National Aeronautics and Space Administration

GMAO RESEARCH BRIEF

#2017-04, May 2017

Using SMAP Soil Moisture Data to Calibrate a Land Surface Model

Randal D. Koster, Sarith P. P. Mahanama, and Rolf H. Reichle

SUMMARY

Soil moisture retrievals from the Soil Moisture Active Passive (SMAP) mission are used to evaluate and calibrate the treatment of soil moisture recharge in the GMAO Catchment land surface model. The improvements lead to better simulations of soil moisture and streamflow, as demonstrated through comparisons against independent in situ data.

BACKGROUND

Satellite data have the potential to transform our understanding of the Earth system through the improved quantification of variations in that system. Much of the GMAO's focus is on assimilating satellite information directly into GMAO models in order to produce optimal, quantitative estimates for a comprehensive set of Earth system states. A complementary approach toward improved state estimation, however, is also worth considering – we can use the satellite information to improve the parameterizations underlying the models themselves.

NASA's Soil Moisture Active-Passive (SMAP) mission provides, among other things, estimates (or retrievals) of the moisture content in the top several centimeters of soil at a resolution of roughly 40 km across the globe and with a revisit time of less than three days. Various aspects of mission design (use of L-band frequencies, reduction of impacts from non-natural radio-frequency interference, etc.) give SMAP data an accuracy, and thus a hydrological utility, not achievable with other satellite-derived soil moisture datasets (see, e.g., Koster et al. 2016). Given that the data have only been available to users since spring of 2015, much of the wealth of hydrological information contained within them is still waiting to be tapped.

The GMAO has a strong connection to the SMAP mission, as it hosts the production of the SMAP Level 4 products – enhanced products (including root zone soil moisture and carbon flux estimates) achieved through data assimilation. Data assimilation, however, is not the only means by which the SMAP data can interact with GMAO modeling. Soil moisture behavior as revealed by the SMAP retrievals could also guide the evaluation and further development of the GMAO land surface model (LSM) itself. This land surface model, referred to here as the Catchment LSM, features state-of-the-art treatments of land processes, including an explicit representation of the effects of subgrid soil moisture heterogeneity on the surface energy and water balances (Koster et al. 2000). This LSM's treatment of near-surface moisture and how it relates to the root zone, however, has never been properly calibrated and thus may benefit from a careful analysis of the SMAP data.

The idea examined here is simple: SMAP provides for the first time a highly accurate global picture of how surface soil moisture varies in time – how it increases, for example, with precipitation and how quickly a rainfall-induced soil moisture anomaly dissipates. Through careful joint analysis of SMAP data and the structure of the Catchment LSM parameterizations, the latter can be modified to behave more realistically.

The model behavior targeted for improvement in the present study is illustrated in Figure 1. Figure 1a shows the time series (May – September, 2016) of SMAP retrievals at a grid cell containing the Little Washita watershed in southwestern Oklahoma (O'Neill et al. 2016), and Figure 1b shows the corresponding time series of soil moisture simulated by the operational version of the GMAO Catchment model when driven with observations-based meteorological forcing (rainfall, air temperatures, wind speeds, etc.) The two-time series have a distinctly different character. Consider, for example, the speed of soil moisture drydown following the rainfall-induced increase on day 43. According to the

SMAP data, the soil dries quickly, losing most of the added moisture within a few days and returning to its driest state in about ten days. In the default land surface model, on the other hand, the drydown following the imposition of rainfall on day 43 is clearly more gradual.

