JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 111, D20102, doi:10.1029/2006JD007190, 2006

Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model

Wenge Ni-Meister,[1] Paul R. Houser,[2] and Jeffrey P. Walker[3]

Received 11 February 2006; revised 30 May 2006; accepted 5 July 2006; published 21 October 2006.

[1] Climate model prediction skill is currently limited in response to poor land surface soil moisture state initialization. However, initial soil moisture state prediction skill can potentially be enhanced by the assimilation of remotely sensed near-surface soil moisture data in off-line simulation. This study is one of the first to evaluate such potential using actual remote sensing data together with field observations. Here the ensemble Kalman filter (Kalman, 1960) is used to assimilate scanning multifrequency microwave radiometer derived near-surface soil moisture data from 1979 to 1987 into the catchment-based land surface model (CLSM). CLSM is used by the NASA Goddard Modeling and Assimilation Office global climate model. Enhancement to land surface soil moisture initialization skill is evaluated for Eurasia using the ground soil moisture measurements collected in Russia, Mongolia, and China. As initial model and observation error predictions were poor, the assimilation improved both the surface and root zone soil moisture estimates only when the observation error was less than the model error. This emphasizes the need for good quality remotely sensed soil moisture data sets, together with reliable observation and model error assessments, in order to ensure improved soil moisture estimates through data assimilation. When the relative magnitude of predicted observation and model error was matched to the error determined from field observation comparison, improvements in root zone and surface soil moisture estimates were guaranteed given unbiased model and satellite observations.

Citation: Ni-Meister, W., P. R. Houser, and J. P. Walker (2006), Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model, *J. Geophys. Res.*, *111*, D20102, doi:10.1029/2006JD007190.

1. Introduction

[2] Accurate land surface moisture initialization in fully coupled climate system models is critical for seasonal-to-interannual climatological and hydrological prediction [*Koster et al.*, 2004]. However, current seasonal climate prediction models do not simulate seasonal or interannual variations of soil moisture that are in agreement with observations, even when a land surface scheme is coupled with the climate model. To improve this, the land surface states are often initialized using outputs generated from an uncoupled land surface model forced by observations (i.e., observed rather than predicted precipitation and radiation [e.g., *Koster et al.*, 2004]). Nonetheless, even with the uncoupled land model simulation, soil moisture predictions are still often poor due to poor model initialization, forcing errors, simplified model physics, and uncertain model parameters [*Houser et al.*, 2001]. Therefore constraining model soil moisture prediction using remotely sensed soil moisture observation assimilation has been proposed as a way to mitigate these errors and improve subsequent predictions [*Walker and Houser*, 2001].

[3] Satellite remote sensing can provide global near surface soil moisture estimates for use in climate model initialization that cannot be obtained through traditional station observation networks. Global $^1/_4$ degree resolution surface soil moisture content has been derived using C band passive microwave observations from the Nimbus 7 satellite scanning multifrequency microwave radiometer (SMMR) for the 1979 to 1987 period of operation [*Owe et al.*, 2001]. Moreover, C band passive microwave soil moisture data is being derived from the Advanced Microwave Scanning Radiometer for the Earth (AMSR-E) observing system launched in 2001 on the Aqua satellite. While no C band measurements are available between SMMR and AMSR-E, low-latitude soil moisture has been estimated using Tropical Rainfall Measuring Mission (TRMM) X band microwave observations [*Bindlish et al.*, 2003; *Gao et al.*, 2004, 2006] since 1998. Further, L band passive microwave soil mois-

[1]Department of Geography, Hunter College and Program in Earth and Environmental Sciences, City University of New York, New York, New York, USA.
[2]Center for Research on Environment and Water, Calverton, Maryland, USA.
[3]Department of Civil and Environmental Engineering, University of Melbourne, Parkville, Victoria, Australia.

0148-0227/06/2006JD007190$09.00

ture data is expected to be available from 2008 with the launch of the Soil Moisture and Ocean Salinity (SMOS) mission. A limitation of these microwave based soil moisture estimates is their confinement to the top few centimeters of soil and subjection to significant vegetation, soils, and roughness error sources. However, it has been demonstrated that assimilation of surface soil moisture observations into land surface models should help mitigate model and observation errors and provide a more accurate root zone soil moisture estimate than modeling alone [*Walker et al.*, 2003; *Reichle and Koster*, 2005], which will be crucial for accurate climate model initialization and prediction.

[4] Data assimilation is the process of merging observations with a model prediction to provide that model with the best estimate of the current state of the natural environment, and is widely used in atmospheric [e.g., *Daley*, 1991] and oceanic modeling studies [e.g., *Bennett*, 1992]. While its use in land surface and hydrological modeling is relatively new, it has shown promise for improving hydrologic predictions by incorporating soil moisture [e.g., *Houser et al.*, 1998; *Walker et al.*, 2003; *Reichle et al.*, 2002a, 2002b; *Zhang et al.*, 2005], snow [e.g., *Sun et al.*, 2004; *Rodell and Houser*, 2004; *Dunne and Entekhabi*, 2006; J. Dong et al., Scanning multichannel microwave radiometer snow water equivalent assimilation, submitted to *Journal of Geophysical Research*, 2006], surface temperature [e.g., *Radakovish et al.*, 2001]. However, the performance of the data assimilation algorithm is highly dependent on reliable estimates of model and observation error, which are difficult to estimate without an estimate of the truth for comparison. Therefore characterizing the model and satellite observation error becomes an important part of assimilation studies that use real satellite data [e.g., *Ni-Meister et al.*, 2005; *Dong et al.*, 2005].

[5] The NASA Global Modeling and Assimilation Office (GMAO; an amalgamation of the former Data Assimilation Office and the NASA Seasonal-to-Interannual Prediction Project) aims to improve seasonal-to-interannual climate predictions through accurate initialization of a global coupled earth system model. To enhance initialization accuracy, innovative data assimilation algorithms are being developed to merge satellite data and model predictions. To this end, *Walker and Houser* [2001] included an extended Kalman filter surface soil moisture data assimilation strategy in the GMAO's catchment-based land surface model (CLSM [*Koster et al.*, 2000]). The ensemble Kalman filter has also been implemented in the CLSM [*Reichle et al.*, 2002a], being simply an alternative methodology for propagating the state covariance matrix that does not require model linearization. A comparison of these two approaches using synthetic data showed very little difference in the results, hence the more generic ensemble Kalman filter has been adopted in this study.

[6] This paper assimilates SMMR-derived surface soil moisture data into the CLSM for the period of 1979–1987. This is one of the first studies to demonstrate the assimilation of real satellite data at a large continental scale over a long period. This study differs from a similar CLSM-SMMR assimilation study by *Reichle and Koster* [2005] undertaken in tandem with this study in three aspects: (1) the satellite-based soil moisture contents in their study were scaled to match the model's climatology before assimilation, (2) their model error was generated based on calibrated model error parameters [*Reichle et al.*, 2002a] and not on realistic model error values, and (3) their evaluation was conducted by comparing the correlations between the assimilated and observed data and between the modeled and observed data. In contrast, this study uses reliable model and satellite observation error information from a ground observation comparison in an earlier study [*Ni-Meister et al.*, 2005]. Moreover, the ensemble Kalman filter soil moisture improvement performance is evaluated using the Eurasian station soil moisture observation network, being the most extensive soil moisture data set available during the SMMR time period.

2. Kalman Filter Assimilation Scheme

2.1. Kalman Filter

[7] The Kalman filter is a linearized statistical approach that provides a statistically optimal estimate of a linear dynamic system, by integrating observations with model predictions (forecast system states) using a weight matrix (Kalman gain) that is based on the relative magnitudes of their respective error covariances [*Gelb*, 1974]. This provides a framework within which the entire system can be modified, with the error covariance representing the reliability of the observations and model prediction. If the observations are unreliable, the model correction will be very small, and if the observations are reliable, the model correction will be such that the model estimate becomes very close to the observed value. Additionally, the Kalman filter can update not only the observed values, but also other correlated state variables.

[8] Through a series of forecasting and update (analysis) steps, the Kalman filter algorithm tracks the conditional mean of the system states and their error covariance. Updates to these forecast state and covariance values are made periodically when observations become available, with the correction being the weighted difference of the observation and model predicted observation. The traditional update equation of the Kalman filter is given by

$$\mathbf{X}^{\mathbf{a}} = \mathbf{X}^{\mathbf{f}} + \mathbf{K}\left(\mathbf{Z} - \mathbf{H}\mathbf{X}^{\mathbf{f}}\right), \tag{1}$$

where the Kalman gain matrix is

$$\mathbf{K} = \mathbf{P}^{\mathbf{f}}\mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{P}^{\mathbf{f}}\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right)^{-1}, \tag{2}$$

and $\mathbf{X}$ is the system state vector, $\mathbf{H}$ is the measurement matrix which linearly relates the observations to the system state, $\mathbf{Z}$ is the observation vector, $\mathbf{P}$ is the model error covariance, $\mathbf{R}$ is the observation error covariance, superscripts f and a indicate forecast (background) and analysis times, respectively, and the superscript $\mathbf{T}$ stands for matrix transpose. The matrix operation $\mathbf{HK}$ yields the ratio of the model error to the total error (model error plus observation error). In the context of this paper the state vector consists of the surface excess, root zone excess and catchment deficit prognostic states used by the CLSM to diagnose surface, root zone and profile soil moisture. Moreover, the observation is a surface soil moisture content and the measurement matrix relates surface soil moisture to the CLSM prognostic states; see *Walker and Houser* [2001] for more details.

