Global Soil Moisture from Satellite Observations, Land Surface Models, and Ground Data: Implications for Data Assimilation

ROLF H. REICHLE

Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, and Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

RANDAL D. KOSTER

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

JIARUI DONG

Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, and Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland

AARON A. BERG

Department of Geography, University of Guelph, Guelph, Ontario, Canada

(Manuscript received 30 September 2003, in final form 19 December 2003)

ABSTRACT

Three independent surface soil moisture datasets for the period 1979–87 are compared: 1) global retrievals from the Scanning Multichannel Microwave Radiometer (SMMR), 2) global soil moisture derived from observed meteorological forcing using the NASA Catchment Land Surface Model, and 3) ground-based measurements in Eurasia and North America from the Global Soil Moisture Data Bank. Time-average soil moisture fields from the satellite and the model largely agree in the global patterns of wet and dry regions. Moreover, the time series and anomaly time series of monthly mean satellite and model soil moisture are well correlated in the transition regions between wet and dry climates where land initialization may be important for seasonal climate prediction. However, the magnitudes of time-average soil moisture and soil moisture variability are markedly different between the datasets in many locations. Absolute soil moisture values from the satellite and the model are very different, and neither agrees better with ground data, implying that a ''correct'' soil moisture climatology cannot be identified with confidence from the available global data. The discrepancies between the datasets point to a need for bias estimation and correction or rescaling before satellite soil moisture can be assimilated into land surface models.

1. Introduction

Interest in global soil moisture observations and data assimilation has been growing steadily over the past few years. Accurate initialization of the vertical profile of soil moisture, for example, may be key to successful seasonal prediction of midlatitude summer precipitation over land. For the best possible soil moisture initial conditions, data assimilation may be used to combine satellite retrievals of surface soil moisture with information from the land surface model and its associated meteorological forcing inputs. The data assimilation system is designed to propagate this surface information into the deeper soil and thereby provide improved estimates of the vertical profile of soil moisture. While there has been considerable progress in the methodological development of soil moisture data assimilation (Houser et al. 1998; Walker and Houser 2001; Margulis et al. 2002; Reichle et al. 2002; Reichle and Koster 2003; Crow and Wood 2003; Seuffert et al. 2003), global observations of soil moisture have been lacking.

It is possible to retrieve surface soil moisture from low-frequency active and passive microwave data collected by satellite with varying degrees of accuracy. Ideally, soil moisture sensors operate in the passive L band (1.4 GHz), but such instruments are still in the development phase (Kerr et al. 2001; Entekhabi et al. 2002). Current spaceborne sensors suitable for soil moisture monitoring include the Advanced Microwave

Corresponding author address: Rolf Reichle, NASA Goddard Space Flight Center, Code 900.3, Greenbelt Road, Greenbelt, MD 20771.
E-mail: reichle@janus.gsfc.nasa.gov

Scanning Radiometers (AMSR) on board the *Aqua* satellite and the *Advanced Earth Observing Satellite* (*ADEOS-II*), respectively. AMSR measures passive microwaves in six channels, with a minimum frequency in C band (6.925 GHz). From October 1978 to August 1987, the Scanning Multichannel Microwave Radiometer (SMMR), a predecessor to AMSR, collected C-band (6.63 GHz) passive microwave data.

Owe et al. (2001) recently developed a novel retrieval algorithm for soil moisture from passive microwave measurements and produced a 9-yr global soil moisture dataset from SMMR observations (de Jeu 2003). A global, multiannual soil water index has also been retrieved from spaceborne radar observations (Wagner et al. 2003). While such active (radar) microwave measurements offer finer spatial resolution, soil moisture retrieval is difficult and prone to errors because of uncertainties in the surface roughness, vegetation, and heterogeneous land cover. Nevertheless, Wagner et al. (2003) find that their soil water index agrees reasonably well with modeled soil moisture in tropical and temperate climates.

Our ultimate goal is to assimilate satellite soil moisture data into a global land model. Here, we take a step in this direction by examining the compatibility of the de Jeu (2003) SMMR soil moisture retrievals with global model soil moisture and ground data where available. Because of the obvious link between precipitation, radiation, and soil moisture, output from a land surface model that is forced with observed meteorological data contains much information about soil moisture. In this study, the land model is forced with a high-quality global dataset of surface meteorological conditions that is based on observations as much as possible. This produces a soil moisture dataset similar to the many model-based soil moisture datasets that can be found in the literature (e.g., Mintz and Serafini 1992; Nijssen et al. 2001). Also, ground-based soil moisture data for the SMMR time period are available for select locations in Eurasia and North America from the Global Soil Moisture Data Bank (Robock et al. 2000).

The satellite, ground-based, and model soil moisture used here are independent data, each with its own set of limitations. It is well known that state-of-the-art land surface models produce widely different soil moisture output even when integrated with identical meteorological forcing inputs (Henderson-Sellers et al. 1995; Koster and Milly 1997; Entin et al. 1999). Errors in C-band surface soil moisture retrievals are generally high, and modest amounts of vegetation obscure the soil moisture signal. Ground-based measurements are sparse and not necessarily representative of large-scale soil moisture. At this time, errors in global soil moisture observation and modeling are so large that there is no universally agreed climatology. Because of these shortcomings, it is not clear a priori that straightforward assimilation of C-band soil moisture retrievals or ground measurements into a land surface model offers the expected benefits. In fact, even for the assimilation of ground data into a highly calibrated model at a single field site, Calvet and Noilhan (2000) find it necessary to rescale model output to observed soil moisture. In this paper, we demonstrate where global soil moisture data from the different sources agree and by how much they can differ. The differences, in particular the discrepancies in the soil moisture climatology, have important implications for soil moisture assimilation.

2. Data

In this study we compare soil moisture derived from satellite data, land model integrations, and ground-based measurements. The three data sources are, of course, fundamentally different. Satellite data infer soil moisture from its impact on the C-band passive microwave signal, whereas the model integrations relate soil moisture to antecedent meteorological forcing. The ground measurements are perhaps most accurate but are far more sparsely distributed in space and in time and not necessarily representative of large-scale soil moisture.

The SMMR satellite retrievals of soil moisture are from de Jeu (2003) and Owe et al. (2001). Their novel retrieval algorithm is based on the polarization difference of the C-band passive microwave signal and simultaneously retrieves surface soil moisture and the vegetation optical depth. The dual polarization approach overcomes the need to specify the vegetation parameter that is required in single-channel algorithms (Jackson and Schmugge 1991) and is difficult to obtain on a global scale. Surface temperature inputs to the Owe et al. (2001) algorithm are estimated from the SMMR 37-GHz channel. Despite global coverage of the satellite, soil moisture retrievals are not available everywhere. Soil moisture retrieval is impossible in areas that contain a significant fraction of surface water (such as along the coast) or when the soil is frozen. Moreover, soil moisture retrieval from C-band passive microwaves is restricted to areas with sufficiently light vegetation cover.

Note that Owe et al. (2001) intentionally did not calibrate their retrieval algorithm. Because of the paucity of calibration and validation data, any calibration to a regional dataset would in effect invalidate the global applicability of the algorithm. In a quality control step, we excluded SMMR soil moisture retrievals associated with a vegetation optical depth greater than 0.6 (simultaneously retrieved) from the analysis. Both SMMR overpasses at local noon and local midnight are used.