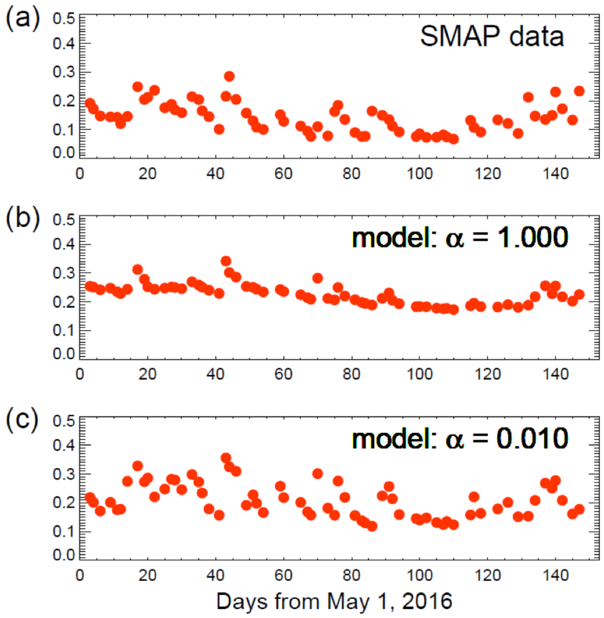

Figure 1. Time series of SMAP retrievals (a) and model-simulated near-surface moisture (b,c) at a grid cell centered on the Little Washita watershed in southwestern Oklahoma. The first model simulated time series uses a value of 1 for the calibrated parameter α (the default value in GEOS-5 systems), and the second uses an α value of 0.01.

The Catchment LSM's formulation of surface soil moisture dynamics includes a treatment of the replenishment of drying soil via recharge from below. A parameter (referred to here as α) was recently incorporated into the code to control this replenishment – α reduces the ability of soil moisture to flow upward against gravity in non-equilibrium situations, to account for the fact that near-surface soils in nature are more heterogeneous than those tested in laboratories. The value of α turns out to have a first order impact on the character of the simulated soil moisture, as illustrated in Figure 1c. At this grid cell, replacing α's default value of 1 with a value of 0.01 produces a better match (in terms of temporal variability, speed of drydown, etc.) with the SMAP data.

In a calibration exercise, a number of simulations, each utilizing a different value for α, generated soil moisture time series for 2015-2016 across the continental US and portions of Canada and Mexico. At each 36km × 36km grid cell, by comparing the different simulated time series to the local time series of SMAP retrievals, we were able to determine the single α value that produces the closest reproduction of the SMAP retrievals. (Closeness here is measured by the temporal correlation coefficient between the SMAP retrievals and the model estimates.) Figure 2 shows a map of these optimized α values. Notice that the default value of 1 (again, standard in GEOS-5 operations) works best at only a handful of locations.

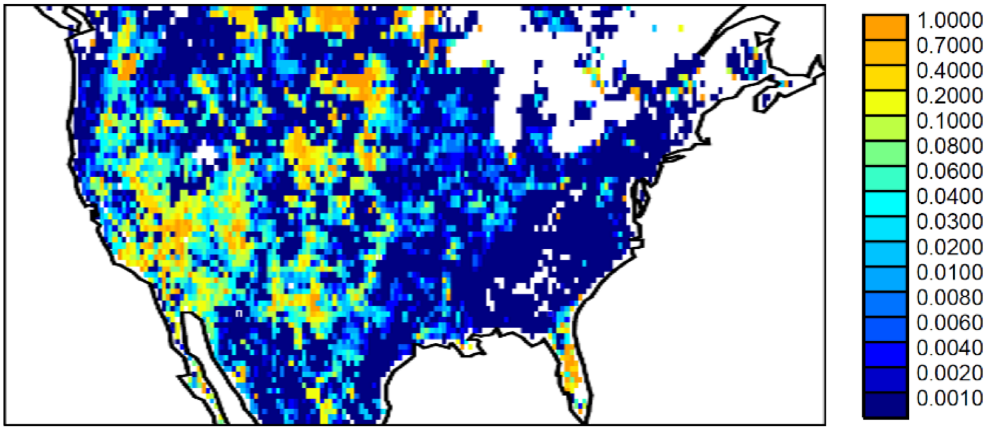

Figure 2. Optimal values of studied model parameter, as determined by optimization against SMAP retrieval time series.