2.2. Ensemble Kalman Filter

[9] To apply the Kalman filter, the equations for evolving the system states must be written in a linear state-space formulation. When these equations are nonlinear, the Kalman filter is called the extended Kalman filter, and is an approximation of the nonlinear system that is based on first-order linearization. *Walker and Houser* [2001] have implemented a one-dimensional version of the extended Kalman filter in the CLSM with the simplifying assumption that errors in different catchments are uncorrelated. The ensemble Kalman filter is an alternative to the extended Kalman filter for nonlinear problems [*Evensen*, 1994; *Houtekamer and Mitchell*, 1998], using a Monte Carlo approach to produce an ensemble of model trajectories which are then used to estimate the model covariances. This approach has been successfully introduced into the GMAO ocean assimilation [*Keppenne*, 2000] and soil moisture [*Reichle et al.*, 2002a] estimation problems.

[10] Using the ensemble Kalman Filter approach the background state covariance matrix may be calculated as [*Evensen*, 1994]

$$\mathbf{P_b} = \frac{\left(\mathbf{X^f} - \overline{\mathbf{X^f}}\right)\left(\mathbf{X^f} - \overline{\mathbf{X^f}}\right)^{\mathbf{T}}}{\mathbf{m} - \mathbf{1}}. \quad (3)$$

This could then be used in equation (2) directly. However, in the ensemble Kalman filter approach, the traditional update (analysis) equation was rewritten as equation (4) so that $\mathbf{P}_b$ and $\mathbf{H}$ are not calculated explicitly [*Evensen*, 1994; *Houtekamer and Mitchell*, 1998; *Keppenne*, 2000]. In this way

$$\mathbf{X^a} = \mathbf{X^f} + \mathbf{B^T b}, \quad (4)$$

where

$$\mathbf{B^T} = \mathbf{P_b H^T},$$

and

$$\mathbf{b} = \left(\mathbf{HP^f H^T} + \mathbf{R}\right)^{-1}\left(\mathbf{y} - \hat{\mathbf{Z}}^{\mathbf{f}}\right),$$

with $\mathbf{y} = \mathbf{Z} + \boldsymbol{\varsigma}$ to account for error in the observation [*Burgers et al.*, 1998], and ς being a normally distributed zero mean perturbation of variance $\mathbf{R}$. The vector $\mathbf{b}$ is calculated for each ensemble member from

$$\left(\mathbf{HP^f H^T} + \mathbf{R}\right)\mathbf{b} = \left(\mathbf{y} - \hat{\mathbf{Z}}^{\mathbf{f}}\right), \quad (5)$$

where $\mathbf{HP^f H^T} = \mathbf{qq^T}/(\mathbf{m} - \mathbf{1})$ is the estimated observation given by $\mathbf{HX}$, and $\hat{\hat{\mathbf{Z}}}$ is the averaged value of the estimated observations. The updates are then made individually to each of the ensemble members. Making a substitution in equation (4) it is also possible to estimate $\mathbf{B}$ from (6) to save computing $\mathbf{HPH^T}$ directly such that

$$\mathbf{B^T} = \frac{\mathbf{X^f} - \overline{\mathbf{X^f}}}{\mathbf{m} - \mathbf{1}}\mathbf{q^T}. \quad (6)$$

See *Reichle et al.* [2002a] for more details on the CLSM ensemble Kalman filter application.

3. Data Sets

[11] Figure 1 shows a map of the Eurasian station measurement network and the corresponding CLSM catchments used in the evaluation. This study used the same ground soil moisture measurements, SMMR derived soil moisture data, and CLSM model inputs as used by *Ni-Meister et al.* [2005] for model and observation error characterization. The following is a summary of the data sets used in this study; a detailed description of these data sets is given by *Ni-Meister et al.* [2005].

3.1. Ground Observations

[12] Historical Eurasian soil moisture observations archived in the Soil Moisture Data Bank (SMDB [*Robock et al.*, 2000]) were used to evaluate the data assimilation results presented in this study. The SMDB covers large areas including 43 Chinese, 36 Mongolian and 130 Russian soil moisture monitoring stations over long time periods, with 1981 to 1991, 1973 to 1997, and 1978 to 1985 periods of record, respectively. Soil moisture profiles were measured biweekly over the top 1 m at 10 cm increments using the standard gravimetric technique. The surface zone soil moisture was defined as the shallowest observation available, which for China is the top 5 cm and for Mongolia and Russia is the top 10 cm. The root zone soil moisture was defined as the top 1m average. The plant available soil moisture observations for Mongolian and Russian sites have been converted to total soil moisture to create a consistent data set for comparison with model and assimilation results [*Ni-Meister et al.*, 2005].

[13] The evaluation of assimilation results from this data set is complicated by differences in surface zone thickness; 5–10 cm for station data, 1 cm for SMMR, and 2 cm for CLSM. Moreover, there are fundamental differences in spatial resolution: station measurements are averages for 0.1 to 20 ha areas, SMMR data are areal average soil moisture estimates over 625 km^2 pixels, and CLSM estimates are averages for up to 10,000 km^2 catchments. Despite these limitations, this is the most comprehensive soil moisture ground truth data set available and some useful conclusions can still be made.

3.2. Satellite Observations and Their Errors

[14] Surface zone (top 1 cm) soil moisture estimates at 25 km spatial resolution have been derived from the Nimbus 7 Scanning Multifrequency Microwave Radiometer (SMMR) observations from 1979 to 1987 [*Owe et al.*, 2001]. The Nimbus 7 had a 2–3 day revisit cycle with noon and midnight overpasses; both day and night data have been used in this analysis when available. Moreover, soil moisture data for pixels with large optical depth or unrealistic "wet" pixels have been removed to eliminate the use of potentially erroneous data due to dense vegetation or divergence of the soil moisture retrieval algorithm.

[15] The quality and error of the SMMR data has been fully analyzed by comparing SMMR-derived surface soil moisture with the SMDB data of *Ni-Meister et al.* [2005] and using station soil moisture observations from Illinois

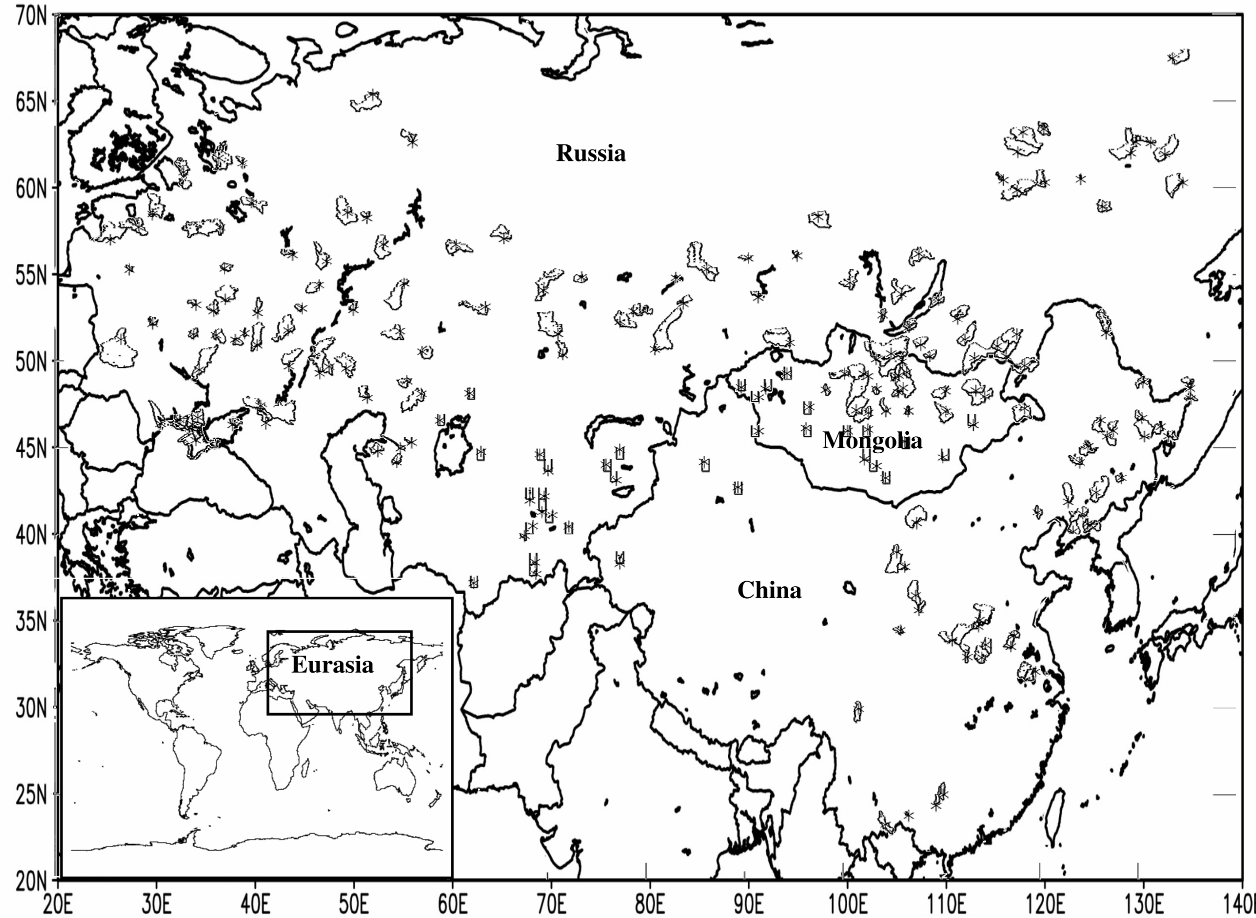

Figure 1. Eurasian catchments and observation stations (shown by stars) used in this study. Only the catchments that contain soil moisture observation station are shown.

