note, however, that ET estimates from the adaptive assimilation system are worse than model estimates when the model skill is very high to begin with. This is not the case with the non-adaptive filter (not shown) and can be explained by minor bias issues associated with the dynamic adjustment of model error parameters, a feature of the adaptive system [*Reichle et al.*, 2008]. These technicalities will be addressed in future research. Nevertheless, with the current system the assimilation of surface soil moisture retrievals yields, on average, modest improvements in ET estimates.

4. Conclusions

[16] With the DA-OSSE framework described here, we can quantify the information added to land assimilation products by satellite retrievals of surface soil moisture. In this paper, we present the general framework and show how the added information varies with both model skill and retrieval skill. The general framework permits detailed and comprehensive error budget analyses for data assimilation products. The framework can be used, for example, to study specific trade-offs in sensor design or ancillary data requirements, assessing the impact of each on the quality of the end-product that will be used in science and applications.

[17] A major component of the DA-OSSE is the determination of experiment-specific input error covariances that enable near-optimal assimilation performance and permit objective comparisons across experiments, as in the surface plots of Figure 2. We are confident that the adaptive EnKF used here [*Reichle et al.*, 2008] adequately meets this requirement of the analysis. The adaptive algorithm may be further improved through continued development and

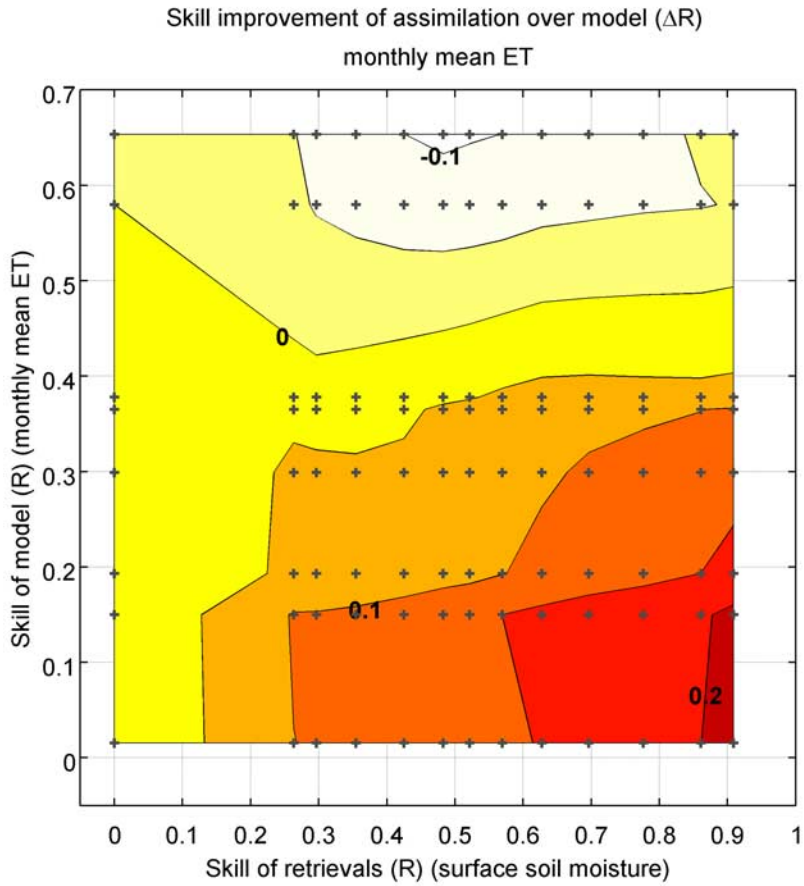

Figure 3. Skill improvement for monthly mean ET assimilation product. Abscissa, ordinate, and plus signs as in Figure 2.

testing, which would ostensibly lead to improved DA-OSSE results.

[18] The contribution of soil moisture retrievals to the skill of land data assimilation products also critically depends on the realism of the imposed model error estimates. If the assumed model errors (stemming from differences between the "truth" land model and CLSM, along with the imposed synthetic model and forcing errors) are not reflective of actual errors, the DA-OSSE may produce overly optimistic improvements in skill that will not be achieved when real observations are assimilated. We note again, however, the consistency between our synthetic results and those from existing AMSR-E and SMMR data assimilation exercises, which do employ real observations.