We now address an important question: when a simulation utilizing the default α value of 1 everywhere (the "control" simulation, run over the years 2001-2016) is compared to one that uses the optimized α values in Figure 2 (the "experiment" simulation, run over the same years), is the latter's skill in reproducing independent (i.e., non-SMAP) hydrological measurements higher? The answer is yes:

a) *In situ measurements of soil moisture at SMAP core validation sites.* The SMAP mission is working with local partners who provide in situ soil moisture measurements suitable for validating SMAP data products. These measurements are comprehensive enough spatially to provide reliable estimates of large-scale, areally-averaged soil moisture. Four of these sites provide data for several years prior to the launch of SMAP. Figure 3 shows the temporal correlations between the Catchment LSM results and the in situ measurements at each of these four sites, with results for the control simulation in red and those for the experiment simulation in blue. In all cases, and especially for Reynolds Creek, Walnut Gulch, and Little River (for which the differences in the correlations shown are significant at the 99.9% confidence level, even under the assumption that the number of independent data values is 1/10 the total number of values), the optimized α values have led to an improved agreement with the in situ measurements for the ~10 year subset of the simulation period over which measurements were available.

The temporal correlation is the measure of simulation skill that we are most interested in improving here, given that the information we want to reproduce most with the Catchment LSM is largely contained within the time variability of the observed soil moisture. An "error" in the absolute magnitude of a simulated soil moisture variable (from any LSM) is in fact largely a reflection of the model-dependent nature of that variable; differences in observational and LSM-generated soil moistures are expected, do not necessarily affect model performance, and are largely correctable as needed after the fact (Koster et al. 2009). This said, it is still of interest to examine the impacts of our calibrated α values on the absolute magnitudes of the simulated soil moistures, as measured by the root mean square error (RMSE) relative to the observations. The RMSE decreased by close to 30% for Reynolds Creek, Walnut Gulch, and Little River, and it increased by about 10% for Little Washita (not shown). This largely reflects a reduction in the overall model bias. Results for the unbiased root mean square error (ubRMSE) were mixed, being better for only 2 out of the 4 sites (not shown). The reader is referred to Entekhabi et al. (2010) for a discussion of the connections between the temporal correlation, the RMSE, and the ubRMSE skill metrics.

b) *In situ measurements of soil moisture at sparse network sites.* A number of in situ soil moisture measurement sites are encompassed by the USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN; Schaefer et al. 2007) and the US Climate Reference Network (USCRN; Diamond et al. 2013, Bell et al. 2013). While these sites, unlike the above core sites, are not designed to provide the large-area estimates produced by both SMAP and the LSM, both networks have the advantage of encompassing the continental US. and thereby covering a broad range of soil textures and background climates (Reichle et al. 2016).

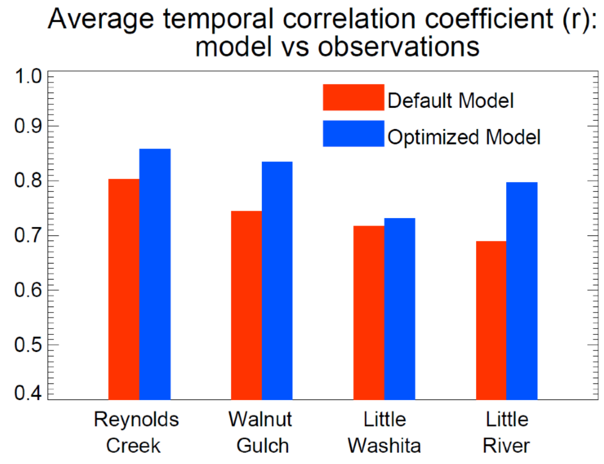

Figure 3. The LSM's ability to reproduce observed surface soil moisture (as measured by the temporal correlation coefficient) at the four SMAP core validation sites with multiple years of measurements. Results are shown for the control simulation (using the default model, red bars) and the experiment simulation (using the optimized model, blue bars).

Figure 4 shows the temporal correlation, averaged over the component sites of each network, between the simulated and observed time series of surface moisture for both the control and the experiment simulations. The model's ability to reproduce the observed surface moisture is improved with the optimized α values. This is particularly true for the SCAN network; the improvement for this network extends well beyond the indicated uncertainties in the averages, which are quantified here as in Reichle et al. (2016). (By these uncertainty measures, the improvement for the USCRN network, while positive, may not be significant.) Results for other soil moisture metrics (not shown) are somewhat mixed but still indicate improved overall performance through the use of the optimized α values. While for both networks the optimized values lead to reduced temporal correlations between simulated root zone moisture contents and corresponding measurements, the reductions are not significant, lying well within the much larger uncertainty levels for the root zone metric – soil moisture measurements in the root zone are fewer in number and in some ways are more difficult to interpret. The absolute bias for surface moisture is significantly improved with the optimized α values for both networks, as is the ubRMSE for the USCRN network.