[*Owe et al.*, 2001], Russia, Mongolia and Turkmenistan [*de Jeu and Owe*, 2003]. The validation results indicated that soil moisture estimation accuracy (RMSE) was approximately 10% v/v. Figure 2a shows the averaged SMMR observation errors (variance) in summer. These errors have a spatial average of 12% v/v which is consistent with the earlier findings. However, it was found that the Chinese SMMR surface zone soil moisture is typically wetter while Mongolian and Russian SMMR surface zone soil moisture is typically drier than SMDB values, except in the northeastern Russian wet climate. Therefore the SMMR soil moisture estimates were found to be biased, less than 5% v/v dry in dry climate areas but over 10% v/v (as high as 20% v/v) wet in wet climate areas. Nonetheless, the seasonal variations were generally well represented. The largest SMMR biases were found in China and northeastern Russia, where precipitation and vegetation cover is also larger.

3.3. CLSM Predictions and Their Errors

[16] Novel CLSM features include its topographically defined catchments [*Koster et al.*, 2000], its explicit subgrid soil moisture variability treatment based on statistical topography induced heterogeneity, and its TOPMODEL [*Beven and Kirkby*, 1979] concept used to relate water table distribution to topography. This leads to the definition of three bulk moisture prognostic variables (catchment deficit, root zone excess and surface excess) with specific moisture transfer between them. Using these three prognostic variables, the catchment is divided into stressed, unstressed and saturated soil moisture regimes with separate evapotranspiration flux calculations for each, and the catchment average surface zone (top 2 cm), root zone (top 1 m) and profile (from 1 to ~3.5 m depending on total soil depth) soil moisture values calculated. The CLSM has been coupled with the GMAO's atmosphere and ocean models (see *Walker and Houser* [2001] for a more detailed summary of the CLSM).

[17] The quality and error of CLSM soil moisture predictions have been fully evaluated by comparing model predicted surface and root zone soil moisture with SMDB data of *Ni-Meister et al.* [2005]. The averaged summer model error (variance) in surface soil moisture for the period of 1979–1987 is shown in Figure 2c, with a spatial average of 10% v/v. Moreover, the CLSM was found to be dry biased in the already relatively dry Mongolian and southern Russian regions, and wet biased along the relatively wet Chinese east coast and the boundary between Mongolia and Russia. The evaluation study also found that the CLSM surface and root zone soil moisture was biased less than 8%

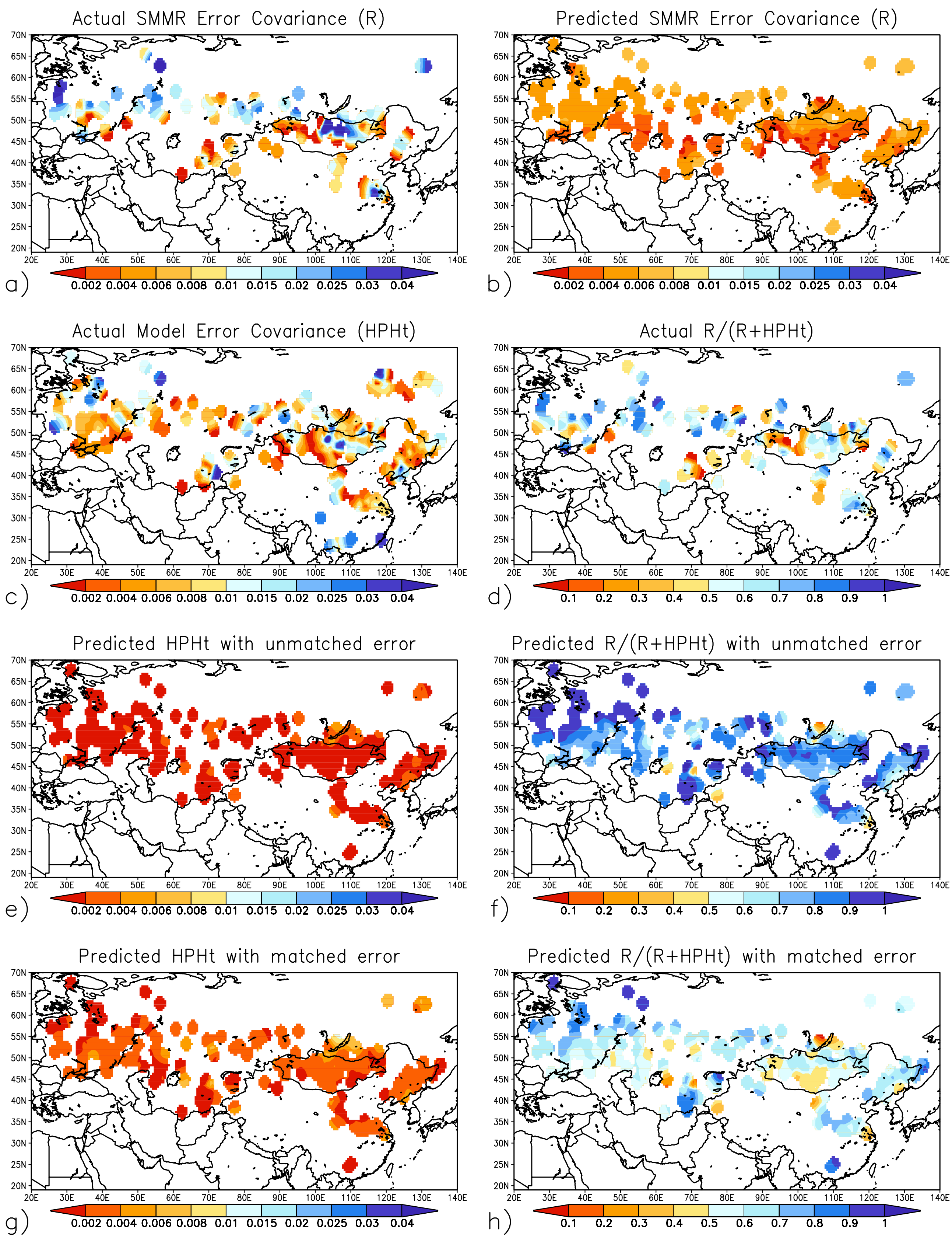

Figure 2. Actual and predicted summertime surface soil moisture observation error covariance (R) and model error covariance (HPHt), and their ratios for the period of 1979–1987.

Table 1. Spatial Average of Different Error Components

	$\sqrt{R}$, v/v	$\sqrt{HPHt}$, v/v	R/(R + HPHt)
Observed	0.12	0.1	0.56
Predicted unmatched error	0.07	0.03	0.82
Predicted matched error	0.07	0.05	0.61

v/v dry in dry climate and frozen soil areas, but biased over 8% v/v (as high as 16% v/v) wet in wet climate areas. Additionally, CLSM suffered from an underestimation in the surface zone seasonal soil moisture variation.

4. Application of the Ensemble Kalman Filter in This Study

[18] Both the model and satellite observation errors, or more accurately their error ratio (i.e., $\mathbf{R}/(\mathbf{R} + \mathbf{HPH}^{\mathbf{T}})$ where $\mathbf{HPH}^{\mathbf{T}}$ is the error covariance of the predicted observation), are required by the Kalman filter. As the previous work by *Ni-Meister et al.* [2005] on true model and satellite observation error (see Figures 2a and 2c) was only able to provide indicative error climatologies, due to the sparse ground observation network and infrequent measurements for each station, it was necessary to obtain continuous through space and time error estimates from alternative measures, rather than use the observed values directly. To this end the error estimates provided with the SMMR-derived soil moisture data set from *Owe et al.* [2001] were used to approximate $\mathbf{R}$, the model ensemble was used to approximate $\mathbf{P}$, and an assessment made against the observed true climatologies.

[19] The SMMR error estimates provided with the data set from *Owe et al.* [2001] were derived as a function of retrieved vegetation optical depth (Figure 2b). These error estimates were found to be smaller than the observed satellite observation errors (Figure 2a), which were estimated by comparing the SMMR derived soil moisture with the station data (see the details in the work by *Ni-Meister et al.* [2005]), having average values of 7% v/v and 12% v/v, respectively (Table 1).

[20] The model error includes three components: (1) initialization error, (2) forcing error, and (3) model parameter and physics error. The error components from 1 and 2 are easily dealt with while the component from 3 is more problematic. The initialization error, or rather the uncertainty in the initial conditions, is represented by perturbing the ensemble of initial state estimates such that the ensemble spread represents this uncertainty. Likewise the forcing error is represented by an ensemble of forcing fields whose spread represent their respective uncertainty. As the level of uncertainty in initialization and model forcing fields is fairly well known, representative perturbations can be reliably estimated; magnitudes of perturbations used in this study were the same as those used in the studies by *Reichle et al.* [2002a] and *Reichle and Koster* [2005], and the reader is referred therein for details. However, the amount of error due to model parameters and physics as apart from initialization and forcing is largely unknown and there are wide ranging recommendations on its inclusion. The approach taken by *Reichle and Koster* [2005] was to use three calibrated parameters that are then used to perturb the three soil moisture states at each time step, without considering how well the predicted error from the ensemble spread (Figure 2e) compared with the actual error based on comparison with station observations (Figure 2c). The calibration was based on optimization of assimilation results for a twin experiment where the truth was generated from the same model as used in the assimilation, just using a different forcing data set.

[21] In this data assimilation study two sets of model error parameters are assessed. The first (Figure 2e) are the calibrated (termed unmatched error) error parameter values by *Reichle et al.* [2002a] and the second (Figure 2g) are error parameter values chosen to maximize the agreement (termed matched error) between predicted and actual mean error ratio; i.e., ratio of the satellite observation error to the total error (satellite observation error + model error). The actual error ratio was calculated based on the SMMR (Figure 2a) and model error (Figure 2c) fields from comparison with station observations. Figures 2e and 2g show the unmatched and matched predicted surface soil moisture error estimates, respectively, and are to be compared with the actual surface soil moisture errors derived from comparison with station data in Figure 2c. The actual mean error ratio estimated from station observations (i.e., Figure 2a data divided by Figure 2c data) is shown in Figure 2d, while the predicted unmatched mean error ratio (i.e., Figure 2b data divided by Figure 2e data) is shown in Figure 2f and the predicted matched mean error ratio (i.e.,. Figure 2b data divided by Figure 2g data) is shown in Figure 2h.

