Bias reduction in short records of satellite soil moisture

R. H. Reichle[1,2] and R. D. Koster[1]

R. H. Reichle and R. D. Koster, NASA-GSFC, Code 900.3, Greenbelt Rd, Greenbelt, MD 20771, USA. (reichle@gmao.gsfc.nasa.gov)

[1]NASA Goddard Space Flight Center,

Global Modeling and Assimilation Office,

Greenbelt, MD, USA.

[2]University of Maryland, Baltimore

County, Goddard Earth Sciences and

Technology Center, Baltimore, MD, USA.

Although surface soil moisture data from different sources (satellite retrievals, ground measurements, and land model integrations of observed meteorological forcing data) have been shown to contain consistent and useful information in their seasonal cycle and anomaly signals, they typically exhibit very different mean values and variability. These biases pose a severe obstacle to exploiting the useful information contained in satellite retrievals through data assimilation. A simple method of bias removal is to match the cumulative distribution functions (cdf) of the satellite and model data. However, accurate cdf estimation typically requires a long record of satellite data. We demonstrate here that by using spatial sampling with a 2 degree moving window we can obtain local statistics based on a one-year satellite record that are a good approximation to those that would be derived from a much longer time series. This result should increase the usefulness of relatively short satellite data records.

1. Motivation

Long-term in situ measurements of soil moisture are limited to parts of Eurasia and small sections of North America [*Robock et al.*, 2000]. To derive global soil moisture distributions, as might be needed for the initialization of seasonal forecast systems [*Koster et al.*, 2004], two alternative data sources are often considered. First, useful global distributions of soil moisture can be produced by a land surface model when forced with observed precipitation, radiation, and other meteorological data [*Rodell et al.*, 2003]. Second, satellite sensors can provide passive C-band (6.6 GHz) or L-band (1.4 GHz) radiance measurements that can be interpreted in terms of surface soil moisture content [*Owe et al.*, 2001; *Jackson et al.*, 2002]. However, the model-based product is subject to the many limitations of the model used, to errors in the specification of vegetation and soil parameters, and to errors in the forcing data. The satellite data, for their part, are not available everywhere and not available continuously. Also, satellite retrievals represent only a shallow near-surface layer and do not provide critical information about soil moisture in the root zone.

Many have argued that a land assimilation system that merges satellite retrievals and model soil moisture will provide optimal global estimates of the state of the land surface. In a data assimilation system, a model-generated soil moisture is "corrected" toward an observational estimate, with the degree of correction determined by the levels of error associated with each. Idealized analyses with large-scale assimilation systems, using synthetic (model-generated) observational data, demonstrate the potential of the approach [*Walker and Houser*, 2001; *Reichle and Koster*, 2003].

Synthetic data studies, however, avoid a fundamental difficulty associated with satellite data assimilation: the strong biases that exist between satellite-based and model-based soil moisture estimates [*Reichle et al.*, 2004]. The top panel of Figure 1 shows, for example, the difference between the mean near-surface soil moisture field retrieved from the C-band Scanning Multichannel Microwave Radiometer (SMMR) over the period 1979–1987 [*De Jeu*, 2003] and that simulated by the NASA Catchment land surface model [*Koster et al.*, 2000] for the same period. Despite global coverage of the satellite, soil moisture retrievals are not available in areas that contain frozen soil, a significant fraction of surface water, or dense vegetation. As for the model, it was forced with reanalysis data that have been corrected by observations as much as possible [*Berg et al.*, 2003]. Precipitation – arguably the most critical input for accurate soil moisture modeling – is based on a merged product of satellite and gauge data from the Global Precipitation Climatology Project (GPCP, Version 2) [*Huffman et al.*, 1997]. Model soil moisture data have been generated at the exact times and locations of SMMR retrievals, to ensure maximum compatibility of the two data sets. The model's computational units are irregularly shaped catchments (or watersheds) with an average area of about 2500km^2 [*Reichle et al.*, 2004].