Figure 1 shows the monthly average number of SMMR soil moisture retrievals that were used in this study. The time average is computed from January 1979 to August 1987. SMMR was flown on a polar-orbiting satellite that passed over a given location in the midlatitudes about once every 3–4 days (or about 8–10 times per month). Also shown in Fig. 1 is the 1983–90 average leaf area index (LAI) derived from data collected by Advanced Very High Resolution Radiometers

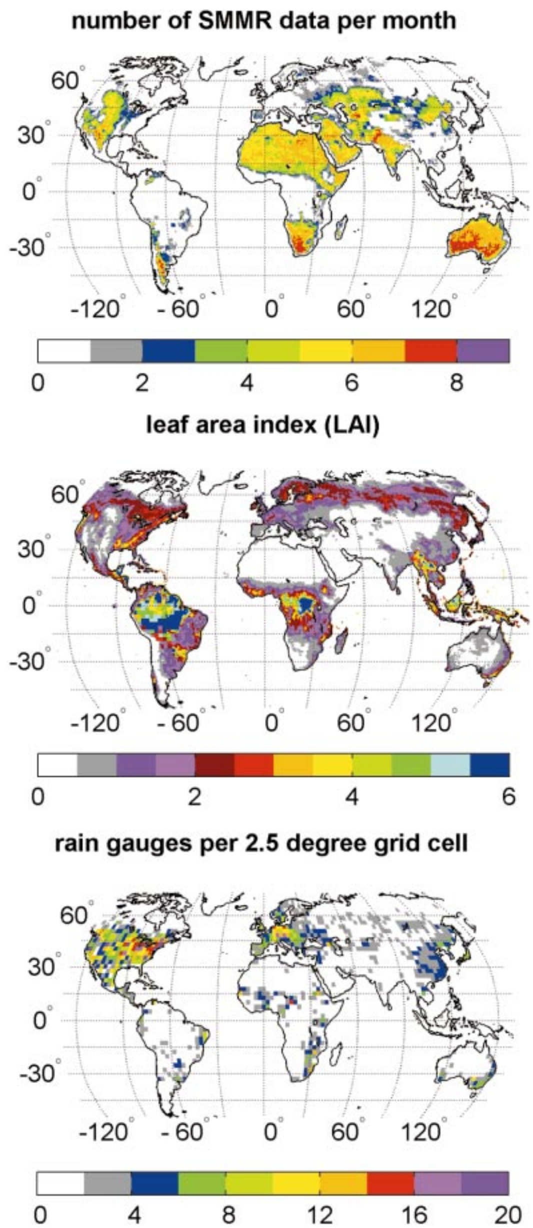

FIG. 1. (top) Jan 1979–Aug 1987 average monthly number of soil moisture retrievals from SMMR after quality control and mapping to catchment space, (middle) 1983–90 average LAI, and (bottom) Jan 1979–Aug 1987 average monthly number of rain gauges per 2.5° grid cell.

(AVHRRs) (Guillevic et al. 2002). Most soil moisture data are available in low-latitude regions with little vegetation, namely, northern and southern Africa and Australia. Data are also available at midlatitudes where vegetation is sparse (U.S. Great Plains, central Eurasia), but here freezing of the soil limits the number of data available in winter, resulting in a lower year-round average. Data are not available in densely forested regions such as the tropical rainforests of South America, Africa, and east Asia or the temperate and boreal forests of North America and Eurasia.

Model soil moisture is obtained from integrations of the National Aeronautics and Space Administration (NASA) Catchment Land Surface Model (hereinafter ''Catchment model'' or CLSM; Koster et al. 2000a; Ducharne et al. 2000). The Catchment model's basic computational unit is the hydrological catchment (or watershed). In each catchment, the vertical profile of soil moisture is determined by the equilibrium soil moisture profile from the surface to the water table and by two additional variables that describe deviations from the equilibrium profile in a 1-m root zone layer and in a 2-cm surface layer, respectively. Unlike traditional, layer-based models, the Catchment model includes horizontal redistribution of soil water within each hydrological catchment based on the statistics of the catchment topography.

The salient feature of the land model integration is that it uses meteorological forcing inputs that rely on observed data as much as possible. The forcing data for the land model are based on the European Centre for Medium-Range Weather Forecasts (ECMWF) 15-yr reanalysis (ERA-15) available from 1979 to 1993. Important corrections using monthly mean observations were applied to the ERA-15 precipitation, radiation, temperature, and humidity data (Berg et al. 2003b). Together, the observation-based corrections ensure that the forcing data and hence the soil moisture output are as close to reality as is possible. Precipitation—arguably the most critical input for accurate soil moisture modeling—has been corrected primarily with a merged product of satellite and gauge data from the Global Precipitation Climatology Project (GPCP, version 2) (Huffman et al. 1997). Figure 1 shows the time-average number of rain gauges per 2.5° grid cell contributing to the merged satellite-gauge product. Gauge density is greatest over North America and Europe, suggesting that soil moisture from the land model is likely most accurate there. In other regions, the precipitation data rely mainly on satellite estimates and are presumably less accurate. Note also that radiation corrections were only available from 1983 on. The model spinup initial condition was derived by repeatedly integrating the model for 10 yr with 1979 forcing.

To examine the model dependence of our results, we repeated all analyses with a traditional, layer-based land surface model, the Mosaic model (Koster and Suarez 1992), using the same observation-based meteorological forcing data. There are substantial differences between soil moisture from the two models, as could be expected from earlier studies in the framework of the Global Soil Wetness Project (GSWP) (e.g., Entin et al. 1999), and the Project for Intercomparison of Land Surface Parameterization Schemes (PILPS) (Henderson-Sellers et al. 1995). In the context of the present study, we find that the use of either model leads to the same general conclusions.

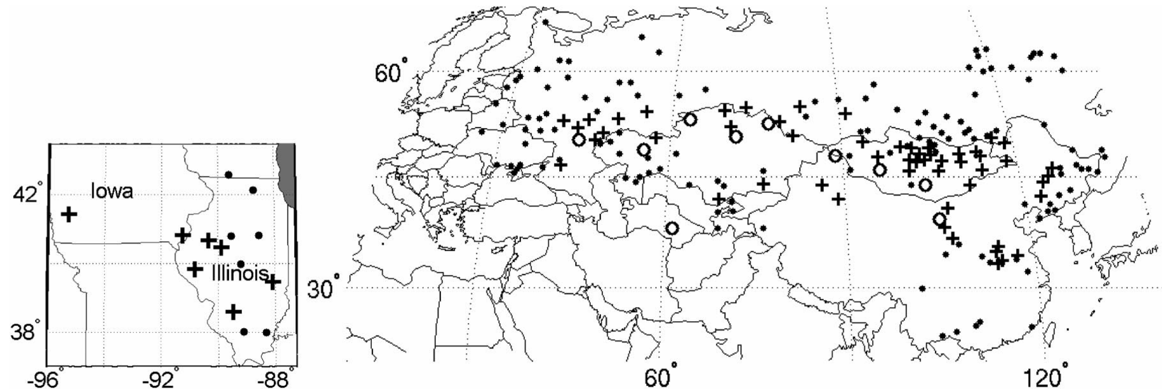

FIG. 2. Catchments that have ground measurements: (plus) included in statistical analysis, (circle) included in statistical analysis, except anomaly correlation analysis because the std dev of monthly mean soil moisture is smaller than the approximate noise level in the monthly mean ground data (0.03 $m^3 m^{-3}$), and (dot) insufficient data for statistical analysis.