[22] As described already, the Kalman filter updates the system states based on the relative magnitudes of the model and satellite observation errors. In this study, both the predicted model and satellite observation errors are allowed to remain smaller than the actual errors, but their ratio is constrained to the actual ratio. As shown in Figure 2, the spatial pattern of the predicted matched error ratio agrees more closely with the actual error ratio, with slightly smaller ratio values for some stations. A more quantitative analysis shows the actual mean ratio of SMMR error to total error is 0.56 and the predicted mean error ratios for the unmatched and matched error sets are 0.82 and 0.61, respectively (see Table 1). A mean error ratio value of 0.56 indicates that the model and SMMR errors are approximately equal, while the unmatched error ratio value of 0.81 suggests a much smaller predicted mean model error than SMMR error, meaning that the data assimilation scheme would trust the model more than the SMMR observation. The matched error ratio of 0.61 is in reasonably close agreement with the actual error ratio, and suggests that the model errors are on average slightly smaller than the SMMR error. Furthermore, the predicted model errors have not been calibrated at individual sites, but have instead been match to the measured error distribution for Eurasia. The spatial patterns of the predicted errors for the satellite observations and model predictions (and their ratios) have not been well represented in the spatial pattern of the actual errors.

[23] A station by station error ratio comparison (Figure 3) shows that both data assimilation error ratios have smaller variations than the observed error ratio, but the ratios for the matched case have slightly larger variations (0.4–1.0 for the matched case and 0.7–1.0 for the unmatched case). Moreover, there is a more even spread of values around the one-to-one relationship between actual and predicted error ratios

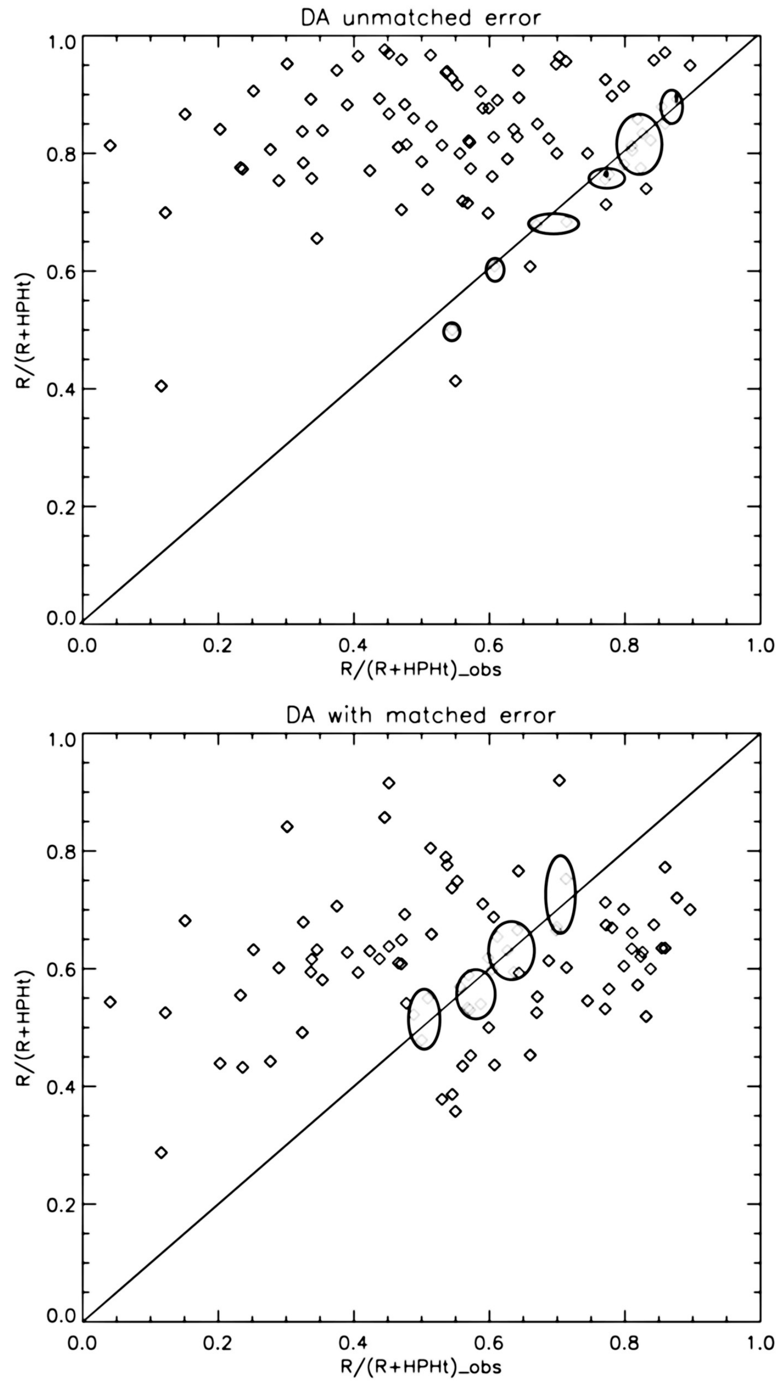

Figure 3. Station by station comparison of predicted (y axis) and actual (x axis) error ratios for matched and unmatched errors. Diamonds in ovals are stations with a modeled and observed error ratio difference of less than 0.05.

in the matched case than for the unmatched case. The following section discusses the implication for using both the unmatched and matched errors in assimilation runs.

5. Data Assimilation Results

5.1. Spatial Pattern Comparisons

[24] The soil moisture estimate improvement through data assimilation was assessed as the difference between the mean absolute error in model predictions with and without assimilation, for both the surface and root zone soil moisture for each of the four seasons. Figure 4 shows the summertime assimilated soil moisture improvement for 1979–1987 for all stations. Blue areas indicate a reduction in soil moisture estimate error through data assimilation while orange areas indicate an increase in soil moisture estimate error due to the data assimilation. Near zero values (within a

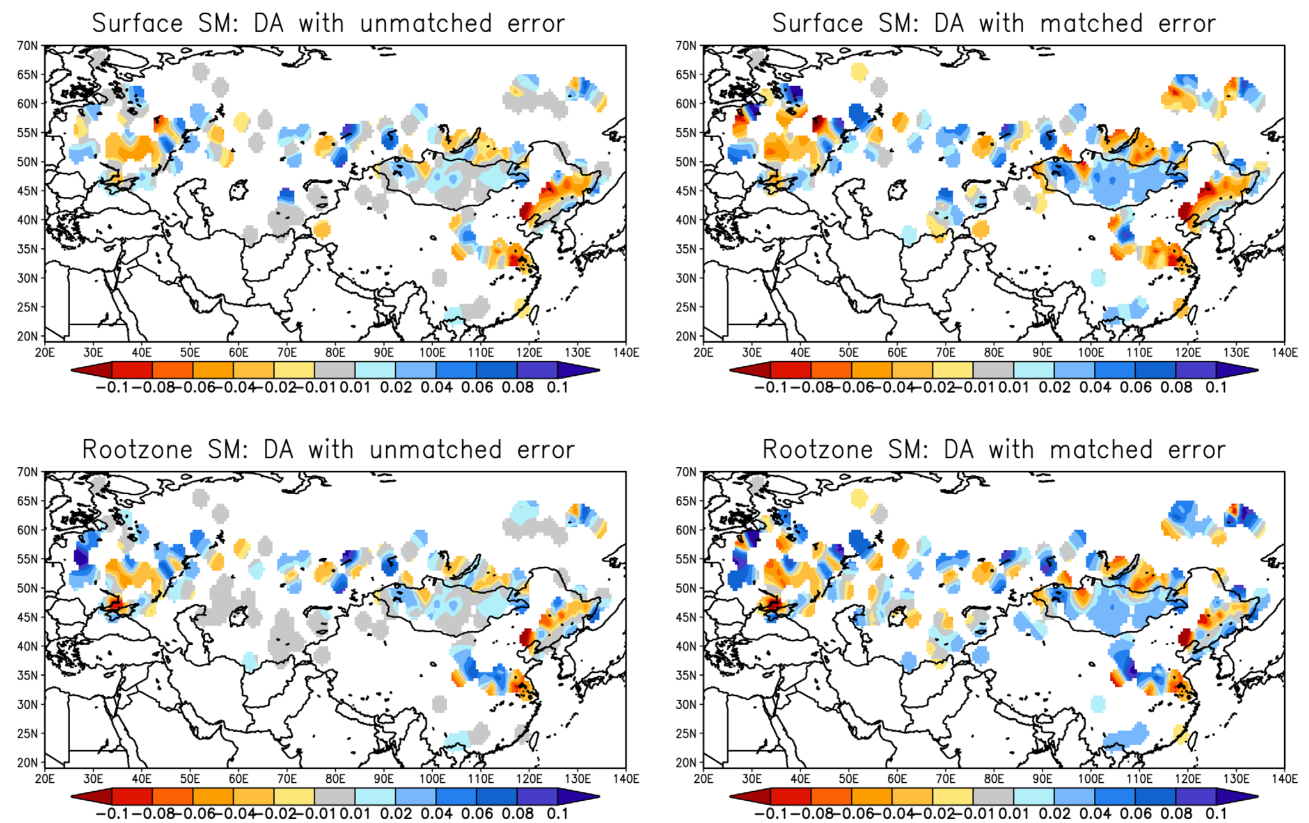

Figure 4. Assimilated summertime soil moisture improvement (absolute model error minus absolute assimilation error) in Eurasia for 1979–1987.

range of −1 and 1% v/v) in gray indicate a negligible impact from data assimilation on the soil moisture estimate. Figure 4 shows more blue than orange areas, particularly for root zone soil moisture, but is mixed with a lot of gray areas.

