Corresponding Author: Dr. Clara Sophie Draper, PhD

Corresponding Author's Institution: NASA GSFC

First Author: Clara Sophie Draper, PhD

Order of Authors: Clara Sophie Draper, PhD; Rolf R Reichle; Richard A de Jeu; Vahid Naeimi; Robert M Parinussa; Wolfgang Wagner

- RMSE is presented as a fraction of the timeseries standard deviation, fRMSE

- Error propagation & triple colocation accurately detect spatial variability in fRMSE

- Triple colocation accurately estimates the magnitude of soil moisture anomaly fRMSE

- Triple colocation is robust to representativity differences between data sets used

- ASCAT and AMSR-E have similar anomaly fRMSE for most land covers in the study domain

Estimating root mean square errors in remotely sensed soil moisture over continental scale domains

Clara Draper[a,b,*], Rolf Reichle[a], Richard de Jeu[c], Vahid Naeimi[d], Robert Parinussa[c], Wolfgang Wagner[e]

[a]*NASA Goddard Space Flight Center, Global Modeling and Assimilation Office, Code 610.1, Greenbelt, 20771, Maryland, USA. (301) 614 6161.*
[b]*GESTAR, Universities Space Research Association, Columbia, Maryland, USA.*
[c]*Earth and Climate Cluster, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands*
[d]*German Remote Sensing Data Centre, German Aerospace Center, Wessling, Germany.*
[e]*Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria.*

Abstract

Root Mean Square Errors (RMSE) in the soil moisture anomaly time series obtained from the Advanced Scatterometer (ASCAT) and the Advanced Microwave Scanning Radiometer (AMSR-E; using the Land Parameter Retrieval Model) are estimated over a continental scale domain centered on North America, using two methods: triple colocation (RMSE^{TC}) and error propagation through the soil moisture retrieval models (RMSE^{EP}). In the absence of an established consensus for the climatology of soil moisture over large domains, presenting a RMSE in soil moisture units requires that it be specified relative to a selected reference data set. To avoid the complications that arise from the use of a reference, the RMSE is presented as a fraction of the time series standard deviation (fRMSE). For both sensors, the fRMSE^{TC}

*Corresponding author
Email address:* `clara.draper@nasa.gov` (Clara Draper)

and fRMSE^{EP} show similar spatial patterns of relatively high/low errors, and the mean fRMSE for each land cover class is consistent with expectations. Triple colocation is also shown to be surprisingly robust to representativity differences between the soil moisture data sets used, and it is believed to accurately estimate the fRMSE in the remotely sensed soil moisture anomaly time series. Comparing the ASCAT and AMSR-E fRMSE^{TC} shows that both data sets have very similar accuracy across a range of land cover classes, although the AMSR-E accuracy is more directly related to vegetation cover. In general, both data sets have good skill up to moderate vegetation conditions.

1. Introduction

Soil moisture is an important control over hydrological and meteorological forecasts, since it can determine the partitioning of energy and moisture incident at the land surface. Increasing recognition of the role of soil moisture has motivated recent developments in globally observing near-surface soil moisture from satellites. These developments have included retrieving soil moisture from already orbiting sensors, such as the Advanced Scatterometer (Wagner et al., 1999; Bartalis et al., 2007) and the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) (Njoku, 1999; Owe et al., 2001). Additionally, several new remote sensors have been specifically designed to sense soil moisture, including the European Space Agency's Soil Moisture Ocean Salinity (SMOS) mission, launched in 2009 (Kerr et al., 2001), and NASA's Soil Moisture Active Passive mission, sched-

uled for launch in 2014 (Entekhabi & coauthors, 2010).

The performance of new remotely sensed soil moisture data sets is benchmarked against predetermined Root Mean Square Error (RMSE) target accuracies (Kerr et al., 2001; Entekhabi & coauthors, 2010) based on comparison to pixel scale near-surface soil moisture observations obtained from either dense networks of in situ sensors (Jackson et al., 2012) or low-level ground-based/airborne microwave sensors (Gherboudj et al., 2012). However, these pixel scale observations are available at only a handful of locations, and further development and application of remotely sensed soil moisture data sets will require a better understanding of their accuracy across the globe.