Ground-based data are available from the Global Soil Moisture Data Bank (GSMDB) (Robock et al. 2000). Data for all or part of the SMMR years are available for the former Soviet Union (130 stations, 1978–85), Mongolia (42 stations, 1964–93), China (43 stations, 1981–91), Iowa (2 stations, 1972–94), and Illinois (19 stations, 1981–96) (Fig. 2). No data are available for the Southern Hemisphere. All measurements have been taken using the gravimetric method except for the Illinois data, for which neutron probes have been used (Hollinger and Isard 1994).

The three data sources—satellite, model, and ground-based soil moisture—are independent. Unfortunately, they also describe different aspects of soil moisture. Most important, there are fundamental differences in the horizontal and vertical scales among the three data sources. For our analysis, we have mapped all data to catchment space. In this space the global land surface is divided into 59 124 catchments or hydrological units (excluding inland water and ice-covered areas). The linear scale of individual catchments ranges from 2 to 250 km with a mean (median) of 47 km (34 km). Model soil moisture is computed directly in catchment space, whereas the underlying horizontal resolution of the SMMR soil moisture data is on the order of 140 km. While model and satellite data are inherently distributed (or areal) data, ground observations are essentially point-scale measurements. The catchment average of the ground data is computed as the arithmetic mean of all available station data that lie within the boundaries of the catchment. However, there is only one station per catchment for 80% of the catchments that have any ground data at all.

The three data sources also differ in their vertical dimension. SMMR soil moisture is shallowest, representing on average only the top 1.25 cm of the soil column. Model surface soil moisture covers the top 2 cm of the soil column. The depth associated with ground-based surface soil moisture varies by location from 5 to 10 cm. Moreover, at many sites ground-based soil moisture is reported as plant-available soil moisture. Unfortunately, measurements of the wilting level are not always available (former Soviet Union data) or are questionable (Mongolian data). It is important to keep these differences in mind when comparing the different soil moisture products, in particular with respect to the general perception of the ground-based data as the ''truth.''

3. Approach

Our analysis is based on monthly mean time series because of the strong variability and noise that are present in surface soil moisture at shorter time scales, and because of the mismatch between satellite and ground observation times. Precipitation events dominate surface soil moisture and introduce strong variability at very short time scales. The exact timing of precipitation events is often wrong in the atmospheric forcing data that drive the land surface model. Last, individual soil moisture retrievals are subject to large errors related to uncertainties in vegetation and surface temperature estimation.

For the satellite data, monthly mean values were computed if at least three data points were available from SMMR, otherwise satellite data for that month were not used. The monthly mean values from the land surface model were computed from the complete model time series. Additional analysis shows that computing the model monthly mean values by extracting the model data only at the SMMR observation times does not alter our conclusions. Monthly mean values for the ground data were computed whenever data were at all available. While 65% of the monthly means used in the analysis were computed from three or more ground observations, 16% of the monthly means were based on just one ground observation.

In addition to the raw time series of monthly means, we also computed anomaly time series by subtracting the monthly climatology (i.e., the average over all monthly means of a given calendar month). In other words, the raw time series include the seasonal cycle, while the anomaly time series describe only deviations from the average seasonal cycle. The monthly climatology is computed if at least three monthly means are available for a given calendar month. For each catchment's raw and anomaly time series of monthly mean soil moisture, we then computed mean values (zero by definition for anomalies), standard deviations, and time series correlation coefficients between the different data types—provided at least 27 monthly mean values (or monthly mean anomalies) were available (out of a maximum possible number of 104). The cutoff excludes catchments for which only few data are available from the analysis, and our results do not depend on the exact number that is chosen. We also refer to the spatially distributed time series mean and variability as the climatology.

Our objective is to ensure a fair comparison between the different data types (satellite, model, and ground data). Because the observation-corrected meteorological forcing data are only available from 1979, we discard the first 3 months of SMMR data and limit our analysis to the time period from January 1979 to August 1987. As discussed earlier, SMMR monthly means were not always available, and at many stations, ground data are only available from April to October. Therefore, in comparisons between model and satellite data, model data were not used whenever monthly mean satellite data were not available. Likewise, in comparisons with the ground data only months with data from both the satellite retrievals and the ground observations are included in the analysis. This approach uses the maximum possible number of data and at the same time ensures that in all comparisons the time series mean and variability of each data type are based on exactly the same months. At each location, however, the statistics are based on different months of data depending on data availability, and therefore should not be interpreted as an estimate of the annual mean or variability.

In a related paper, Berg et al. (2003a, manuscript submitted to *J. Geophys. Res.*, hereinafter BFR) also examine the consistency between SMMR soil moisture retrievals, ground data from the Global Soil Moisture Data Bank, and model soil moisture that is based on the same observed meteorological forcing inputs. There are many differences in the analysis; however, BFR focus on the validation of the observation-based forcing dataset, not on implications for data assimilation. Moreover, BFR use a different land surface model at much larger spatial scales (2° latitude by 2.5° longitude) but at a much finer temporal resolution of 6 h. The spatial aggregation of BFR or the aggregation to monthly mean values (in the present paper) are important and eliminate some of the noise in the satellite data. Last, BFR construct the anomaly time series by subtracting 3-month climatological mean values (as opposed to a monthly climatology), and the comparison with ground data in BFR is done for root zone soil moisture (as opposed to surface soil moisture). Despite these differences, BFR find correlations between the datasets that are very similar to the correlations presented here. This means that the results presented here are robust with respect to the spatial and temporal scales, the land surface model being used, and other processing details.

4. Results and discussion

The satellite and model data represent two components of a data assimilation system that is designed to produce an optimal merged product. Proper merging, however, requires a full understanding of how the datasets compare, most importantly with respect to potential biases in the mean and variability. In the first part of our study, we focus on analyzing the differences between the satellite and model data. We withhold judgment on which data are more correct, since both contain independent and valuable information from observations (observed precipitation and radiation in the case of the model data). As we show in the second part of this study, a comparison of both datasets with the available ground observations does not indicate that one is superior.

a. Comparison of satellite and model data

1) SOIL MOISTURE CLIMATOLOGY

Figure 3 shows the time-average surface soil moisture from January 1979 to August 1987 for SMMR retrievals and the Catchment Land Surface Model. For ease of comparison, model soil moisture is not plotted where satellite data are not available. In both cases, global wetness and dryness patterns agree very much with expectations—the driest places are in the Sahara Desert, the Arabian Peninsula, and central Australia, whereas the wettest regions are typically at higher latitudes (the two maps correlate with $R^2 = 0.57$). The model, however, does not agree well with the satellite data on the absolute level of surface soil moisture. Mean soil moisture values differ by several (volumetric) percent, with the maximum differences spanning more than half of the dynamic range of soil moisture. The global and time-average Catchment model soil moisture is drier than that of SMMR (by 0.069 $m^3\ m^{-3}$). Here and in the remainder of the text, all ''global'' averages are area-weighted averages across all catchments for which we have both satellite and model data.