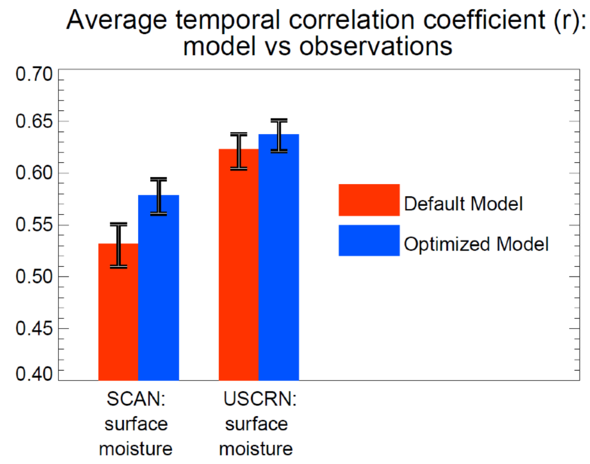

Figure 4. The land model's ability to reproduce observed surface soil moisture at two networks of sites covering the continental US, as measured by the average value of the temporal correlation coefficient over the sites. Results are shown for the control simulation (using the default model, red bars) and the experiment simulation (using the optimized model, blue bars). Error bars indicate 95% confidence intervals.

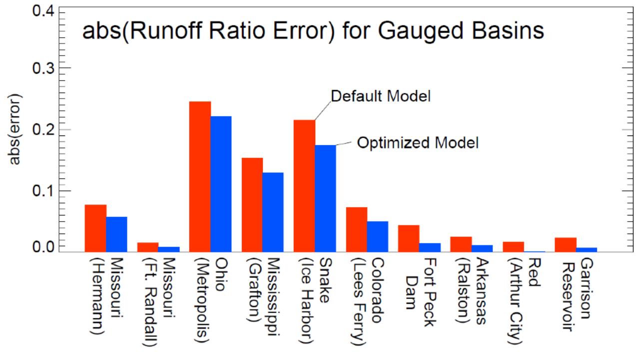

Figure 5. Absolute errors in runoff ratio for the ten largest basins examined by Mahanama et al. (2012). (Results for the smaller basins are similar.) Red bars are for the control simulation and blue bars are for the experiment simulation.

c) Observed runoff ratios. Mahanama et al. (2012) evaluated the streamflows produced by the GMAO's Catchment land surface model against time series of naturalized streamflows (derived from stream gauge observations) in a number of hydrological basins in the continental US. These naturalized observations are used in Figure 5 to evaluate the relative performance of the control and experiment simulations in reproducing observed runoff ratios (long-term average streamflow divided by the long-term average rainfall in the area upstream of the stream gauge site). While the experiment simulation still produces large errors in runoff ratio in three of the basins (an error that can be treated using other approaches – see Koster and Mahanama 2012 for further discussion of this error), in every case the experiment simulation improves over the control simulation – the SMAP-based tuning of the recharge parameter has translated into an improvement in streamflow simulation.

In summary, utilizing SMAP data to calibrate a parameter in the surface recharge module of the GMAO's Catchment land surface model leads to improved model performance, as demonstrated by comparisons of simulated hydrological states and fluxes to several sets of SMAP-independent observations (as well as to others not shown here). Figure 6 puts this result in perspective, highlighting the complementary nature of this calibration effort and the data assimilation work also underway in the GMAO. Figure 6a shows the standard land modeling approach to producing soil moisture states; a land model driven with observations-based meteorological forcing (rainfall, etc.) produces, as a matter of course, estimates for hydrological states (e.g., soil moisture) and fluxes (e.g., evapotranspiration and runoff). In the GMAO, SMAP data are assimilated into the system (Figure 6b), guiding the states toward more realistic values. This is quite distinct from the calibration effort described above (Figure 6c) wherein the SMAP data are used to change the model itself. We can in fact speculate that a combination of the two efforts (or, more ambitiously, the inclusion of parameter estimation along with soil moisture state estimation in an enhanced GMAO data assimilation effort) would lead to state and flux estimates of even higher quality (Figure 6d).