Evaluating soil moisture over continental scale domains is not straight forward, since the true global soil moisture is unknown, due to the systematic differences between soil moisture estimates obtained from different remote sensors and numerical models (Reichle et al., 2004). These systematic differences can arise from i) differences in the soil and vegetation parameters assumed, or ii) representativity differences, for example due to differences in horizontal, vertical, and temporal support (Vinnikov et al., 1999; Reichle et al., 2004), or differences in the soil moisture processes resolved by each soil moisture estimate (Koster et al., 2009)

In the literature a common approach to evaluating soil moisture over continental scales has been to use the Root Mean Square Difference (RMSD) with an alternative soil moisture estimate, for example from a model (dall'Amico et al., 2012), or from networks of sparse in situ soil moisture sensors (Wagner et al., 1999; Reichle et al., 2007; Draper et al., 2009). However, this approach generates misleading results, since the errors in the alternative data set are

included in the RMSD (hence, the use of root mean square *difference*, rather than *error*).

Consequently, this study investigates recently developed methods to estimate distributed Root Mean Square Errors (RMSE) in remotely sensed soil moisture over continental-plus scale domains. The focus is on the RMSE for consistency with the metric specified for remote sensing target accuracies. Also, the RMSE is useful for specifying observation error variances for data assimilation. The RMSE is estimated for two remotely sensed soil moisture products: the Surface Degree of Saturation (SDS) retrieved from active microwave ASCAT observations (Wagner et al., 1999; Bartalis et al., 2007), and the X-band passive microwave AMSR-E soil moisture retrieved with the Land Parameter Retrieval Model (LPRM; Owe et al. (2001); de Jeu & Owe (2003)). While neither of these missions were designed to sense soil moisture, both have been providing useful soil moisture observations (Draper et al., 2012), with the advantage of a relatively long data record.

Two methods for estimating the RMSE of the ASCAT and AMSR-E soil moisture data are investigated. The first method is triple colocation (Stoffelen, 1998; Scipal et al., 2008b), which combines three independent estimates of a state variable to calculate the errors in each, by assuming an additive error model. The second method is error propagation through the model used to retrieve soil moisture from the microwave observations, as developed by Naeimi et al. (2009) for the ASCAT SDS and Parinussa et al. (2011b) for the AMSR-E LPRM retrievals. The error estimates are investigated over a continental scale domain, between 25-50°N in North America.

Due to the systematic differences between large scale soil moisture esti-

4

mates, different soil moisture data sets describe different climates as measured by their central moments. Without knowledge of the true soil moisture climate, these differences cannot be attributed to errors in a particular data set. Consequently, when comparing soil moisture data sets over large domains, the systematic differences between their mean and variance (and possibly higher-order moments) are typically eliminated by rescaling all data sets to have statistics consistent with an arbitrarily selected reference data set (Reichle & Koster, 2004; Scipal et al., 2008a). Hence over large domains, soil moisture RMSEs estimated by comparing different data sets are based on rescaled data sets, and are then presented relative to the climatology of the reference (e.g., Scipal et al. (2008b); Draper et al. (2009); Dorigo et al. (2010); dall'Amico et al. (2012)). Before the triple colocation and error propagation RMSE estimates are presented in this study, the effect of this rescaling to a reference data set on the subsequent RMSE is demonstrated, to establish how the RMSE should be interpreted.

The remainder of this paper is structured as follows. The soil moisture data sets and RMSE estimation methods are reviewed in Sections 2 and 3, respectively. The latter includes the introduction of statistical uncertainty estimates for the triple location RMSE, and the development of a strategy to compare RMSE estimates calculated over large domains from rescaled soil moisture data sets. The ASCAT and AMSR-E triple colocation and error propagation RMSE estimates are then examined in Section 4.1 to establish how useful the two methods might be for evaluating remotely sensed soil moisture. Also, the assumptions underlying triple colocation are tested in Section 4.2, by examining the dependence of the estimated RMSE on the

triplet of data sets used. Finally, a discussion of the implications of the results, and the conclusions drawn from this study are presented in Sections 5 and 6, respectively.

2. Data

2.1. Remotely sensed soil moisture data sets

ASCAT is a C-band scatterometer, orbiting in a sun-synchronous orbit on EUMETSAT's MetOp satellite. The soil moisture data used here were retrieved from ASCAT backscatter observations at the Vienna University of Technology, using the semiempirical change detection approach of Wagner et al. (1999) and Bartalis et al. (2007) (WARP 5.4 version). This yields an observation of the surface degree of saturation, ranging between 0 and 100%, representing the driest and wettest observation at each location, respectively. While the SDS must be multiplied by the porosity to give a soil moisture value, it will be referred to here as a soil moisture observation for convenience. The ASCAT SDS relate to soil moisture over a $\sim$1 cm deep surface layer, with a spatial resolution of 25 km (reported on a 12.5 km grid).