Regional biases are also strong, with a spatial standard deviation of 0.053 $m^3\ m^{-3}$ for the difference between the time-average fields. An important difference between satellite and model soil moisture can be seen in North America, where the model data show a strong west-to-east gradient. SMMR data also show a gradient, but it is much weaker and shifted to the east. This implies that the model is much drier than SMMR in the western United States, but much wetter in the central

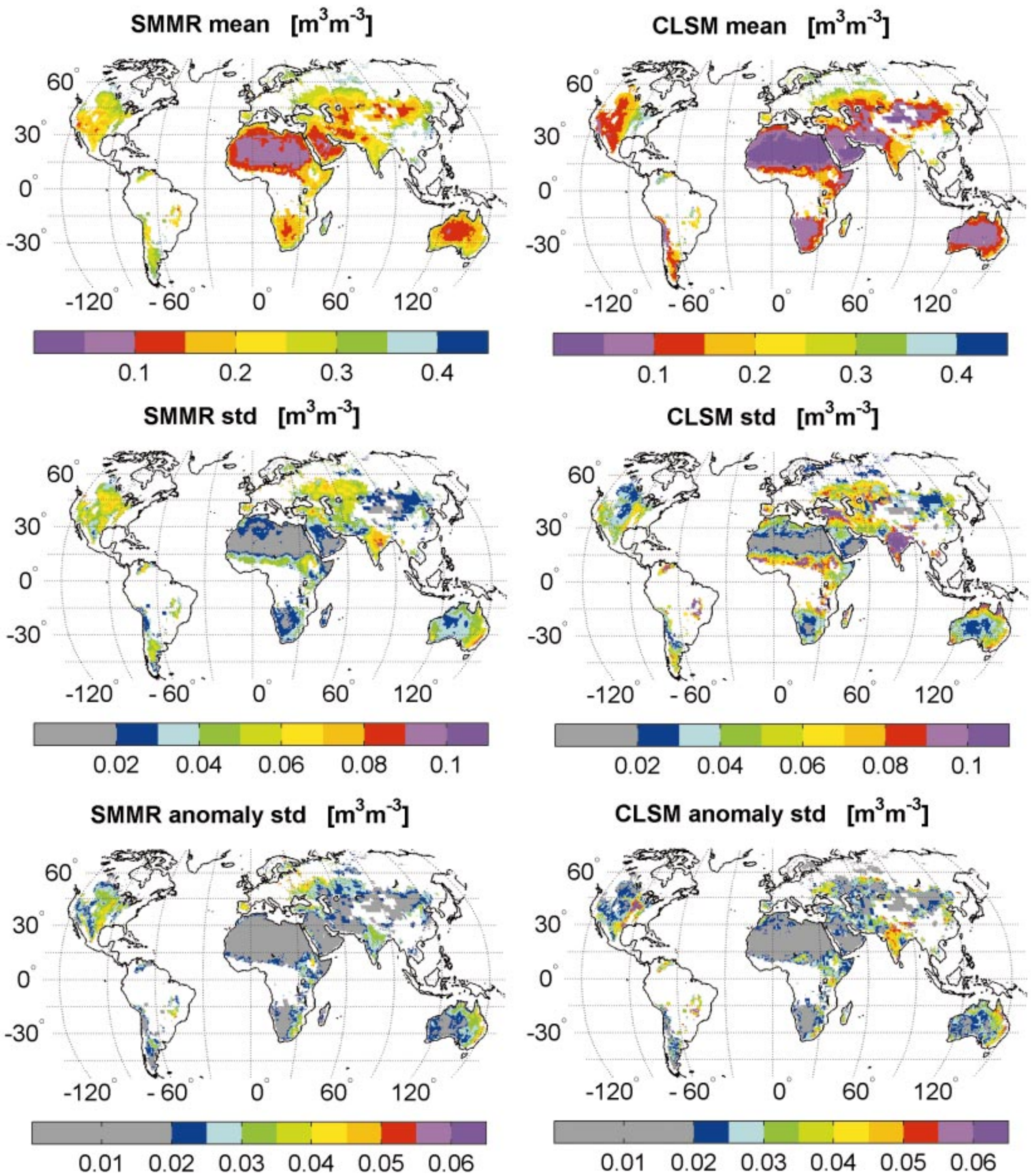

FIG. 3. Climatology of surface soil moisture (Jan 1979–Aug 1987) from (left) SMMR and (right) the CLSM: (top) mean, (middle) std dev, and (bottom) anomaly std dev.

United States. Similar regional biases exist in eastern Africa and central Eurasia.

One reason for the observed biases may be the (intentional) lack of calibration of the soil moisture retrieval algorithm (section 2). The global pattern of bias between SMMR and the model appears to be related to LAI (Fig. 1). Where LAI is less (greater) than 1, SMMR soil moisture is biased wet (dry) against the model. While it can be expected that SMMR soil moisture is less accurate for larger LAI, it is counterintuitive that SMMR retrievals should be biased dry for highly vegetated areas. Rather, we would have expected that SMMR picks up too much signal from the vegetation water and would therefore overestimate soil moisture over dense vegetation. Nevertheless, the close relationship between the bias in soil moisture and vegetation parameters suggests that at least some of the bias can be attributed to the SMMR retrieval algorithm. On the other hand, it must be acknowledged that the bias could just as well be due to problems in the model formulation, model parameters (in particular LAI), or forcing inputs. Given the paucity of validation data, it is not possible to say with confidence whether the retrieval algorithm or the model is to blame for the bias, or, in other words,

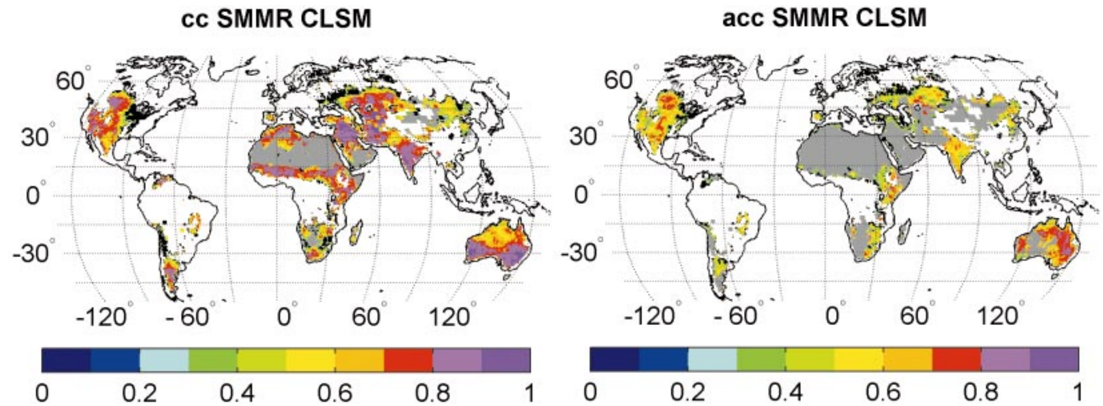

FIG. 4. Time series correlation between surface soil moisture from SMMR and the Catchment model (Jan 1979–Aug 1987): (left) correlation coefficient (cc) and (right) anomaly correlation coefficient (acc). Correlation coefficients that are not different from zero at 5% significance are plotted black. Catchments with std dev less than the approximate noise level in the SMMR monthly mean data (0.02 $m^3 m^{-3}$) are shown gray.

whether SMMR or model soil moisture is closer to the true climatology.