[25] For a more quantitative analysis, Table 2 lists the fraction and number (in brackets) of stations with improvement, no change, and degradation for Figure 4. The analysis shows that with unmatched error 46% (81 stations out of the total of 176) of the stations had an improvement in surface zone soil moisture and 39% (74 stations out of 176) of the stations had an improvement in root zone soil moisture. Using matched error the percentage of stations with an improvement in surface zone soil moisture was unchanged (47% verses 46%), but the percentage of stations showing an overall root zone soil moisture improvement increased from 39% to 57% (from 74 to 107 stations). However, there was also an increase in the number of stations that had degraded soil moisture, particularly in the root zone (from 40 to 60). The overall error improvement through data assimilation for matched and unmatched errors is 0.5% v/v (ranging from 0.45% v/v to 0.8% v/v).

[26] Figure 5 and Table 3 are similar to Figure 4 and Table 2, except they only show results for the stations which have a smaller climatological mean SMMR error compared to the model error. In contrast to the results for all stations, the unmatched error results show 78% of the stations having an improvement in the surface zone soil moisture estimate and 72% of the stations having an improvement in the root zone soil moisture estimate. With the matched error this increases to 83% of the stations having an improvement in the surface zone soil moisture and 87% of the stations having an improvement in the root zone soil moisture. For both matched and unmatched error cases, only small fractions of the stations (ranging from 5% to 8%) had degraded soil moisture as a result of the assimilation, with the overall average soil moisture improvement increasing from 0.5% v/v to around 3% v/v. The improvement increased by a further 1% v/v (from 3% v/v to 4% v/v for surface soil moisture and 2% v/v to 3% v/v for root zone soil moisture) when using the matched error as compared to the unmatched error. Spatial analysis of Figures 4 and 5 shows soil moisture improvement through data assimilation for most stations in Mongolia and Russia and a few stations in China; many stations in China have wet biased SMMR observations. Also

Table 2. Fraction and Number of Stations With Soil Moisture Improvement Through Data Assimilation Using the Matched and Unmatched Error Statistics During Summers of 1979–1987

DA Results	Improved	No Change	Worse	DA Improvement
Error difference	>0.01	[−0.01,0.01]	<0.01	
		Unmatched Error		
Surface SM	0.46(81)	0.18(31)	0.36(64)	0.0045
Root zone SM	0.39(74)	0.40(75)	0.21(40)	0.0048
		Matched Error		
Surface SM	0.47(82)	0.15(27)	0.38(67)	0.0045
Root zone SM	0.57(107)	0.12(22)	0.32(60)	0.0080

Units are v/v, and values in parentheses are number of stations.

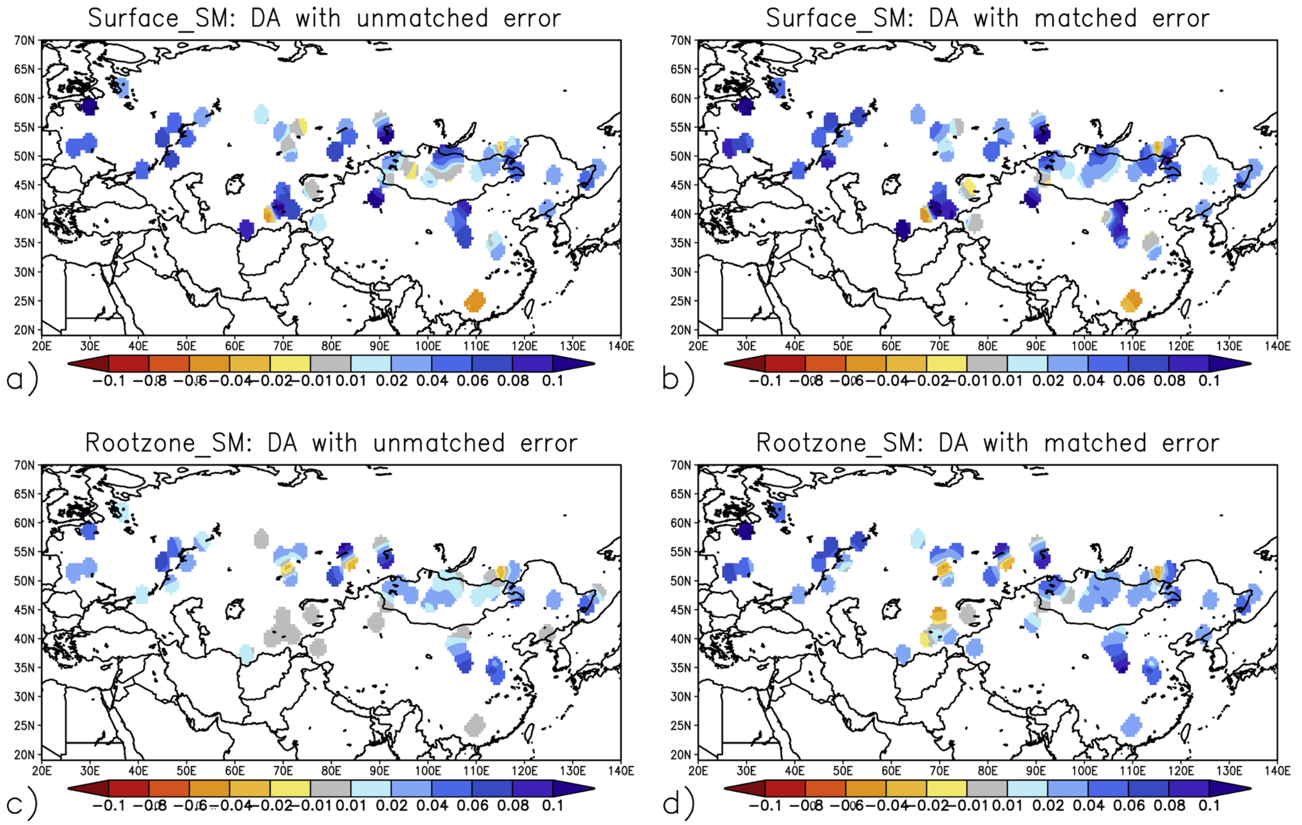

Figure 5. As for Figure 4, but only for cases where the SMMR error is less than the model error.

note that uncertainty exists in these comparisons since the model and SMMR derived soil moisture data at 25 km scales for top 1–2 cm are compared with top 10 cm ground data at a point.

5.2. Comparisons for All Stations

[27] Figure 6 shows the comparison of root-mean-square error (RMSE) in modeled surface zone and root zone soil moisture throughout Eurasia for the period of 1979–1987. The RMSE was calculated from the difference between the modeled and station measured soil moisture with and without assimilation. The scatterplot for the matched error shows a wider spread along the 1:1 line than for the unmatched error case, confirming the earlier assertion that smaller model error estimates in the unmatched case will result in smaller updates to the predicted soil moisture. Moreover, many Chinese stations show a data assimilation improvement in root zone soil moisture with corresponding degradation in surface zone soil moisture. Previous analysis of model results has shown that the CLSM simulates a depressed vertical change of soil moisture in this region, with a dry biased root zone soil moisture prediction [*Ni-Meister et al.*, 2005]. The assimilation of wet biased surface soil moisture observations in this case was able to improve the root zone soil moisture estimate through data assimilation, while making the surface zone soil moisture wet biased; this was due to the opposite sign of the biases. However, root zone soil moisture improvement adds significant value toward improved seasonal climate prediction of precipitation through its greater evapotranspiration feedback [*Koster et al.*, 2004].

[28] To better analyze the impact of SMMR error on data assimilation results, Figure 6 identifies the stations with SMMR errors that were less than the model error when no assimilation is performed. It can be seen that most of these stations are located below the 1:1 line, indicating that data assimilation has improved both the surface zone and root zone soil moisture estimates for both matched and unmatched cases when reliable satellite observations of surface soil moisture are available.

[29] To better analyze the impact of using better matched error ratio on the soil moisture estimates, Figure 7 compares the RMSE of soil moisture estimate with and without data assimilation for stations with the best match of the error ratio to the actual error ratio. The stations shown in Figure 7 are those within ovals in Figure 3, being stations with an error ratio difference of less than 0.05 between predicted and actual values. In this case data assimilation is constrained to places where the errors predictions are closer to

Table 3. Same as Table 2 but Only for Cases When the SMMR Error is Less Than the Model Error

DA Results	Improved	No Change	Worse	DA Improvement
Error difference	>0.01	[−0.01, 0.01]	<0.01	
		Unmatched Error		
Surface SM	0.78(47)	0.13(8)	0.8(5)	0.03
Root zone SM	0.72(43)	0.23(14)	0.5(3)	0.02
		Matched Error		
Surface SM	0.83(50)	0.10(6)	0.7(4)	0.04
Root zone SM	0.87(52)	0.5(3)	0.8(5)	0.03