The AMSR-E instrument, orbiting on NASA's Aqua satellite in a sun-synchronous orbit, observed at six-dual polarized frequencies of which the two lowest (C- and X-band) are routinely used to infer soil moisture. The AMSR-E soil moisture data used here were retrieved at the VU University Amsterdam from X-band brightness temperatures using the LPRM (Owe et al., 2001; de Jeu & Owe, 2003). At X-band, AMSR-E observations relate to a surface layer depth slightly less than 1 cm with a horizontal resolution close to 40 km, although the swath data (reported every 5-10 km) were used

6

here.

The maximum available coincident data record, spanning ∼4.75 years, from January 2007 (first ASCAT data) to October 2011 (failure of AMSR-E) has been used. To avoid complications from the differing statistical moments of day- and nighttime observations, only nighttime data have been used. On average the nighttime crossing over North America occurs at 3 UTC (9 pm) for the (ascending) ASCAT overpass, and at 9 UTC (1 am) for the (descending) AMSR-E overpass. Both satellite overpasses were assumed to occur at 6 UTC, and have been interpolated to a 25 km grid, before being cross-screened to retain only locations and times for which both data sets are available.

For ASCAT, locations with dense vegetation were screened using the error propagation RMSEs provided with the data (see Section 3.2), following Mahfouf (2010) and Dharssi et al. (2011). An upper limit of 14% (in SDS units) was applied. For AMSR-E, dense vegetation was screened using an upper threshold of 0.8 for the vegetation optical depth, which is retrieved in parallel with the soil moisture (Owe et al., 2001). Both soil moisture data sets were also screened to remove grid cells with a wetland fraction above 10%, or where the Catchment land surface model (Section 2.2) indicates frozen conditions, snow cover, or precipitation. Additionally, the ASCAT soil moisture observations were discarded where the topographic complexity was above 10% (Draper et al., 2012), and LPRM observations flagged as having moderate or strong radio frequency interference were also discarded. Finally, a lower cut-off of 100 coincident data was imposed at each grid cell.

Figure 1 shows a map of the land cover classes for the regions where re-

motely sensed data are available after the above quality control. On average, there were 272 coincident data at each grid cell plotted. The quality control has screened out most of the grid cells with densely vegetated classes, however small pockets of deciduous broadleaf, evergreen needleleaf, and woody savanna remain, as well as large regions of mixed forest, and crop/natural mix in the east. The ASCAT and AMSR-E soil moisture data are not expected to have any skill over these densely vegetated land cover classes, and an additional screening is usually applied to the soil moisture data based on independent vegetation data (e.g., Draper et al. (2012)). However, this was not done here, to test whether the error estimation methods under investigation can detect the larger errors expected for densely vegetated conditions.

2.2. Catchment model soil moisture

Soil moisture simulations from NASA's Catchment land surface model (Koster et al., 2000) were used as the third data set in the triple colocation calculations. Catchment was run on a 25 km grid over the experiment domain, using meteorological data from the NASA Modern-Era Retrospective analysis for Research and Applications (MERRA) (Rienecker et al., 2011), with the precipitation forecasts corrected towards rain gauge observations. The near-surface soil moisture (0-2 cm) simulated at 6 UTC each day was then extracted for comparison to the remotely sensed data.

2.3. In situ soil moisture data

In situ soil moisture observations were used as an alternate data set to test the assumptions underlying the triple colocation method at the SCAN/SNOTEL (Schaefer et al., 2007) sites shown in Figure 1. At each of these sites a daily

8

time series of near-surface (0-5 cm) soil moisture observations at 6 UTC was sampled from the hourly SCAN/SNOTEL observations. After cross-screening the in situ observations for the availability of ASCAT and AMSR-E observations and applying a lower cut-off of 100 coincident observations, 57 SCAN/SNOTEL sites were included in this study (Figure 1), with an average of 261 coincident observations at each site.

3. Methods

3.1. Triple colocation

Triple colocation has been used to estimate the errors in the soil moisture anomaly time series from ASCAT (θ_A), AMSR-E (LPRM) (θ_L), and the Catchment model (θ_C), using the method described by Stoffelen (1998). For each data set the soil moisture anomaly time series was constructed using the difference of the raw data from their multi-year, seasonally varying climatology. The seasonal climatology was computed as the 31 day moving average, with the moving averages based on data from all years for the 31 day period surrounding each day of year.