The time series standard deviation of the monthly means is largely a measure of the strength of the seasonal cycle of soil moisture (at least where the seasonal cycle is stronger than any anomalies). Figure 3 also shows the time series standard deviation (January 1979 to August 1987) for SMMR and Catchment model soil moisture. Again, the patterns of relatively strong and weak seasonal cycles largely agree between SMMR and the model, with strong seasonal cycles in India, the central United States, and central Eurasia (the two maps correlate with $R^2 = 0.30$). However, the strength of the average seasonal cycle in SMMR soil moisture is weaker than in the Catchment model, by 0.01 $m^3 m^{-3}$ on average across the globe. The differences are particularly pronounced in North America, India, the Sahel, central Eurasia, and northern Australia.

When the seasonal cycle is removed from the time series, the standard deviation becomes a measure of the strength of soil moisture anomalies, which are critical for seasonal climate forecasting. Figure 3 also compares the anomaly standard deviations of SMMR and Catchment model soil moisture. Again, SMMR and the model agree that the strongest anomalies are in the central United States, central Eurasia, India, and eastern Australia (the two maps correlate with $R^2 = 0.35$). When compared to the Catchment model, however, absolute SMMR soil moisture anomalies are much weaker in India, generally stronger in central Eurasia, and both stronger and weaker in different parts of North America.

2) TIME SERIES CORRELATIONS

The model does not agree well with the SMMR retrievals in terms of absolute magnitudes. Nevertheless, soil moisture data from SMMR and the model do largely agree in terms of their global patterns. As will be shown next, we also find some agreement in the temporal signals, as indicated by an examination of time series correlations. By definition, the correlation coefficient measures agreement in the temporal variation of two time series, regardless of their individual mean values and standard deviations. Given certain assumptions, the square of the correlation coefficient (typically known as R^2) can be interpreted as the variance of one time series explained by another time series.

Figure 4 shows global maps of correlation coefficients between SMMR and Catchment model soil moisture. Whenever the computed correlation coefficient is not statistically different from zero at 5% significance, it is plotted black—confidence intervals were computed as a function of the number of data and vary by location, with higher confidence associated with larger data volumes. In the gray areas of Fig. 4, correlations are meaningless because the variability in the time series and anomaly time series is smaller than the noise (i.e., the measurement error). Since there is considerable noise in the instantaneous SMMR soil moisture retrievals (between 0.05 and 0.09 $m^3 m^{-3}$), and since the temporal sampling rate is low, SMMR monthly means will also be noisy. Typically, there are about nine SMMR retrievals per month available for months having data. (Note that Fig. 1 shows the average number of data per month over the entire SMMR history, including months in which the soil was frozen or vegetation grew too dense and monthly mean data were not computed). From this information, we (optimistically) estimate the approximate noise level in the monthly mean soil moisture from SMMR to be 0.02 $m^3 m^{-3}$.

Figure 4 shows that correlation coefficients are strongest in North America, Patagonia, the Sahel, central Asia, India, and Australia. While correlations are strong in the U.S. Great Plains, there appears to be no agreement between SMMR and Catchment model soil moisture to the east of the Great Plains. The global average time series correlation coefficient between SMMR and the Catchment model soil moisture is 0.63, indicating that the seasonal cycle in SMMR agrees reasonably well with model results, with typically about 40% of the variance explained where data are available.

When the seasonal cycle is removed, the standard deviation of the anomaly time series is smaller than the noise level of SMMR monthly mean data (0.02 $m^3 m^{-3}$) in much larger areas, further shrinking the regions where we can meaningfully conduct the correlation analysis. Figure 4 shows that the strongest anomaly correlations are again found in North America, India, Australia, and central Eurasia, with some positive correlations also present in southern Africa and Patagonia. The globally averaged anomaly correlation coefficient between SMMR and the Catchment model is 0.50.

Reasonable agreement between satellite and model soil moisture can only be expected if both datasets are at least somewhat skillful. The dominant errors in the satellite and model soil moisture are related to vegetation and precipitation, respectively. Soil moisture retrievals from satellite tend to be less accurate with increasing LAI because vegetation obscures the moisture signal from the soil. Model soil moisture, on the other hand, can only be as good as the input precipitation data. Oki et al. (1999) show that at least 30 rain gauges per 10^6 km^2 are necessary for accurate streamflow simulation. As a first approximation, we can assume that a comparable density of about two gauges per 2.5° GPCP grid cell is required for soil moisture modeling.

By comparing LAI and rain gauge density (Fig. 1) with the time series correlations between SMMR and model soil moisture (Fig. 4), the global patterns of strong and weak correlations and anomaly correlations can partly be explained. Of all the regions for which we have sufficient SMMR data, the U.S. Great Plains are the only large region that has both a high rain gauge density and low vegetation cover. It also shows the highest agreement between SMMR and model soil moisture. While the rain gauge density increases to the east of the Great Plains, so does the vegetation cover, and correlations between SMMR and the model are no longer significant. In fact, there appears to be no agreement between SMMR and model soil moisture wherever the time average LAI is greater than 1. By contrast, low gauge density does not always preclude reasonable agreement between the satellite and the model. There are large regions (including Australia, India, and central Eurasia) with fewer than two rain gauges per 2.5° grid cell that show reasonable agreement between SMMR and model soil moisture, despite the fact that in these regions the precipitation data are primarily based on satellite-derived estimates.

It is intriguing that some of the areas of reasonable agreement between the SMMR and Catchment model anomaly time series coincide with areas that are important for soil moisture memory and short-term climate predictability. Figure 5 shows regions where land initialization is likely to have a significant impact on precipitation forecasts for Northern Hemisphere summer

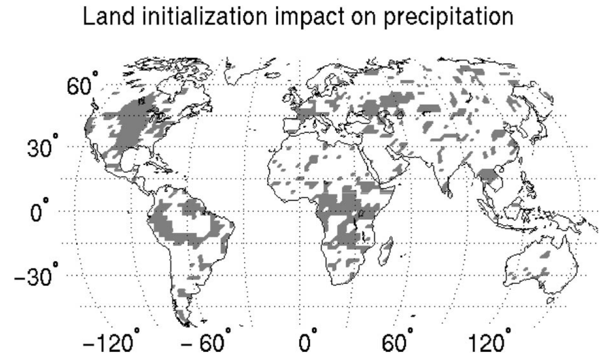

FIG. 5. Regions where land initialization had an impact on boreal summer (JJA) precipitation at least once during the 5 yr studied at the 1% significance level [reproduced from Koster and Suarez (2003), upper-right panel of their Fig. 4].