Figure 1 shows that across the globe, SMMR retrievals are typically wetter than model soil moisture, except in the eastern half of North America, northern Eurasia, and the Sahel. The bottom panel of Figure 1 shows the corresponding differences in the standard deviation (std) of the instantaneous fields, that is the bias in the std. SMMR retrievals exhibit more variability than model soil moisture across North America, in northern Eurasia, southern Africa, and southern Australia. Elsewhere, particularly in India, SMMR re-

trievals are less variable in time than model soil moisture. (Note that *Reichle et al.* [2004] used monthly data as opposed to instantaneous data. Consequently, the time series std in the present paper is about twice as large.)

The satellite and model data clearly differ in their statistical moments. These biases are not uniform but are spatially distributed with complex patterns and with magnitudes on the order of the dynamic range of the signal. Furthermore, the relative accuracy of the two datasets cannot be objectively determined. *Reichle et al.* [2004] demonstrate that neither is clearly superior when compared to the limited array of in situ point observations. Such bias is unavoidable, both now and in the foreseeable future. Even if the satellite retrievals could be considered unbiased relative to nature, simulated soil moisture contents reflect the many necessary simplifications imposed in the land surface model and should arguably be considered model-specific "indices of wetness" rather than quantities that can be measured in the field [*Koster and Milly*, 1997]. (See also [*Entin et al.*, 1999] for a strong demonstration of the model-specific nature of simulated soil moisture.) To merge successfully the satellite observations with the model data, biases across the statistical moments must be quantified and corrected. In effect, the satellite-based moisture contents must be converted ("scaled") into moisture contents consistent with the land surface model used.

Herein lies a major problem. In order to correct the biases, the temporal statistical moments of both the simulated soil moisture and the satellite-derived soil moisture must be well-established, and without further assumptions, this would require many years of data for each. While such data exist for the model-generated estimates, the passive C-band

Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) has become operational only in June 2002. Two passive L-band sensors, the Soil Moisture and Ocean Salinity (SMOS) mission [*Kerr et al.*, 2001] and the Hydrosphere State (HY-DROS) mission [*Entekhabi et al.*, 2004], are still in their planning stages. Moreover, the expected lifetime of these sensors is only a few years. Given the tremendous investment placed in the satellites, researchers are pressured to use the satellite products in a data assimilation system as soon as they are produced.

We thus require a strategy for making use of a short record of satellite data under the constraint that we do not have global estimates of the data's temporal statistical moments. (Knowledge of the data's uncertainty does not ameliorate the problem, since we also do not know the true statistical moments.) Here, we present a viable strategy involving the ergodic substitution of variability in space for variability in time. To demonstrate the strategy's effectiveness, we use a single year of the SMMR soil moisture record to determine scaling parameters that convert an instantaneous field of SMMR retrievals into a soil moisture field consistent with the land surface model used. These scaling parameters are then applied to the full 9 years of SMMR data. When the statistical moments of the 9 years of scaled satellite data are compared to those of the simulated soil moisture fields, the biases in the mean and std are seen to be much smaller than those in Figure 1, indicating that the scaling, based on a single year of data, was a success. These scaled data can be merged more reliably with land model simulations in a data assimilation system.

2. Approach

Our strategy for bias removal is to match the cumulative distribution function (cdf) of the satellite retrievals to the cdf of the model soil moisture. Similar cdf matching techniques have been used, for example, to establish reflectivity-rainfall relationships for calibration of radar or satellite observations of precipitation [*Atlas et al.*, 1990; *Anagnostou et al.*, 1999]. Our approach is illustrated in Figure 2, which shows cdf's of surface soil moisture at a particular location in the Northern Great Plains (46N, 100W). At this location, SMMR retrievals are considerably wetter and exhibit more variability than model soil moisture. The scaled satellite retrieval $x\prime$ is given by the solution to

$$cdf_m(x\prime) = cdf_s(x), \qquad (1)$$

where cdf_s and cdf_m denote the cdf's of the satellite and model soil moisture, respectively, and x is the unscaled satellite soil moisture. Since assimilation systems ingest instantaneous satellite retrievals at the local scale, equation (1) is solved at each location after estimating the corresponding local cdf's. The bold arrows in Figure 2 illustrate schematically how the unscaled satellite retrieval x is converted into the scaled retrieval $x\prime$ (using the "ideal" cdf estimated from 1979–1987 SMMR retrievals.) Note that cdf matching corrects all moments of the distribution function regardless of its shape, subject to statistical errors associated with a limited sample size. In practice, we can expect meaningful estimates only for the first few moments, and limit ourselves to analyzing the mean, std, and skewness.