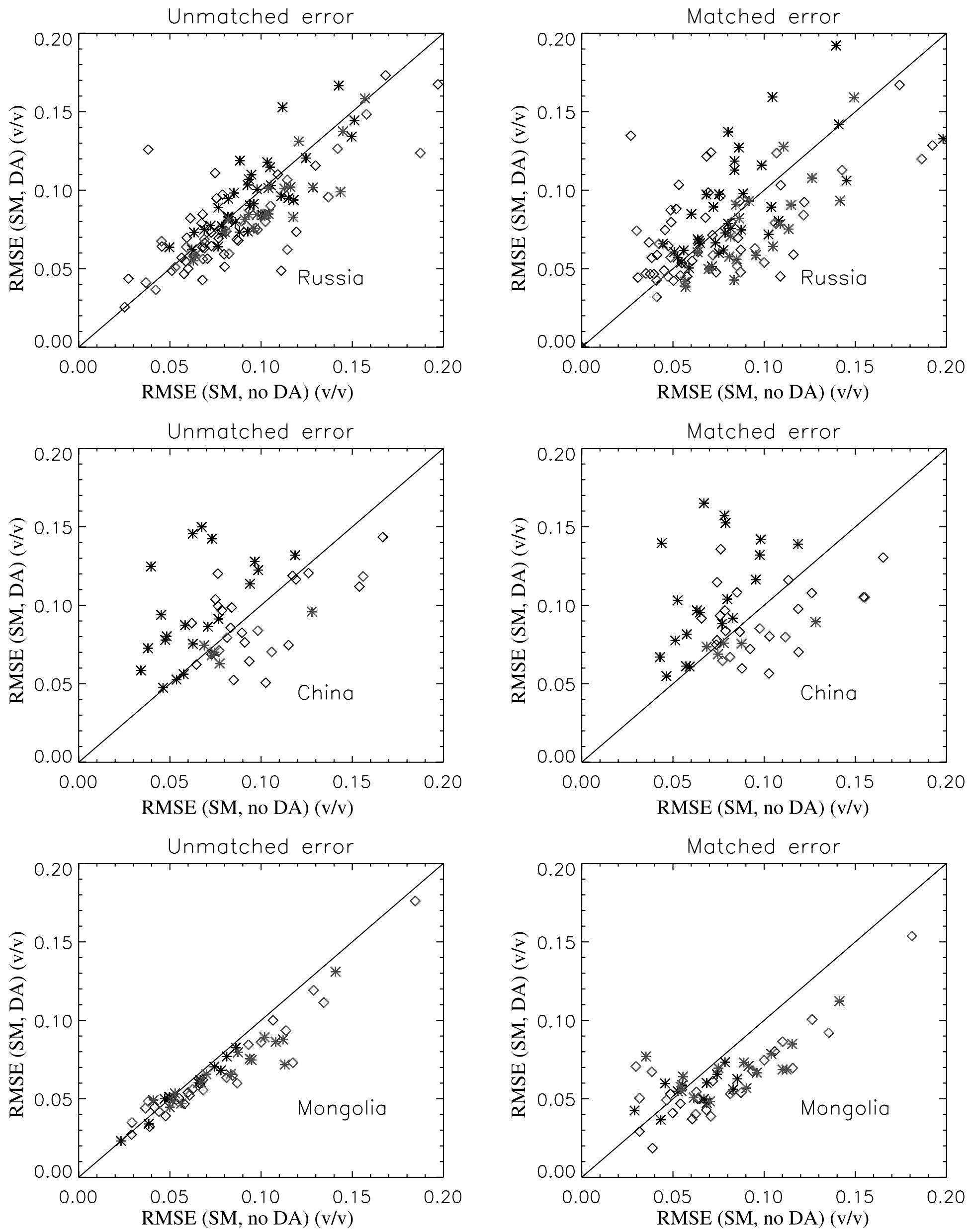

Figure 6. Summertime RMSE comparison of assimilated and modeled surface (stars) and root zone (diamonds) soil moisture in Eurasia for the period of 1981–1987 for (left) unmatched error and (right) matched error. Cases of SMMR error less than model error are in red and the rest in black. Data in the lower triangle show an improvement.

the actual model and satellite observation error ratios; ultimately this is how assimilation should be undertaken everywhere. Figure 7 shows that in the case of the matched errors (right panel) an improvement in soil moisture estimates was achieved all of the time, whereas the unmatched case made the results worse almost all of the time.

[30] Figure 8 shows a scatterplot of the surface and root zone soil moisture improvement due to assimilation; stations with SMMR error smaller or larger than the model error are differentiated. Most stations with small SMMR errors are located in the upper right quadrant, indicating that when satellite observation error is smaller than the model error, data assimilation leads to reduced surface and root zone soil moisture bias as well as reduced error. However, when the SMMR mean error is greater or equal to the model error, most stations are located in the lower left and lower right quadrants, indicating that data assimilation typically leads to an increase in soil moisture bias, particularly in the surface soil moisture estimates; most of the stations in China show degraded surface soil moisture estimates. As discussed earlier, previous work has shown significant SMMR-derived surface soil moisture bias in China. In this case bias correction is required before the observations can be rigorously used in the ensemble Kalman filter, since one assumption of the Kalman filter is that both the model and satellite observation errors are unbiased.

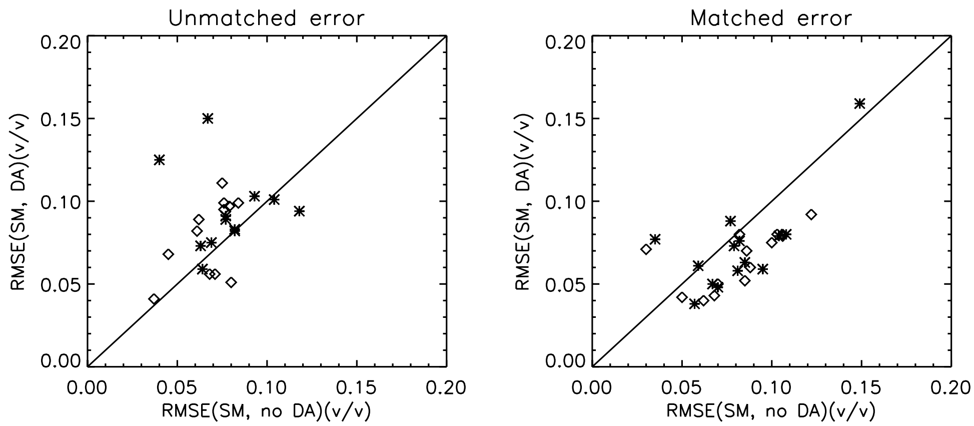

Figure 7. RMSE comparison of assimilated and modeled surface (stars) and root zone (diamonds) soil moisture in Eurasia for the period of 1981–1987 for locations where the modeled and observed error ratio difference was less than 0.05 (the diamonds in ovals shown in Figure 3).

5.3. Individual Station Comparisons

[31] To investigate the temporal response of soil moisture data assimilation, modeled soil moisture with and without assimilation for both matched and unmatched error cases are compared with satellite-derived and station-measured surface zone and root zone soil moisture for two Russia stations (Ogurtsovo from eastern Russia and Penza from western Russia), one China station (Zhenzhou) and one Mongolia station (Khalkhgol; see Figures 9–12). These stations were chosen as representative examples of catchments from these four regions.

[32] For the Ogurtsovo, Russia station comparison (Figure 9), both the modeled surface and root zone soil moisture underestimates the station data, while both the assimilated surface and root zone soil moisture estimates agree better with the station observations. In winter, there are not enough SMMR measurements due to frozen and/or snow covered soil, with the assimilated surface soil moisture tracking the ground observations less closely then during other times of the year. However, the results are still closer to the observations compared to the model results without assimilation, but not so wet as the ground observations; the ground observations show very high wintertime surface soil moisture values. For root zone soil moisture estimates, both assimilated soil moisture cases agree well with the station measurements, with poorer results obvious during wintertime when there are no observations available for assimilation. The SMMR data agree well with the ground measurements at this station.

[33] For the Penza, Russia station comparison (Figure 10), the model overestimates both the surface and root zone soil moisture compared to the station measurements. SMMR-derived surface soil moisture are closer to the ground observations and assimilating SMMR data into the catchment model results in better surface zone and root zone soil moistures agreement with the ground observations. The results from the matched error case show a slightly better agreement with the surface soil moisture measurements and the results from the unmatched error case show a better agreement with the root zone soil moisture measurements. Assimilating the SMMR data has also led to a larger temporal variation of surface soil moisture estimates than model estimates.

[34] For the Zhenshou, China station comparison, SMMR derived surface soil moisture overestimates ground observations, leading to the overestimation of surface soil moisture after assimilation. However, the model underestimates the root zone soil moisture, and assimilation helps bring the underestimated soil moisture up to the observed values, resulting in a somewhat better agreement with the station observations. Moreover, the assimilated root zone soil moisture from the matched error case does not show a better agreement with the station measurements than the ones with the unmatched error.

[35] For the Khalkhgol, Mongolia station comparison, Figure 12 shows the model underestimates both the surface zone and root zone soil moisture comparing to the ground measurements while the SMMR data match the ground measurements reasonably well. Both assimilated surface and root zone soil moisture show good ground measurement agreement. Again the assimilated surface and root zone soil moisture with the matched error agree better with the ground measurements than the unmatched case. Improvement in surface soil moisture estimate through assimilation also led to the improvement of root zone soil moisture estimate.