[June–July–August (JJA)]. For ease of comparison, we reproduced Fig. 5 with original data from Koster and Suarez (2003) (upper-right panel in their Fig. 4) in the map projection used here. In the midlatitudes, regions of greatest interest for land initialization of summer precipitation forecasts include the U.S. Great Plains and parts of Eurasia just north of the Black and Caspian Seas. Moreover, additional predictability experiments for boreal winter suggest that soil moisture initialization may also have an impact in eastern Australia [see lower-right panel of Fig. 9 in Koster et al. (2000b)]. The tropical areas in South America, Africa, and south Asia shaded in Fig. 5 are not of interest here because soil moisture cannot be measured by satellite in such highly vegetated areas.

Here, we focus on the U.S. Great Plains and parts of Eurasia just north of the Black and Caspian Seas. Koster and Suarez (2003) show that these regions have (i) a strong potential for large soil moisture anomalies (as confirmed in Fig. 3), (ii) sensitivity of evaporation to soil moisture (so that a soil moisture anomaly can induce an evaporation anomaly), and (iii) sensitivity of precipitation to evaporation (so that an evaporation anomaly can induce a precipitation anomaly). Together, these three factors represent a mechanism of possible land–atmosphere feedback and are typically found in transition regions between wet and dry climates.

Likewise, the present study naturally focuses on the transition regions between dry and wet climates. On the one hand, primarily wet regions are typically covered by dense vegetation, and soil moisture cannot be observed by SMMR. On the other hand, very dry (and very wet) regions usually show very little variability in monthly means of surface soil moisture, which disqualifies these regions from our correlation analysis because of the high level of noise in the SMMR retrievals. It is thus fortunate that areas where soil moisture retrievals from the C-band satellite instrument appear to be most consistent with model data (in particular the U.S. Great Plains) coincide with regions of most interest for seasonal precipitation forecasts.

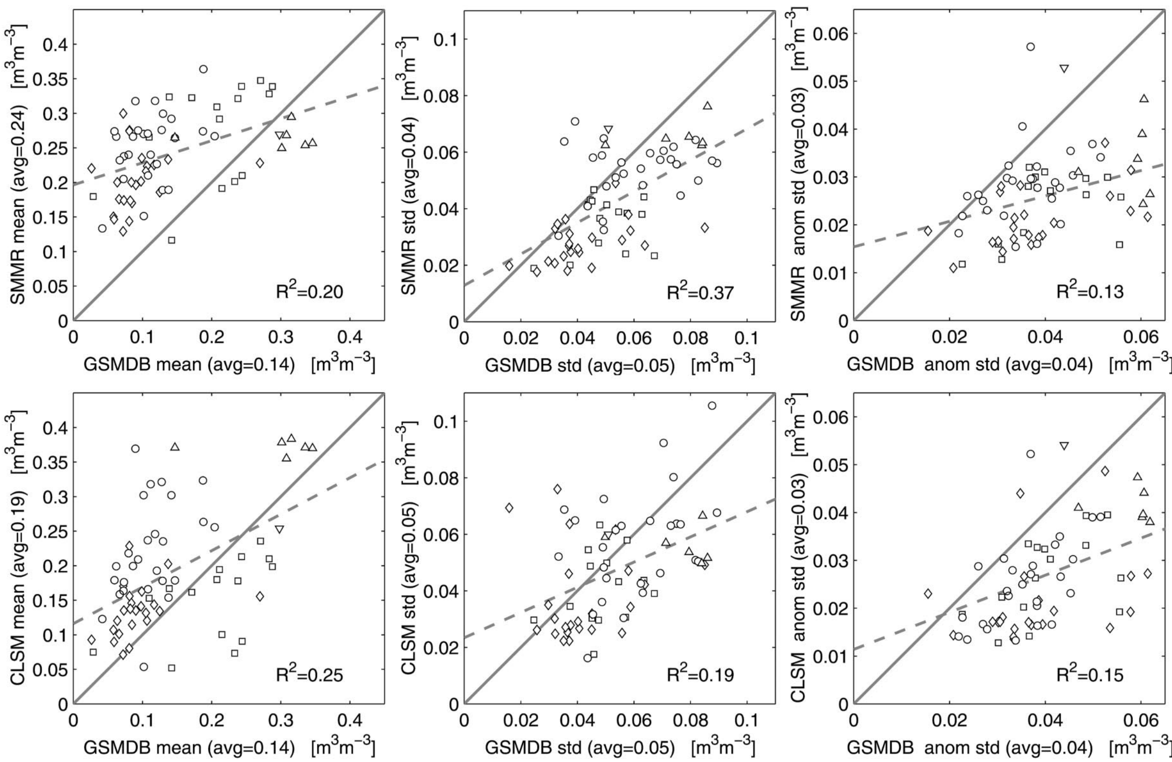

FIG. 6. Climatology of surface soil moisture from (top) SMMR and (bottom) CLSM against the climatology of the ground observations (GSMDB): (left) mean, (middle) std dev, (right) anomaly std dev. The solid line is the one-to-one line. The dashed line is a linear regression, with R^2 as indicated in the lower-right-hand corner of each plot. Symbols indicate GSMDB dataset: (square) China, (diamond) Mongolia, (circle) former Soviet Union, (triangle pointing up) Illinois, and (triangle pointing down) Iowa. Period is Jan 1979–Aug 1987.

b. Comparison with ground data

So far we have only compared SMMR and Catchment model soil moisture without knowing which (if either) is true. In an attempt to determine which is correct, we now compare these datasets with ground-based observations from the Global Soil Moisture Data Bank (Robock et al. 2000). After mapping to catchment space, ground data are available for 207 catchments (Fig. 2). Of these, only 71 catchments have enough ground and SMMR data during the analysis period to allow for a statistically useful analysis. Furthermore, of the 71 catchments, 10 (indicated by circles in Fig. 2) are excluded from the anomaly correlation analysis because the variability of the anomalies in the ground measurements was smaller than the noise level of the ground data. The noise in the catchment-average monthly mean soil moisture was approximated as 0.03 $m^3 m^{-3}$, striking a compromise between the generally higher accuracy of individual ground measurements and their poorer spatial and temporal representativeness compared to SMMR soil moisture.

Figure 6 shows the time-average soil moisture of SMMR and the Catchment model versus the ground observations (January 1979–August 1987). It is immediately obvious that neither SMMR nor the Catchment model agree very well with the ground data. Both SMMR and Catchment model soil moisture exhibit a large wet bias relative to the ground observations and are scattered widely. The bias between SMMR and the ground data (0.10 $m^3 m^{-3}$) is even larger than between the Catchment model and ground data (0.05 $m^3 m^{-3}$), and there is more scatter in SMMR soil moisture than in the Catchment model when plotted against ground data. The wet bias is remarkable because the surface layer associated with the SMMR and model data is shallower than the measurement depth of the ground data (section 2), and a dry bias would be more in line with expectations.

Obviously, regional differences in time-average soil moisture are also strong. Mongolian ground data, for example, are very poorly matched by SMMR retrievals. The Catchment model appears too dry in China and too wet in Mongolia and the former Soviet Union when compared to the ground data. Such regional biases could in part be due to uncertainties in the wilting point, because the Eurasian ground data are reported as plant-available soil moisture, and simultaneous wilting-level measurements are not even available for the former Soviet Union data. In any case, there is scant agreement in the absolute level of soil moisture between datasets.