Our goal is to obtain an acceptable estimate of the cdf used for scaling from only the first year of SMMR data. In order to control statistical noise in the cdf estimate, we estimate

the temporal statistics at a given site by using observations at neighboring locations that are within a chosen distance from the site. In other words, we apply a moving spatial sampling window to the computation of the statistics and implicitly assume some degree of ergodicity in the data. We then use this approximate estimate of the cdf (based on just one year of SMMR data) to solve equation (1) and obtain 9 years of scaled SMMR retrievals from the 9 years of unscaled SMMR data. Finally, we compare the statistics of the scaled dataset to those of the model soil moisture. Note that the model cdf used for scaling is based on model soil moisture from 1979 to 1987.

3. Results

Robust estimation of statistics requires sufficient data. Our cutoff criterion for estimating the local cdf is that at least 100 measurements must be available within the spatial sampling window. Naturally, the degree of global coverage of cdf estimates obtained in this way increases rapidly with the size of the window, but so does the error associated with the ergodicity assumption. We are thus faced with a trade-off between coverage and error. To quantify this trade-off, we tried several spatial sampling windows with radii ranging from 0 to 5 degrees.

Since the ergodicity error increases monotonically with the window size, a reasonable approach is to use the minimum window size for which the coverage of the approximate cdf estimates (obtained from one year of SMMR data) is almost complete relative to the coverage obtained when the cdf is estimated from 9 years of SMMR data without spatial sampling. For SMMR, this approach suggests that the optimal spatial sampling window has a radius of 2 degrees. The approximate SMMR cdf based on 1979 data only and

using a 2 degree spatial sampling window is illustrated in Figure 2 for the representative location in the Northern Great Plains. The rough agreement with the full SMMR cdf is an indication of the validity of the ergodicity assumption. When 9 years of SMMR retrievals are scaled using this approximate cdf estimate, the cdf of the resulting scaled SMMR retrievals (also shown in Figure 2) is much closer to the model cdf than before scaling.

Figure 3 shows global maps of the biases obtained when the statistics of the scaled SMMR retrievals (using approximate cdf estimates) are compared to those of the model soil moisture. As in Figure 1, the biases in Figure 3 are computed for the period from 1979 to 1987. While there is some bias left, scaling with the approximate cdf based on just one year of satellite data clearly removes much of the bias seen in Figure 1. The biases after scaling depend only weakly on the particular year used for estimating the cdf. This is not surprising, given that the bias in the mean is much larger than the interannual variability. Globally averaged, the bias in the mean (or std; or skewness) is reduced by 80% (or 55%; or 25%) when only a single year of SMMR retrievals is used to estimate the cdf used for scaling. Since cdf estimation involves finite size bins, even scaling with the "ideal" cdf that is computed from the entire SMMR history does not completely eliminate the biases, particularly in the higher moments. In the ideal case, the bias in the mean (or std; or skewness) is reduced by 98% (or 90%; or 55%).

4. Conclusions

We use the 9-year SMMR record to demonstrate that temporal sampling of SMMR soil moisture retrievals can be traded off against spatial sampling. Robust estimation

of the statistics for bias removal via cdf matching was accomplished using only a one-year satellite record. When only one year of data is available and the cutoff criterion for computation of statistics is set to 100 data points, a reasonable approach is to estimate the cdf used for scaling by applying a spatial sampling window with a 2 degree radius. In this case, the global average bias in the mean of the scaled SMMR 9-year dataset (relative to model soil moisture) is reduced by 80% when compared to the original bias of the unscaled SMMR retrievals. For the bias in the std (skewness), cdf matching permits bias reduction by 55% (25%). With our method, current and future satellite retrievals of soil moisture can be assimilated more confidently in near-real time using only a one-year climatology.