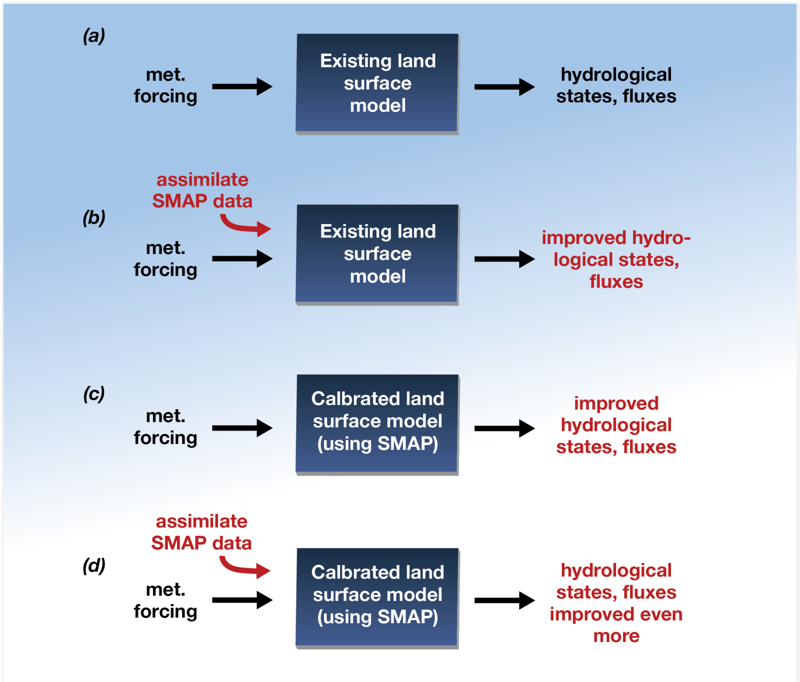

Figure 6. Schematic showing means by which SMAP data can improve model-based estimates of hydrological states and fluxes

References

Bell, J., and Coauthors, 2013: U.S. Climate Reference Network soil moisture and temperature observations. *J. Hydrometeor*., **14**, 977–988, doi:10.1175/JHM-D-12-0146.1.

Diamond, H., and Coauthors, 2013: U.S. Climate Reference Network after one decade of operations: Status and assessment. *Bull. Amer. Meteor. Soc*., **94**, 485–498, doi:10.1175/BAMS-D-12-00170.1.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010: Performance metrics for soil moisture retrievals and application requirements. *J. Hydromet.*, 11, 832-840.

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, 24809-24822.

Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *J. Climate*, 22, 4322-4335.

Koster, R. D., and S. P. P. Mahanama, 2012: Land surface controls on hydroclimatic means and variability. *J. Hydromet.*, 13, 1604-1620.

Koster, R. D., L. Brocca, W. T. Crow, M. S. Burgin, and G. J. M. De Lannoy, 2016: Precipitation estimation using L-band and C-band soil moisture retrievals. *Water Resources Research*, 52, 7213–7225, doi:10.1002/2016WR019024.

Mahanama, S., B. Livneh, R. Koster, D. Lettenmaier, and R. Reichle, 2012: Soil moisture, snow, and seasonal streamflow forecasts in the United States. *J. Hydromet.*, 13, 189-203.

O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish. 2016: *SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture, Version 4*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: 10.5067/XPJTJT812XFY. Accessed October 2016.

Reichle, R. H., G. J. M. De Lannoy, Q. Liu, J. V. Ardizzone, F. Chen, A. Colliander, A. Conaty, W. Crow, T. Jackson, J. Kimball, R. D. Koster, and E. B. Smith (2016): Soil Moisture Active Passive Mission L4_SM Data Product Assessment (Version 2 Validated Release). *NASA GMAO Office Note, No. 12* (Version 1.0), National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 55pp.

Schaefer, G. L., M. H. Cosh, and T. J. Jackson, 2007: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *J. Atmos. Oceanic Technol*., **24**, 2073– 2077.