Figure 6 also shows the variability of the SMMR and Catchment model soil moisture versus the variability of

TABLE 1. Spatially averaged time series cc and acc between surface soil moisture from SMMR, GSMDB, and CLSM. Correlation coefficients (cc) are in regular font below the diagonal; anomaly correlation coefficients (acc) are in italics above the diagonal. Values are averages over the catchments for which ground data are available, except values in parentheses, which are global averages.

cc\acc	SMMR	CLSM	GSMDB
SMMR	—	*0.45* (*0.50*)	*0.36*
CLSM	0.43 (0.63)	—	*0.38*
GSMDB	0.45	0.45	—

the ground data. Again, there is precious little agreement between SMMR or the Catchment model on the one hand and the ground observations on the other hand. The standard deviation of the SMMR time series (or anomaly time series) is smaller than that of the ground data by 0.01 $m^3 m^{-3}$ on average. When compared to the ground data, the Catchment model exhibits similar average variability in the time series, but less variability in the anomaly time series (by 0.01 $m^3 m^{-3}$). Some of the differences in variability may be related to the different depths that are associated with each data type (Section 2). However, despite the fact that the measurement depth of the ground data varies regionally, we do not find any obvious regional dependence when comparing the time series variability of the ground data to the satellite and model soil moisture. In summary, the scatter is considerable in all cases, and our comparison of the ground data, SMMR retrievals, and the model data demonstrates that each dataset has its own unique climatology.

Last, we analyze the time series correlation of SMMR and Catchment model soil moisture with the ground data. Spatial averages of the time series correlation coefficients are summarized in Table 1. The average correlation coefficient between SMMR (the Catchment model) and the ground data is 0.45 (0.45). This compares to an average correlation coefficient of 0.43 between SMMR and the Catchment model for the catchments of Fig. 2. For the anomalies, the average correlation coefficient between SMMR (the Catchment model) and the ground data is weaker at 0.36 (0.38), compared to an average correlation coefficient of 0.45 between SMMR and the Catchment model. Although these correlation coefficients are somewhat weaker than the global average correlation coefficients between SMMR and the Catchment model, they still indicate that there is reasonable agreement between the three datasets.

Last, Fig. 7 compares for each catchment the time series correlation coefficients between SMMR and the ground data to the correlation coefficients between the Catchment model and the ground data. Note that the shape of the symbols reveals the region and that the symbols are shaded according to the corresponding time series correlation coefficients between SMMR and the Catchment model. Also shown in Fig. 7 is the average 95% confidence interval for all data points. This confidence interval depends on the correlation coefficient itself and on the number of available monthly mean data. Even though the confidence interval is quite large, most of the data points exceed zero at a statistically significant level. In other words, there is some agreement between

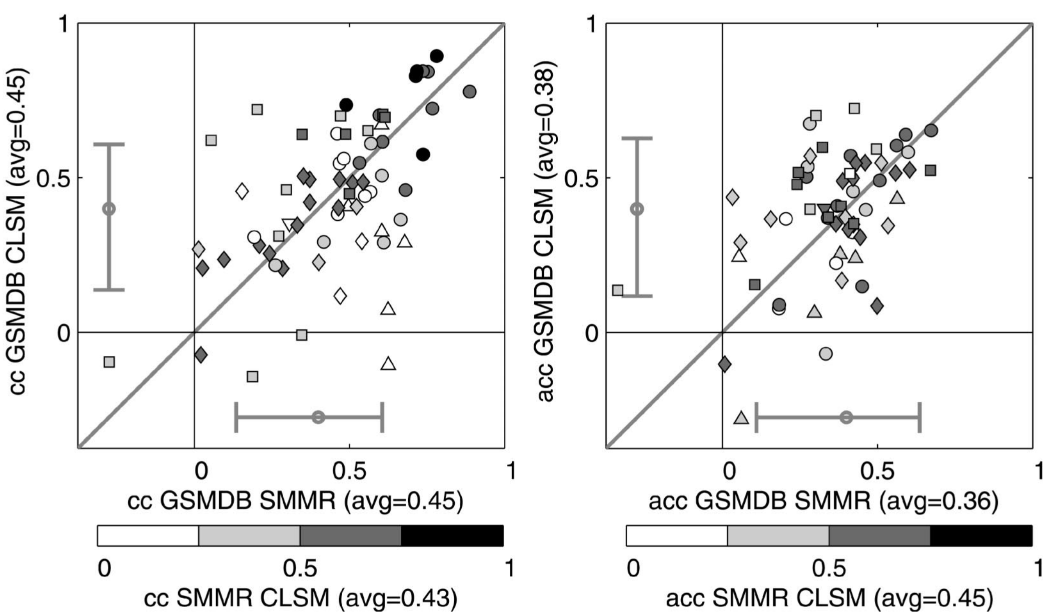

FIG. 7. Time series correlation of surface soil moisture: (left) cc between the Catchment model and ground observations vs cc between SMMR retrievals and ground observations, and (right) same, but for acc. Also shown are average 95% confidence intervals. Shading of symbols is according to the time series correlation between SMMR and the Catchment model. The shape of symbols is the same as in Fig. 6.

SMMR and the ground data and between the Catchment model and the ground data.

There are notable differences in the correlation coefficients between the various regions (Fig. 7). The strongest correlations between all three datasets are typically found in the former Soviet Union, which includes the transition region between wet and dry climates in central Eurasia that are of interest for seasonal climate prediction. Furthermore, SMMR and the ground data correlate very well in Illinois and Iowa, where the Catchment model correlates reasonably with the ground data and poorly with SMMR. Note that in terms of anomaly correlations, all datasets correlate comparably in Illinois and Iowa. Finally note that for Mongolia, SMMR and the Catchment model correlate better than either dataset correlates with ground data.

An important result from Fig. 7 is that the correlation coefficients are scattered evenly around the one-to-one line. This means that SMMR and the Catchment model agree equally well (or equally poorly) with the ground data. Note also that the correlation coefficients between SMMR (or model) soil moisture and ground data are comparable to the correlations between SMMR and the model data (Table 1). Taken together, these results suggest that no dataset agrees with any other dataset particularly well and that from the available data we cannot determine a single ''true'' soil moisture climatology. Given the large and approximately equal errors in all datasets, it is also impossible to say which dataset could offer a ''true'' anomaly time series.

5. Discussion and summary

We have compared global soil moisture data from the SMMR satellite instrument, from model integrations of observed antecedent meteorological forcing, and from available ground observations. Our analysis is based on monthly mean surface soil moisture time series from January 1979 to August 1987. The three types of soil moisture datasets all rely on independent observations. We find that the datasets largely agree in the global patterns of wet and dry regions. The time-average fields from the satellite and the model correlate well spatially. Nevertheless, there are strong differences in the statistics of the datasets. In many regions, both the time-average soil moisture and the temporal standard deviation of soil moisture differ by several volumetric percent between the datasets. We obtain similar differences when either satellite or model data are compared to ground observations.