6. Discussion and Conclusions

[36] A SMMR derived surface soil moisture product has been assimilated into the GMAO CLSM using an ensemble Kalman filter. The Kalman filter corrects model-generated soil moisture toward the satellite observation, with the size of the correction dependent upon the relative magnitudes of the satellite observation and model errors. Satellite observation error estimates provided with the soil moisture product were estimated as a function of vegetation optical depth (average 7% v/v), but these were shown to underestimate the true error in the SMMR derived soil moisture (average 12% v/v). Moreover, two types of predicted model error estimates were tested. The first used unmatched errors based on model error parameter values from *Reichle et al.* [2002b]. In this case the predicted model error estimates

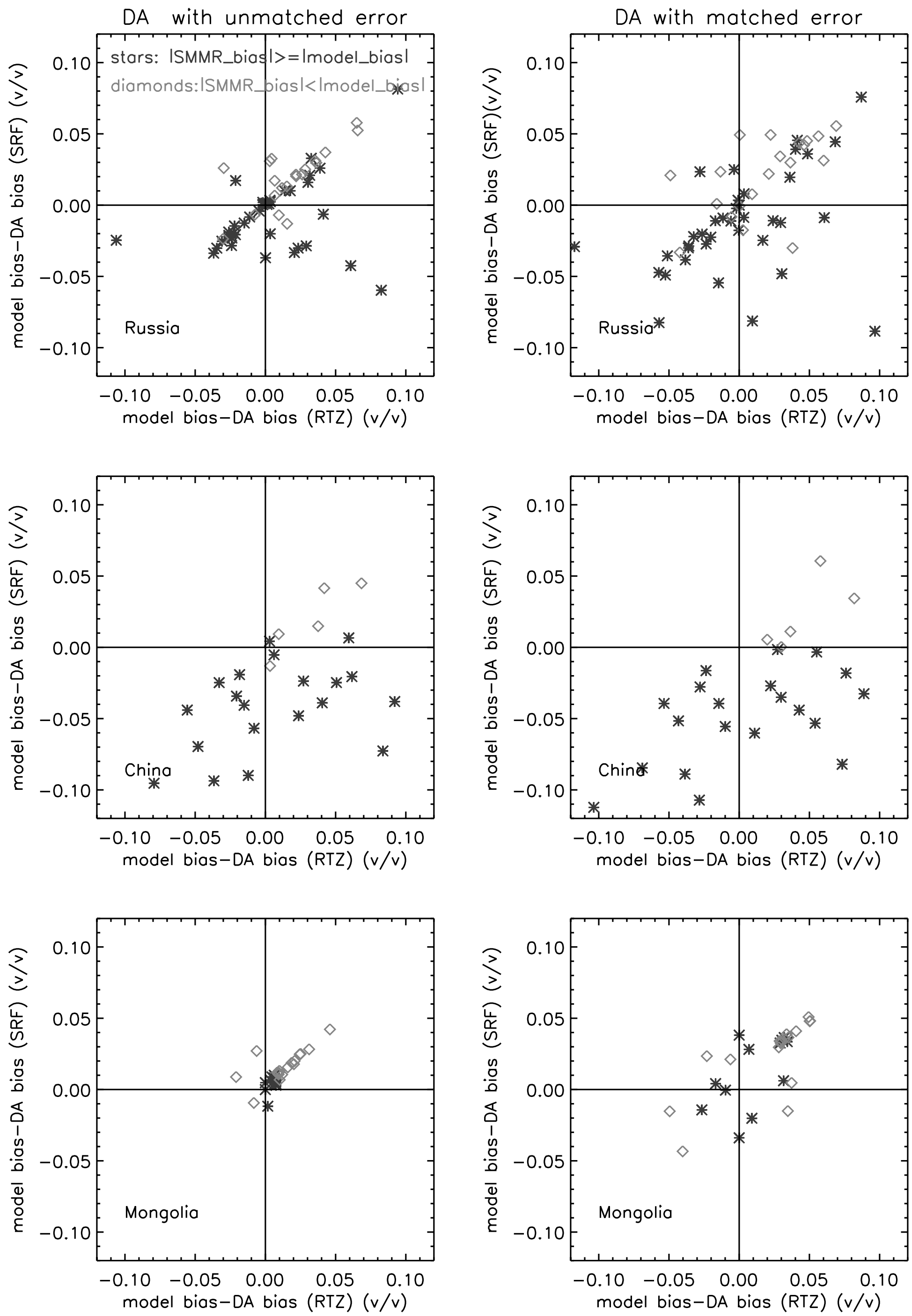

Figure 8. Comparison of surface and root zone soil moisture bias improvement in Eurasia for the period of 1981–1987. The stations with an average SMMR error less than model error are shown by diamonds; all other stations are shown by stars. Data located in the top right-hand quadrant show an improvement in both the surface and root zone soil moisture estimates.

were too small (average 3% v/v) relative to the actual error in model predicted soil moisture (average 10% v/v). The second used matched errors based on model error parameters that gave the best agreement of SMMR error to total error (model error plus SMMR error) ratio with the actual error ratio. In this case the resultant model and SMMR error magnitudes were approximately equal on average. The data assimilation results from both cases were then compared with ground observations.

[37] It has been shown that with the matched error, data assimilation provided more accurate root zone soil moisture estimates for 57% of stations as compared to 39% of stations with the unmatched error. This improvement in root zone soil moisture estimates has great implications for

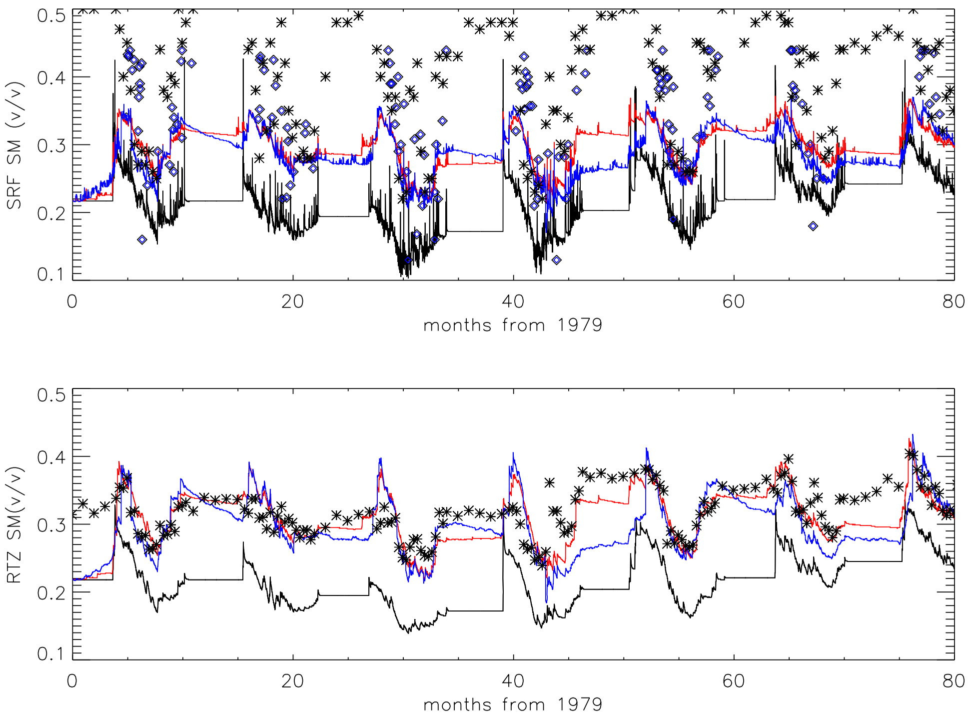

Figure 9. Comparison of (top) assimilated and modeled surface and (bottom) root zone soil moisture with station measurements and SMMR observations at Penza in Russia for the period of 1979–1985. Black stars are station measurements, black lines are modeled soil moisture, blue diamonds are SMMR observations, red lines are assimilated soil moisture with unmatched error, and blue lines are assimilated soil moisture with matched error.

seasonal climate prediction improvement, since better knowledge on root zone soil moisture is more important than surface zone soil in terms of land surface–atmosphere interactions. Moreover, these results highlight the need for accurate model and observation error estimates. Analysis of the impact of SMMR error on the assimilation results showed that when the climatologic mean SMMR error was less than the model error, the data assimilation

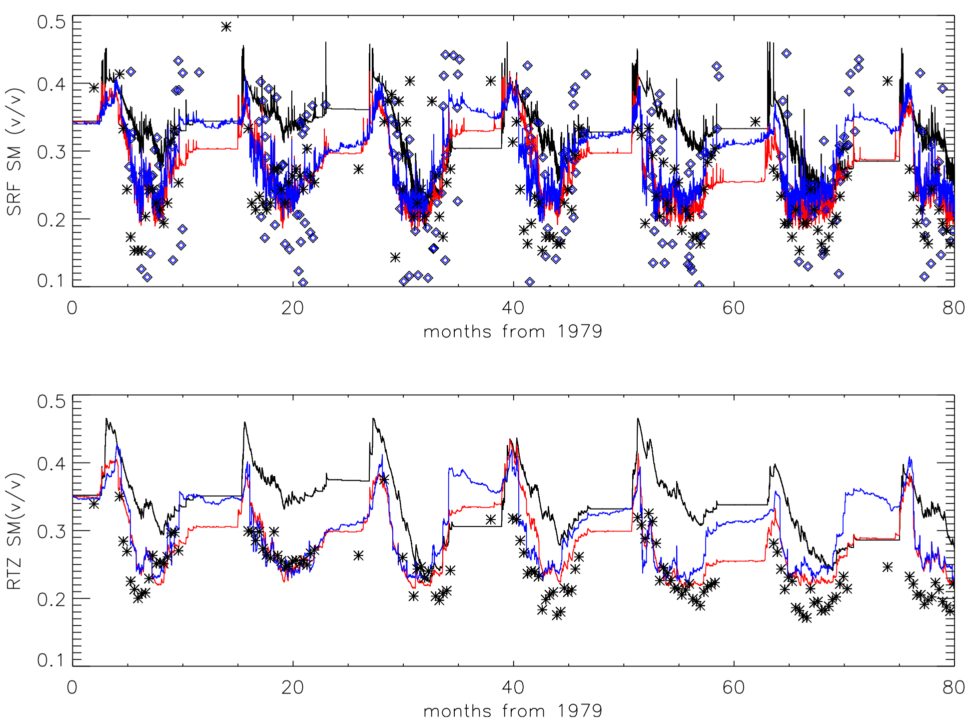

Figure 10. As for Figure 9 but for Ogurtsovo, Russia.