Although differences in the spatial and temporal mean and variability between state-of-the-art land surface modeling systems are substantial, our method does not depend on the particular model used precisely because we scale the satellite retrievals to be consistent with the given model. Finally, AMSR-E and future sensors yield improved measurements of brightness temperatures compared to SMMR. Most importantly, AMSR-E offers higher sampling rates than SMMR (around 2.5 times higher spatial resolution and wider swath width), which may permit reducing the size of the spatial sampling window and hence the ergodicity error. Nevertheless, the retrievals used here are based on a state-of-the-art algorithm, as is the modeling system. Therefore, the underlying errors in the retrieval algorithm, the land surface model, and the surface meteorological forcing data are unlikely to change significantly in the near future. Our approach presents a valuable tool for the imminent operational use of AMSR-E and future soil moisture retrievals.

References

Anagnostou, E. N., A. J. Negri, and R. F. Adler, Statistical adjustment of satellite microwave monthly rainfall estimates over amazonia, *J. Appl. Meteorol.*, *38*(11), 1590–1598, 1999.

Atlas, D., D. Rosenfeld, and D. B. Wolff, Climatologically tuned reflectivity-rain rate relations and links to area-time integrals, *J. Appl. Meteorol.*, *29*(11), 1120–1135, 1990.

Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser, The impact of bias correction to reanalysis products on simulations of north american soil moisture and hydrological fluxes, *J. Geophys. Res.*, *108*(D16), 4490, DOI 10.1029/2002JD003334, 2003.

De Jeu, R. A. M., Retrieval of land surface parameters using passive microwave remote sensing, PhD dissertation, Vrije Universiteit Amsterdam (Netherlands), Dept. of Geo-Environmental Sciences, 2003.

Entekhabi, D., et al., The Hydrosphere State (HYDROS) mission concept: an Earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *Trans. Geos. Remote Sens.*, in press, 2004.

Entin, J. K., A. Robock, K. Y. Vinnikov, V. Zabelin, S. Liu, A. Namkai, and T. Adyasuren, Evaluation of Global Soil Wetness Project soil moisture simulations, *J. Meteorol. Soc. Japan*, *77*(1B), 183–198, 1999.

Huffman, G. J., et al., The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, *Bull. Amer. Meteorol. Soc.*, *78*(1), 5–20, 1997.

Jackson, T. J., A. Gasiewski, A. Oldak, M. Klein, E. Njoku, A. Yevgrafov, S. Christiani, and R. Bindlish, Soil moisture retrieval using the C-band polarimetric scanning radiometer during the Southern Great Plains 1999 Experiment, *IEEE Trans. Geosci. Remote Sens.*, *40*(10), 2151–2161, 2002.

Kerr, Y. H., P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger, Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1729–1735, 2001.

Koster, R. D., and P. C. D. Milly, The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models, *J. Climate*, *10*(7), 1578–1591, 1997.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, A catchment-based approach to modeling land surface processes in a general circulation model, 1: Model structure, *J. Geophys. Res.*, *105*(20), 24,809, 2000.

Koster, R. D., M. J. Suarez, P. Liu, U. Jambor, A. A. Berg, M. Kistler, R. H. Reichle, M. Rodell, and J. Famiglietti, Realistic initialization of land surface states: Impacts on subseasonal forecast skill, *J. Hydrometeor.*, in press, 2004.

Owe, M., R. de Jeu, and J. Walker, A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1643–1654, 2001.

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

Reichle, R., R. D. Koster, J. Dong, and A. A. Berg, Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation, *J. Hydrometeor.*, *5*(3), 430–442, 2004.

Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, The global soil moisture data bank, *Bull. Amer. Meteorol. Soc.*, *81*, 1281–1299, 2000.

Rodell, M., et al., The Global Land Data Assimilation System, *Bull. Amer. Meteorol. Soc.*, *85*(3), 381–394, DOI 10.1175/BAMS-85-3-381, 2003.