At each grid cell, the anomaly soil moisture time series are assumed to include a signal of the true soil moisture anomalies (θ) plus a zero-mean error ϵ:

$$\theta_A = \alpha(\theta + \epsilon_A) \tag{1}$$

$$\theta_L = \lambda(\theta + \epsilon_L) \tag{2}$$

$$\theta_C = \gamma(\theta + \epsilon_C) \tag{3}$$

where α, λ, and γ are the triple colocation calibration constants, used to rescale the data sets to eliminate the systematic differences in their variability. A bias term was not included, since anomaly time series have been used. There are insufficient degrees of freedom to solve for all terms, and so one data set is selected as the reference and the remaining two are calibrated to be consistent with this reference. For example, if θ_A is the reference, $\alpha = 1$, and the remaining calibration constants are estimated:

$$\frac{<\theta_L \theta_C>}{<\theta_A \theta_C>} = \frac{\lambda <\theta^2 + \theta\epsilon_L + \theta\epsilon_C + \epsilon_L\epsilon_C>}{<\theta^2 + \theta\epsilon_A + \theta\epsilon_C + \epsilon_A\epsilon_C>} \tag{4}$$

$$\frac{<\theta_L \theta_C>}{<\theta_A \theta_L>} = \frac{\gamma <\theta^2 + \theta\epsilon_L + \theta\epsilon_C + \epsilon_L\epsilon_C>}{<\theta^2 + \theta\epsilon_A + \theta\epsilon_L + \epsilon_A\epsilon_L>} \tag{5}$$

where $<\dot{}>$ represents the long-term mean. If the errors in each data set are not correlated with each other or with the true soil moisture state, then the ratio of the expected sums on the right hand side of the above equations becomes one. The left hand side of each equation then provides the estimated calibration constants, $\hat{\lambda}$ and $\hat{\gamma}$. The calibrated data sets can then be combined to give:

$$< (\theta_A - \frac{\theta_L}{\hat{\lambda}}).(\theta_A - \frac{\theta_C}{\hat{\gamma}}) > = <\epsilon_A^2> - <\epsilon_A\epsilon_L> - <\epsilon_A\epsilon_C> + <\epsilon_L\epsilon_C> \tag{6}$$

$$< (\frac{\theta_L}{\hat{\lambda}} - \theta_A).(\frac{\theta_L}{\hat{\lambda}} - \frac{\theta_C}{\hat{\gamma}}) > = <\epsilon_L^2> - <\epsilon_A\epsilon_L> - <\epsilon_L\epsilon_C> + <\epsilon_A\epsilon_C> \tag{7}$$

$$< (\frac{\theta_C}{\hat{\gamma}} - \theta_A).(\frac{\theta_C}{\hat{\gamma}} - \frac{\theta_L}{\hat{\lambda}}) > = <\epsilon_C^2> - <\epsilon_A\epsilon_C> - <\epsilon_L\epsilon_C> + <\epsilon_A\epsilon_L> \tag{8}$$

If the errors are again assumed to be mutually uncorrelated, the last three terms in each equation become zero, and the square root of the left hand side gives the triple colocation estimate of the RMSE ($RMSE^{TC}(A)$). Here the

10

A in parentheses indicates that these estimates were obtained using θ_A as the reference data set. The RMSE can be converted to use another reference by multiplication with the appropriate calibration constant, or by repeating the calculation with an alternative calibration constant set to one.

Since the triple colocation was based on soil moisture anomalies from the seasonal cycle, the RMSE^{TC} represent only the errors in the soil moisture anomaly time series, or equivalently the anomalies from the mean seasonal cycle in the RMSE time series. That is, the mean seasonal cycle in the errors and the long-term mean error (bias) are not included in the RMSE^{TC}. Anomalies from the seasonal cycle were used following Miralles et al. (2010), who found anomalies to be more consistent with the triple colocation assumptions than raw soil moisture time series. The importance of using anomalies from the seasonal cycle will be confirmed in Section 4.2.

To date, most soil moisture triple colocation studies have excluded the calibration constants from equations 1-3, and instead rescaled the data sets with the ratio of their standard deviations prior to applying the above error model (e.g., Miralles et al. (2010); Dorigo et al. (2010); Parinussa et al. (2011a)). However, as discussed by Stoffelen (1998) and Yilmaz & Crow (2013), this results in biased calibration constants, which will then lead to biased RMSE estimates. In this study, standard deviation scaling would have resulted in many unphysically large RMSE estimates (exceeding the soil moisture anomaly time series standard deviation by up to 50%).

Triple colocation relies on all three data sets observing the same variable and having mutually uncorrelated errors. These assumptions have yet to be thoroughly tested for soil moisture, and will be checked in Section 4.2. For

soil moisture, a particular concern is the representativity differences between different soil moisture data sets. Appendix A examines how these representativity differences will contaminate the triple colocation RMSE estimates. In summary, where there are representativity differences between the data sets used, the triple colocation will favor the two most similarly defined data sets.