The analysis of available in situ soil moisture data does not allow us to determine whether SMMR or model data are closer to the truth and shows that transferring soil moisture data from satellite to models and between models is fraught with risk. A given absolute value of soil moisture in satellite data might represent dry conditions, while in one model the same absolute soil moisture might indicate wet conditions, and in another model it might represent just average soil moisture. The lack of a single, agreed climatology poses a severe problem for soil moisture data assimilation. In particular, when used for seasonal forecast initialization, data assimilation requires some method of bias correction or soil moisture scaling at various steps in the process. In essence, the soil moisture climatologies from the satellite data, the uncoupled land model (used in data assimilation for initialization), and the coupled atmosphere–land model (used for prediction) must all be sensibly translated into each other. A simple yet promising approach is to rescale each soil moisture time series by subtracting the local mean and dividing by the local standard deviation and then use these normalized variables as states in the data assimilation. Preliminary tests with the seasonal forecasting system of the NASA Seasonal-to-Interannual Prediction Project (NSIPP) show an improvement in the skill of precipitation hindcasts when rescaling is applied.

Our analysis of time series correlation between the different datasets suggests some agreement between the various datasets. Average correlation coefficients between SMMR and the model are on the order of 0.6 for the soil moisture time series (including the seasonal cycle) and around 0.5 for the anomaly time series. In many regions, the agreement between satellite, model, and ground data is statistically significant at the 5% level, suggesting that SMMR retrievals and the land surface model integration of antecedent meteorological forcing data are giving consistent information. This serves as a first demonstration, at the global scale, of useful soil moisture information in the SMMR data. Since AMSR uses essentially the same frequency as SMMR, our results bode well for the use of AMSR data in the years to come.

Our analysis of SMMR data is most relevant in the transition regions between wet and dry climates because (i) the presence of dense vegetation in wet regions makes satellite remote sensing of soil moisture impossible, and (ii) in predominantly dry regions the high level of noise in the satellite data drowns out the typically small variability signal. Fortunately, these same transition regions are of most interest to soil moisture initialization in seasonal climate forecasting, as demonstrated in earlier studies (Koster et al. 2000b; Koster and Suarez 2003).

In summary, we obtain three key results: 1) The surface soil moisture climatologies of SMMR retrievals, model integrations of observed antecedent meteorological forcing data, and ground measurements are markedly different. 2) Even so, temporal correlations between satellite and model soil moisture are significant, suggesting that SMMR soil moisture retrievals contain useful information and can be assimilated following rescaling. 3) Temporal correlations between soil moisture data from all sources are strongest in the regions of greatest interest for seasonal forecast initialization.

Our results also show clearly that there is a need for global soil moisture retrievals of higher quality than can be obtained from SMMR, which was not specifically designed for soil moisture monitoring. The current AMSR sensors on the *Aqua* and *ADEOS-II* platforms are promising because of the greater spatial resolution and repeat frequency. However, because AMSR is based on essentially the same frequency as SMMR (C band), the soil moisture signal will still be weaker than that obtained with L-band sensors. Two planned L-band satellites—the Soil Moisture and Ocean Salinity (SMOS) mission (Kerr et al. 2001) and the Hydrosphere State (HYDROS) mission (Entekhabi et al. 2002)—should, if successful, provide important advances in soil moisture monitoring and associated climate and forecasting studies.

For the merging of the disparate datasets into a single unified time series of soil moisture anomaly fields through data assimilation we will rely on the scaling approach suggested above. The results presented here will also aid in determining the model and observation error parameters needed for the data assimilation system. Perhaps the most intriguing verification of soil moisture data, in particular the anomaly time series, would be a measurable improvement in the skill of coupled climate forecasts.

Acknowledgments. This research was sponsored by NASA Grant NRA-00-OES-07 (GWEC) and the NASA Seasonal-to-Interannual Prediction Project. We thank Manfred Owe and Richard de Jeu for the SMMR soil moisture data and the Global Soil Moisture Data Bank for the ground measurements.

REFERENCES

Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser, 2003: The impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes. *J. Geophys. Res.,* **108,** 4490, doi:10.1029/2002JD003334.

Calvet, J.-C., and J. Noilhan, 2000: From near-surface to root-zone soil moisture using year-round data. *J. Hydrometeor.,* **1,** 393–411.

Crow, W. T., and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.,* **26,** 137–149.

de Jeu, R. A. M., 2003: Retrieval of land surface parameters using passive microwave remote sensing. Ph.D. dissertation, Vrije Universiteit Amsterdam, Netherlands, 122 pp.

Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model. 2: Parameter estimation and model demonstration. *J. Geophys. Res.,* **105,** 24 823–24 838.

Entekhabi, D., E. Njoku, and P. Houser, 2002: The Hydrosphere State (HYDROS) mission approach to high resolution and frequent revisit data products for hydrometeorological and hydroclimatological applications. *Eos, Trans. Amer. Geophys. Union,* **83** (Fall Meeting Suppl.), Abstract H72A-0845.

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

Guillevic, P., R. D. Koster, M. J. Suarez, L. Bounoua, G. J. Collatz, S. O. Los, and S. P. P. Mahanama, 2002: Influence of the interannual variability of vegetation on the surface energy balance—A global sensitivity study. *J. Hydrometeor.,* **3,** 617–629.

Henderson-Sellers, A., A. Pitman, P. Love, P. Irannejad, and T. Chen, 1995: The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.,* **76,** 489–504.

Hollinger, S. E., and S. A. Isard, 1994: A soil moisture climatology of Illinois. *J. Climate,* **7,** 822–833.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich, 1998: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resour. Res.,* **34,** 3405–3420.

Huffman, G. J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. *Bull. Amer. Meteor. Soc.,* **78,** 5–20.

Jackson, T. J., and T. J. Schmugge, 1991: Vegetation effects on the microwave emission of soils. *Remote Sens. Environ.,* **36,** 203–212.

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

Koster, R. D., and M. J. Suarez, 1992: Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *J. Geophys. Res.,* **97,** 2697–2715.

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

——, and M. J. Suarez, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeor.,* **4,** 408–423.

——, ——, A. Ducharne, M. Stieglitz, and P. Kumar, 2000a: 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.

——, ——, and M. Heiser, 2000b: Variance and predictability of precipitation at seasonal-to-interannual timescales. *J. Hydrometeor.,* **1,** 26–46.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne, 2002: Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment. *Water Resour. Res.,* **38,** 1299, doi:10.1029/2001WR001114.

Mintz, Y., and Y. V. Serafini, 1992: A global monthly climatology of soil moisture and water balance. *Climate Dyn.,* **8,** 13–27.

Nijssen, B., R. Schnur, and D. P. Lettenmaier, 2001: Global retrospective estimation of soil moisture using the Variable Infiltration Capacity land surface model, 1980–93. *J. Climate,* **14,** 1790–1808.

Oki, T., T. Nishimura, and P. Dirmeyer, 1999: Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP). *J. Meteor. Soc. Japan,* **77,** 235–255.

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

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

——, D. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 103–114.

Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, 2000: The Global Soil Moisture Data Bank. *Bull. Amer. Meteor. Soc.,* **81,** 1281–1299.

Seuffert, G., H. Wilker, P. Viterbo, J.-F. Mahfouf, M. Drusch, and J.-C. Calvet, 2003: Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data. *Geophys. Res. Lett.,* **30,** 1498, doi:10.1029/2003GL017128.

Wagner, W., K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf, 2003: Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *J. Geophys. Res.,* **108,** 4611, doi:10.1029/2003JD003663.

Walker, J. P., and P. R. Houser, 2001: A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *J. Geophys. Res.,* **106,** 11 761–11 774.