Walker, J. P., and P. R. Houser, A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *J. Geophys. Res.*, *106*(11), 11,761–11,774, 2001.
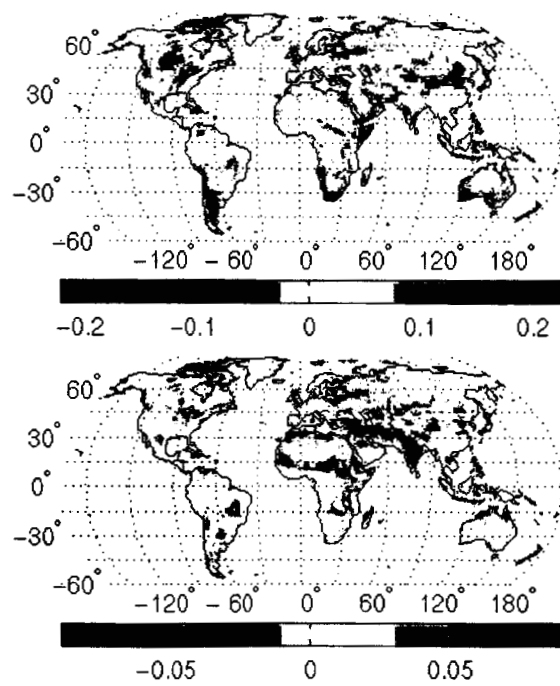
Figure 1: Difference in 1979–1987 (top) mean and (bottom) standard deviation of SMMR soil moisture retrievals and model soil moisture [m^3m^{-3}].
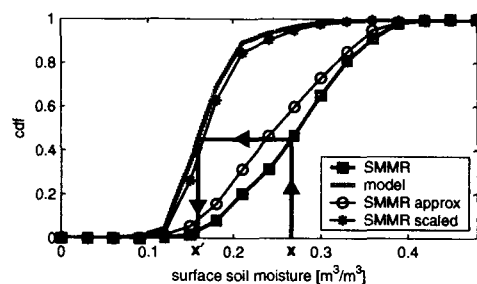
Figure 2: Cdf estimates at 46N, 100W: (Squares) 1979–1987 SMMR retrievals, (Solid line, no marker) 1979–1987 model soil moisture, (Circles) 1979 only SMMR retrievals using a spatial sampling window of 2 degree radius (approximate cdf), (Stars) 1979–1987 SMMR retrievals scaled with approximate cdf.
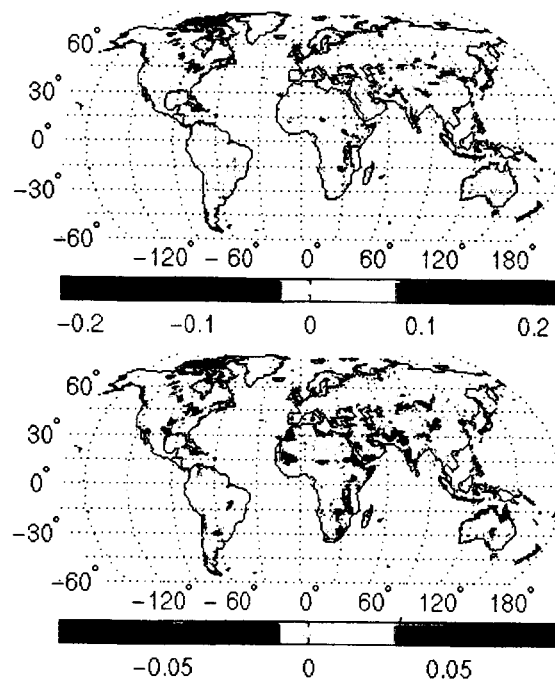
Figure 3: Same as Figure 1 except that SMMR retrievals were scaled with an approximate cdf estimated from 1979 only SMMR data using a spatial sampling window (2 degree radius).

"Bias reduction in short records of satellite soil moisture"

Reichle, R H and R D Koster

Abstract

Although surface soil moisture data from different sources (satellite retrievals, ground measurements, and land model integrations of observed meteorological forcing data) have been shown to contain consistent and useful information in their seasonal cycle and anomaly signals, they typically exhibit very different mean values and variability. These biases pose a severe obstacle to exploiting the useful information contained in satellite retrievals through data assimilation. A simple method of bias removal is to match the cumulative distribution functions (cdf) of the satellite and model data. However, accurate cdf estimation typically requires a long record of satellite data. We demonstrate here that by using spatial sampling with a 2 degree moving window we can obtain local statistics based on a one-year satellite record that are a good approximation to those that would be derived from a much longer time series. This result should increase the usefulness of relatively short satellite data records.