3.2. Error propagation through the retrieval models

For remotely sensed soil moisture retrievals, the soil moisture error associated with the uncertainty in the instrument measurements and the retrieval model parameters can be estimated by propagating these uncertainties through the retrieval model. For ASCAT, error estimates (Naeimi et al., 2009) are produced in parallel with the SDS data using Gaussian error propagation. For AMSR-E, Parinussa et al. (2011b) propagates the input errors through the LPRM model using the partial derivatives of the radiative transfer equation. These error propagation techniques generate an expected RMSE for each soil moisture observation, giving a time series of the expected RMSEs. At each grid cell, the square root of the mean of the squared error time series has been used as the error propagation RMSE estimate (RMSE^{EP}).

It is unclear whether the error propagation RMSE estimates better represent absolute soil moisture errors, or anomalies from the seasonal cycle of the error time series. The RMSE time series have a clear seasonal cycle associated with the seasonal cycle in the sensitivity of retrieval model parameters to various errors, indicating that at least some of the seasonal scale errors are included. However, error propagation cannot measure other aspects of the

12

longer-term errors. For example, errors in the retrieval model structure, such as in the separation of the vegetation and soil moisture signals, are a major source of seasonal to annual scale errors, and cannot be detected by error propagation. Nor does the error propagation include the long-term (length of the full data record) bias. In the absence of clear evidence either way, the error propagation results are assumed to relate to the anomalies from the seasonal cycle of the error time series, consistent with the ϵ defined for the triple colocation (in equations 1-3).

3.3. Confidence intervals of the triple colocation RMSE

For the error propagation, only one realization of the RMSE time series is available and so fRMSE^{EP} confidence intervals cannot be estimated. For triple colocation, by contrast, fRMSE^{TC} confidence intervals can be estimated using boot strapping, following Caires & Sterl (2003). Boot strapping is useful for estimating the standard error of statistics for which the population distribution is unknown or complex. The sample itself is used to approximate the population, and an empirical population distribution of the test statistic is constructed by resampling the original sample multiple times, with replacement to preserve the sample size. A test of the impact of the number of resamples on the estimated confidence intervals indicated stable results after approximately 500 resamples, and so a conservative count of 1000 resamples has been used, consistent with Wilks (2006). The required percentiles for the test statistic (the RMSE) have then been estimated directly from the boot-strapped distribution.

To estimate the confidence limits for the mean triple colocation RMSE over multiple grid cells, two different approaches have been used. When

the mean is estimated over contiguous spatial areas, such as over a land cover class in Section 4.1, all of the contiguous grid cells are conservatively assumed not to be independent. For a contiguous region covering n grid cells, the mean RMSE^{TC} has then been estimated in the usual way, using $\sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(RMSE_i)^2}$. The 90% confidence interval for the mean is then calculated separately for the upper (95th percentile minus median) and lower (median minus 5th percentile) intervals, and for both the mean of the contributing intervals is used. In contrast, for calculating the mean RMSE^{TC} and its confidence interval over the SCAN/SNOTEL sites in Section 4.2, the results at the individual SCAN/SNOTEL sites are assumed to be independent so long as they are sufficiently separated. Hence, the domain was divided into 5°x5° grid cells, and the SCAN/SNOTEL sites within each of these grid cells were assumed to lack independence, while the results for each 5° grid cell were assumed to be independent. Within each 5° grid cell, the mean RMSE ($\mathrm{RMSE}_{5°}$) and the width of the upper and lower confidence intervals were estimated as described above for contiguous areas. The domain-wide mean RMSE^{TC} over the m 5° grid cells containing SCAN/SNOTEL sites was then estimated as $\sqrt{\frac{1}{m}\Sigma_{i=1}^{m}(RMSE_{5°,i})^2}$. The width of the upper and lower confidence intervals for the mean were each then calculated as the mean of respective intervals for the m contributing 5° grid cells, divided by the square root of m.

3.4. Fractional RMSE (fRMSE)

As outlined in Section 1, when two soil moisture data sets are compared over large spatial domains the systematic differences between their central moments are usually removed by rescaling each data set to have statistics

14

consistent with a chosen reference data set (e.g., by using the calibration constants defined by equations 1-3). This has several consequences for the interpretation of the resulting RMSE. Most obviously, since the mean difference between the data sets has been removed, the resulting RMSE does not include the bias. Additionally, the RMSD estimated by comparing two fields, A and B, with equivalent means is a function of the standard deviation of each field (σ_A and σ_B) and the correlation (R) between them:

$$RMSD(A, B) = \sqrt{< (A - B)^2 >} = \sqrt{\sigma_A^2 + \sigma_B^2 - 2R\sigma_A\sigma_B} \qquad (9)$$

A RMSE based on rescaled soil moisture data then depends on the standard deviation of the reference data set, and the correlation with the validating truth: note that the signal of the accuracy of the data is derived from the latter.