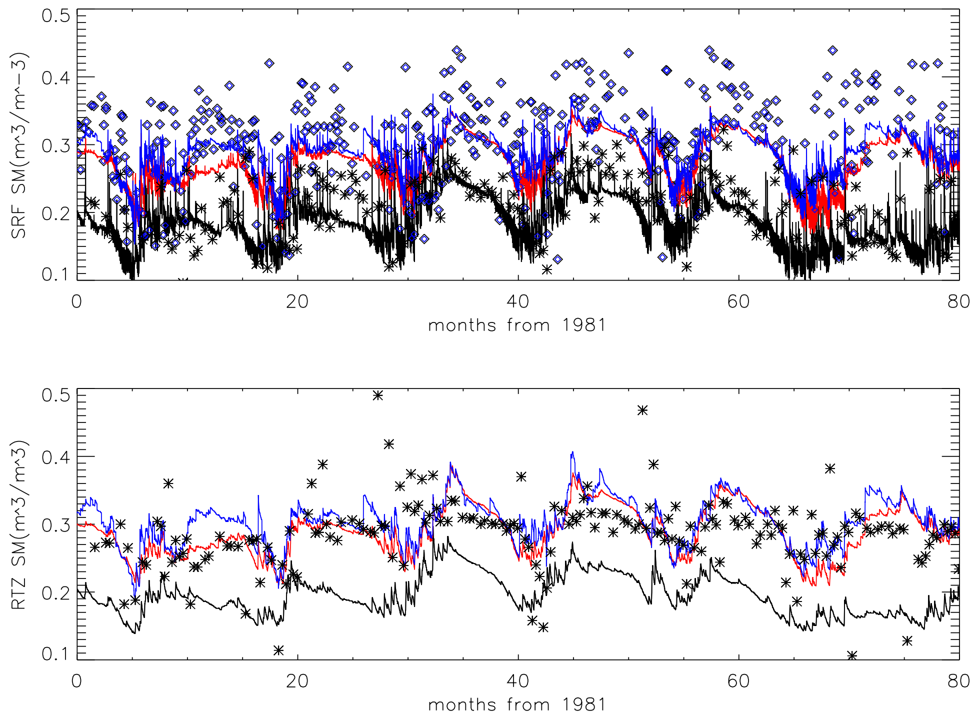

Figure 11. As for Figure 9 but for Zhenzhou, China for the period of 1981–1987.

improved both surface zone and root zone soil moisture estimates, with the greatest improvement being in the root zone. Moreover, the overall accuracy improvement of both surface and root zone soil moisture estimates increased from less than 0.9% v/v to approximately 3% v/v, indicating that good quality satellite products are required to ensure significant improvement in both surface zone and root zone soil moisture estimates over large regions. Additionally, most Mongolia and Russia stations showed soil moisture improvement as a result of data assimilation, while most China stations resulted in biased surface soil moisture estimates as a result of wet biased SMMR observations in

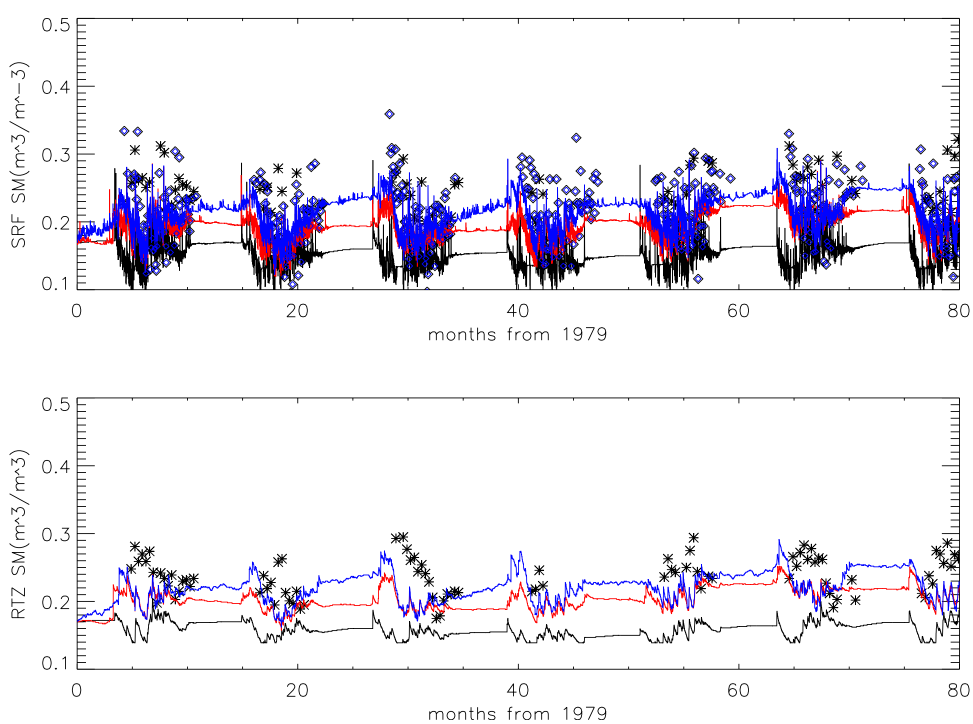

Figure 12. As for Figure 9 but for Khalkhgol, Mongolia.

that region. This indicates that bias must be removed before assimilation to ensure improvement in soil moisture estimates.

[38] **Acknowledgments.** This work was funded by NASA contract NNG04GL08G. The authors gratefully acknowledge A. Robock's group for developing the Soil Moisture Data Bank. We also thank Randy Koster for his many valuable CLSM discussions and Rolf Reichle for providing ensemble Kalman filter code, Manfred Owe and Richard de Jeu for providing the SMMR soil moisture data, Aaron Berg for providing the forcing data, and Sarith Mahanama for providing the CLSM soil and vegetation parameters.

References

Bennett, A. F. (1992), *Inverse Models in Physical Oceanography*, 246 pp., Cambridge Univ. Press, New York.

Beven, K. J., and M. J. Kirkby (1979), A physically based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*(1), 43–69.

Bindlish, R., T. J. Jackson, E. Wood, H. L. Gao, P. Starks, D. Bosch, and V. Lakshmi (2003), Soil moisture estimates from TRMM microwave imager observations over the southern United States, *Remote Sens. Environ.*, *85*(4), 507–515.

Burgers, G., P. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*, 1719–1724.

Daley, R. (1991), *Atmospheric Data Assimilation*, Cambridge Univ. Press, New York.

de Jeu, R., and M. Owe (2003), Further validation of a new methodology for surface moisture and vegetation optical depth retrieval, *Int. J. Remote Sens.*, *24*, 4559–4578.

Dong, J., J. P. Walker, and P. R. Houser (2005), Factors affecting remotely sensed snow water equivalent uncertainty, *Remote Sens. Environ.*, *97*(1), 68–82, doi:10.1016/j.rse.2005.04.010.

Dunne, S., and D. Entekhabi (2006), Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment, *Water Resour. Res.*, *42*, W01407, doi:10.1029/2005WR004334.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162.

Gao, H., E. F. Wood, M. Drusch, W. Crow, and T. J. Jackson (2004), Using a microwave emission model to estimate soil moisture from ESTAR observations during SGP99, *J. Hydrometeorol.*, *5*, 49–63.

Gao, H., E. F. Wood, M. Drusch, T. Jackson, and R. Bindlish (2006), Using TRMM/TMI to retrieve soil moisture over the southern United States from 1998 to 2002, *J. Hydrometeorol.*, *7*, 23–38.

Gelb, A. (1974), *Applied Optimal Estimation*, 373 pp., MIS Press, New York.

Houser, P. R., W. J. Shuttleworth, H. V. Gupta, J. S. Famiglietti, K. H. Syed, and D. C. Goodrich (1998), Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resour. Res.*, *34*(12), 3405–3420.

Houser, P. R., et al. (2001), The global land data assimilation system, *GEWEX News*, *11*(2), 11–13.

Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter techniques, *Mon. Weather Rev.*, *126*, 796–811.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*, 35–45.

Keppenne, C. L. (2000), Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter, *Mon. Weather Rev.*, *128*, 1971–1981.

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

Koster, R. D., M. J. Suarez, P. Liu, U. Jambor, A. A. Berg, M. Kistler, R. H. Reichle, M. Rodell, and J. Famiglietti (2004), Realistic initialization of land surface states: Impacts on subseasonal forecast skill, *J. Hydrometeorol.*, *5*(6), 1049–1063.

Ni-Meister, W., J. P. Walker, and P. R. Houser (2005), Soil moisture initialization for climate prediction: Characterization of model and observation errors, *J. Geophys. Res.*, *110*, D13111, doi:10.1029/2004JD005745.

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

Radakovish, J. D., P. R. Houser, A. da Silval, and M. G. Bosilovich (2001), Results from global land-surface data assimilation methods, paper presented at AMS 5th Symposium on Integrated Observing Systems, Am. Meteorol. Soc., Albuquerque, N. M., 14–19 Jan.

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

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002a), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*(1), 103–114.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended vs. ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, *3*(6), 728–740.

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. Am. Meteorol. Soc.*, *81*, 1281–1299.

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

Sun, C., J. P. Walker, and P. R. Houser (2004), A methodology for snow data assimilation in a land surface model, *J. Geophys. Res.*, *109*, D08108, doi:10.1029/2003JD003765.

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.

Walker, J. P., N. Ursino, R. B. Grayson, and P. R. Houser (2003), Australian root zone soil moisture: Assimilation of remote sensing observations, in *Proceedings of the International Congress on Modelling and Simulation (MODSIM)*, edited by D. Post, pp. 380–385, Modell. and Simul. Soc. of Aust. and N. Z., Inc., Townsville, Australia.

Zhang, S. W., H. R. Li, W. D. Zhang, C. J. Qiu, and X. Li (2005), Estimating the soil moisture profile by assimilating near-surface observations with the ensemble Kalman filter (EnKF), *Adv. Atmos. Sci.*, *22*(6), 936–945.

P. R. Houser, Center for Research on Environment and Water, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705-3106, USA.

J. P. Walker, Department of Civil and Environmental Engineering, University of Melbourne, Parkville, Victoria 3010, Australia.

W. Ni-Meister, Department of Geography, Hunter College of City University of New York, 695 Park Ave., New York, NY 10021, USA. (wenge.ni-meister@hunter.cuny.edu)