[19] Conceptually, extending the DA-OSSE to continental or global scales is straightforward, but computational costs may prohibit a comparable analysis (e.g., a continental-scale version of Figure 2). Similarly, it is straightforward to include higher-resolution soil moisture retrievals, such as those obtained from radar backscatter measurements, in the retrieval algorithm. An active-passive sensor is at the core of the SMAP mission concept. Again, however, computational costs limit the resolution that can be examined for a given domain size. Another possible extension of the DA-OSSE framework is to assimilate microwave brightness temperatures directly (as opposed to surface soil moisture retrievals) and then examine how uncertainties in the retrieval process may be mitigated through use of a priori information from the land surface model, notably surface soil temperature. As the focus on data assimilation products grows in future land surface satellite missions, our DA-OSSE framework presents an important end-to-end tool for mission planning and uncertainty assessment.

[20] **Acknowledgments.** We thank J. Sheffield, L. Tan, E. Wood, and P. Dirmeyer for the forcing data. Computing was supported by the NASA High End Computing Program.

References

Bindlish, R., T. J. Jackson, E. F. Wood, H. Gao, P. Starks, D. Bosch, and V. Lakshmi (2003), Soil moisture estimates from TRMM microwave imager observations over the southern United States, *Remote Sens. Environ.*, *85*, 507–515.

Crow, W. T., M. Drusch, and E. F. Wood (2001), An observation system simulation experiment for the impact of land surface heterogeneity on AMSR-E soil moisture retrievals, *IEEE Trans. Geosci. Remote. Sens.*, *39*, 1622–1632.

Crow, W. T., T. Chen, D. Entekhabi, P. Houser, A. Hsu, T. Jackson, E. Njoku, P. O'Neill, J. Shi, and X. Zhan (2005a), An observing system simulation experiment for Hydros radiometer-only soil moisture products, *IEEE Trans. Geosci. Remote Sens.*, *43*, 1289–1303.

Crow, W. T., R. D. Koster, R. H. Reichle, and H. O. Sharif (2005b), Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products, *Geophys. Res. Lett.*, *32*, L24405, doi:10.1029/2005GL024889.

Dirmeyer, P. A., and L. Tan (2001), A multi-decadal global land-surface dataset of state variables and fluxes, *COLA Tech. Rep. 102*, 43 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md.

Drusch, M. (2007), Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set, *J. Geophys. Res.*, *112*, D03102, doi:10.1029/2006JD007478.

Entekhabi, D., et al. (2004), The Hydrosphere State (Hydros) mission concept: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *IEEE Trans. Geosci. Remote Sens.*, *42*, 2184–2195.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, *105*, 24,809–24,822.

Njoku, E. G., T. L. Jackson, V. Lakshmi, T. Chan, and S. V. Nghiem (2003), Soil moisture retrieval from AMSR-E, *IEEE Trans. Geosci. Remote Sens.*, *41*, 215–229.

Peters-Lidard, C. D., M. S. Zion, and E. F. Wood (1997), A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes, *J. Geophys. Res.*, *102*, 4303–4324.

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

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe (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, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, doi:10.1029/2007WR006357, in press.

Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, and D. L. Toll (2003), The Global Land Data Assimilation System, *Bull. Am. Meteorol. Soc.*, *85*, 381–394, doi:10.1175/BAMS-85-3-381.

Sharif, H. O., W. Crow, N. L. Miller, and E. F. Wood (2007), Multidecadal high-resolution hydrologic modeling of the Arkansas-Red River Basin, *J. Hydrometeorol.*, *8*, 1111–1127, doi:10.1175/JHM622.1.

Sheffield, J., G. Goteti, and E. F. Wood (2006), Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, *J. Clim.*, *19*, 3088–3111.

Space Studies Board (2007), *Earth Science and Applications From Space: National Imperatives for the Next Decade and Beyond,* 400 pp., Natl. Acad. of Sci., Washington, D. C.

W. T. Crow, Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Room 104, Building 007, BARC-W, Beltsville, MD 20705-2350, USA.

R. D. Koster, S. P. P. Mahanama, and R. H. Reichle, NASA Goddard Space Flight Center, Code 610.1, Greenbelt Road, Greenbelt, MD 20771, USA. (rolf.reichle@nasa.gov)

H. O. Sharif, Civil Engineering Department, University of Texas at San Antonio, San Antonio, TX 78249, USA.