To highlight the dependence of the RMSE on the reference standard deviation, Figures 2 compares the time series standard deviation of the soil moisture anomalies for ASCAT, AMSR-E, and Catchment, to the triple colocation ASCAT error estimates, represented using each of these data sets as the reference. There are considerable differences in the σ for each data set, with the mean varying between 14% SDS (or $0.07m^3m^{-3}$ assuming a porosity of $0.5m^3m^{-3}$) for ASCAT, $0.07m^3m^{-3}$ for AMSR-E , and $0.03m^3m^{-3}$ for Catchment. The spatial patterns described by each are also very different. The absolute values of the ASCAT RMSE^{TC} also differ depending on which data set is used as the reference, with the mean RMSE varying between 9% SDS ($0.05m^3m^{-3}$), $0.04m^3m^{-3}$, and $0.03m^3m^{-3}$ when ASCAT, AMSR-E, and Catchment are used, respectively. The spatial patterns in the ASCAT

15

fRMSE^{TC} also differ depending on which reference was used, and in each case there are clear features of the reference σ in the RMSE maps.

While at individual locations the ratio of the RMSE between different data sets does not depend on the selected reference data set, the ratio and even the ranking of the domain-averaged RMSE does depend on the reference. For example, Table 1 lists the mean RMSE^{TC} across the domain for each data set, presented using each data set as the reference. With AS-CAT as the reference, the mean RMSE^{TC} for AMSR-E and Catchment are both more than 50% higher than the ASCAT RMSE^{TC}, and Catchment has the highest mean RMSE^{TC}. However with AMSR-E as the reference, the difference between the ASCAT RMSE^{TC} and the AMSR-E and Catchment RMSE^{TC} is reduced. Of more consequence, with Catchment as the reference the ranking of the RMSE^{TC} changes, and AMSR-E has the highest mean RMSE^{TC}, although the differences are now small.

This study investigates the spatial variability in remotely sensed soil moisture RMSEs. If the RMSE were presented in soil moisture units (relative to a reference data set), the spatial variability in each would be very similar, due to the common signal of the reference standard deviation. Hence, the fractional RMSE (fRMSE) is introduced for examining the RMSE:

$$fRMSE_X = RMSE_X(X)/\sigma_X \qquad (10)$$

The fRMSE is obtained by presenting the RMSE for data set X ($RMSE_X$) using itself as the reference ($RMSE_X(X)$), and then dividing this by the standard deviation of X (σ_X). With the signal of the standard deviation effectively removed from equation 9, the fRMSE statistic is consistent with

16

the common use of correlation statistics to evaluate soil moisture (e.g., Reichle et al. (2007); Scipal et al. (2008a); de Jeu et al. (2008); Parinussa et al. (2011a); Draper et al. (2012)).

The fRMSE has several advantages over presenting the RMSE using an arbitrary reference. It is self contained, and has a well defined range between 0 (perfect estimates) and 1 (noise, with no signal of the truth), with values greater than $1/\sqrt{2}$ ($\sim$ 0.7) indicating an error variance that exceeds the variance of the true time series. Users of a specific data set need only multiply the fRMSE by the standard deviation of that data set to obtain a RMSE in soil moisture units, rather than requiring access to the arbitrary reference data set. The fRMSE also allows more flexibility in comparing different error estimates, since it does not rely on being able to convert all error estimates to a common reference (which will allow the inter-comparison of the triple colocation RMSE obtained with different data triplets in Section 4.2).

A potential disadvantage of the fRMSE, however, is that by converting the RMSE for each data set to its own climatology, the ratio of the RMSEs for two data sets at a given location is not conserved, although the ranking between them is. While the RMSE ratio could be preserved by converting each RMSE to a common reference (say data set Y), and then dividing by the standard deviation of that reference, this leads to the inclusion of the errors in the reference data set in the statistic (since $RMSE_X(Y)/\sigma(Y) = \sqrt{\epsilon_X^2(Y)/(\sigma_T(Y) + \epsilon_Y(Y))^2}$, where σ_T is the standard deviation of the true soil moisture). The result is no longer self contained and can generate unexpected results.

For the remainder of this paper, the RMSE estimates are presented using

17

the fRMSE. While the RMSE^{EP} are not based on rescaled data sets, they implicitly reflect the climatology of the data set to which they apply, and a fRMSE has been calculated by dividing the error propagation RMSE by the standard deviation of the soil moisture anomaly time series. Note that reporting the errors as a fRMSE magnifies differences between the error estimates. Soil moisture, and consequently the error in soil moisture, is usually reported with a precision of 0.01 $m^3 m^{-3}$. Based on the (spatial) mean standard deviation of $0.08 m^3 m^{-3}$ for AMSR-E in Figure 2b, a RMSE of 0.01 $m^3 m^{-3}$ is equivalent to a fRMSE of 0.1.

4. Results

4.1. fRMSE over the domain

As discussed above, in Section 3.4, the error estimates are presented here in terms of the fRMSE (equation 10). Figure 3 shows maps of the ASCAT and AMSR-E fRMSE calculated from triple colocation (fRMSE^{TC}) and error propagation (fRMSE^{EP}). Spatially, the fRMSE^{EP} is smoother than the fRMSE^{TC}, likely due to noise in the triple colocation estimates, and the dependence of the error propagation on coarsely defined model parameters. The most obvious feature of the four maps is that the AMSR-E fRMSE^{EP} are unphysically large, with values consistently above two (i.e., a RMSE more than double the time series standard deviation). In contrast, the ASCAT fRMSE^{EP} are within the expected range, and tend to be slightly lower than the fRMSE^{TC}.

The error propagation methods were developed with a focus on predicting the temporal and spatial variability in the RMSE of a specific data set. The

18

magnitude of the fRMSE^{EP} depends on the magnitude of the uncertainties specified for the retrieval model input and parameters, however the uncertainties in the retrieval model parameters are not well understood at scales relevant to remote sensing, and so are specified somewhat arbitrarily. Hence, little weight should be placed on the magnitude of the error propagation output, and the unrealistic absolute values obtained for the AMSR-E fRMSE^{EP} are not surprising.

In terms of the spatial variability in the RMSE, Figure 3 shows a general agreement in the broad patterns described by the fRMSE^{TC} and fRMSE^{EP} for each data set, all of which are consistent with expectations. All four maps show the expected increase in fRMSE toward the more vegetated east of the domain, although for the ASCAT fRMSE^{EP}(Figure 3b) the eastward increase is weaker than for the other maps.

The ASCAT and AMSR-E errors can only be compared based on the triple colocation results, given the uncertain magnitude of the error propagation output. In Figure 3 the ASCAT and AMSR-E fRMSE^{TC} appear to be very similar across the domain, except over the croplands in the Mid-West of the US where the ASCAT fRMSE^{TC} are much lower than the AMSR-E fRMSE^{TC}, and immediately to the east of the Rocky Mountains where the AMSR-E fRMSE^{TC} are lower than the ASCAT fRMSE^{TC}.

To establish whether these differences in the fRMSE^{TC} are significant, Figure 4 shows the width of the 90% confidence interval for the fRMSE^{TC} estimates, while Figure 5 indicates regions where the (one sided) differences between the ASCAT and AMSR-E fRMSE^{TC} are significant (at 5%). In Figure 4 the confidence intervals in the Mid-West of the US and in the northeast

19

of the domain exceed 0.5 for ASCAT (that is, the 90% confidence interval spans more than 50% of the standard deviation), and also for AMSR-E over a subregion of the Mid-West. Over the rest of the domain, the confidence intervals are between 0.1-0.3, with a tendency for the ASCAT and AMSR-E intervals to offset each other (i.e., one is reasonably high where the other is reasonably low). Despite the large uncertainties, Figure 5 shows that the lower ASCAT fRMSE^{TC} (compared to the AMSR-E fRMSE^{TC}) is significant across much of the Mid-West. Figure 5 also clearly shows the tendency for the AMSR-E fRMSE^{TC} to be lower than the ASCAT fRMSE^{TC} in the west of the plotted domain, while the ASCAT fRMSE^{TC} tends to be lower in the east of the domain. This east-west difference has been reported by previous triple colocation studies comparing passive and active microwave soil moisture retrievals (Scipal et al., 2008b; Dorigo et al., 2010).

Figure 6 shows the mean fRMSE by land cover, for each land cover class with at least 100 grid cells in Figure 3, along with 90% confidence intervals for the fRMSE^{TC} estimates (Section 3.3). At the microwave frequencies observed by ASCAT and AMSR-E, interference from vegetation is a major source of error in soil moisture retrievals. Hence, the mean LAI over each land cover class is also included in Figure 6 to provide a proxy for the vegetation interference. As expected, there is a general pattern across the land cover classes of increasing mean fRMSE with increasing LAI.

For ASCAT, the agreement between the fRMSE^{TC} and fRMSE^{EP} is in general very good, except that over the five most densely vegetated categories (woody savanna, ever-green needleleaf, deciduous broadleaf, forest, and croplands/natural cover), the mean fRMSE^{EP} is lower than the mean fRMSE^{TC},

and is even below the $1/\sqrt{2}$ line (hence the relatively low fRMSE^{EP} in the east in Figure 3b). The analysis was repeated without discarding ASCAT data with high error propagation errors (see Section 2.1), and this separate analysis (not shown) confirms that this quality control step was not the cause of the low ASCAT fRMSE^{EP} in the east of the domain.

For AMSR-E, the relationship between the fRMSE^{TC} and fRMSE^{EP} is less consistent. For triple colocation, the variability between the mean AMSR-E fRMSE^{TC} for each land cover class directly reflects the variability in the mean LAI. However, for error propagation the mean AMSR-E fRMSE^{EP} are grouped into two bins: the three land cover classes with the lowest LAI were effectively assigned similar and relatively low mean fRMSE^{EP}, while the remaining five land cover classes were effectively assigned similar relatively high mean fRMSE^{EP}. This tendency to assign the errors to one of two modes is also evident in the lack of graduated colors in Figure 3d.

The ASCAT SDS retrieval model includes a semiempirical vegetation correction that removes the climatological seasonal cycle of the vegetation signal from the observed backscatter. This correction is thought to be reasonably effective over moderate vegetation conditions, and in theory moderate vegetation will be less detrimental to the accuracy of the ASCAT SDS than to the accuracy of the AMSR-E soil moisture. This is consistent with Figure 6 which shows that the relationship between the mean LAI and the mean fRMSE is much stronger for AMSR-E than for ASCAT. For ASCAT, factors other than vegetation can also contribute to the errors in the soil moisture retrievals. For example, open shrubs have the lowest mean LAI in Figure 6, yet the mean ASCAT fRMSE estimates over the open shrubs are very high

21

(close to $1/\sqrt{2}$) for both methods. The open shrub grid cells are in the arid southwest of the domain (Figure 1), where the coverage in Figure 3 is poor (229 out of nearly 7000 grid cells). The poor performance of the SDS in arid environments is an established (although not well understood) limitation of the ASCAT change-detection model (Wagner et al., 2003). Additionally, over grasslands and croplands both fRMSE^{TC} and fRMSE^{EP} indicate similar AS-CAT fRMSE, despite the croplands having much higher LAI. This reasons for this difference are not known.

In terms of the relative performance of the ASCAT and AMSR-E soil moisture, while there are some differences in their mean fRMSE^{TC} over different land cover classes in Figure 6, none of these differences are significant. As was noted above, over the croplands the fRMSE^{TC} for ASCAT is quite low, and much lower than the AMSR-E fRMSE^{TC}. While this result is not statistically significant, the enhanced ASCAT skill is supported by the mean fRMSE^{EP} also being relatively low for croplands. For the three least vegetated land cover classes (open shrubs, grassland, and crops), at least one of the ASCAT and AMSR-E fRMSE^{TC} are significantly less than $1/\sqrt{2}$, indicating an ability to accurately detect soil moisture anomalies. These three land cover classes constitute 63% of the domain with fRMSE values in Figure 3. Over the five more densely vegetated land cover classes, the fRMSE^{TC} is generally above or close to $1/\sqrt{2}$, indicating poor skill with errors exceeding the true soil moisture variability, and confirming the usual practice of screening the ASCAT and AMSR-E data at these locations (Section 2.1).

Finally, in Figures 3 and 4 the plotted coverage is less than that of the quality controlled data in Figure 1, due to the triple colocation having pro-

duced negative mean square errors at 17% of the locations that passed the quality control procedures described in Section 2.1. These locations are generally adjacent to regions where data have been screened by the quality control, or are barren/open shrub grid cells in the arid southwest where the ASCAT errors are very large. This suggests that triple colocation may require a minimum skill from all three data sets, which is consistent with the assumption that all three data sets observe the same variable.

4.2. Dependence of $RMSE^{TC}$ on the data triplet

Triple colocation is based on the assumption that all three data sets observe the same variable and have mutually uncorrelated errors. However, representativity differences (extending beyond systematic differences in the central moments) are inevitable between global soil moisture data sets, and will lead to violations of these assumptions. The consequences of this are investigated here, by testing how the fRMSE^{TC} estimates differ when different data sets are used in the triple colocation data triplet. In situ soil moisture from the 57 SCAN/SNOTEL sites have been used as an alternate data set.

Figure 7 shows the mean fRMSE^{TC} averaged over the SCAN/SNOTEL sites, calculated with different data triplets selected from the ASCAT, AMSR-E, Catchment, and SCAN/SNOTEL soil moisture anomalies. For a given data set the differences between the fRMSE^{TC} estimates are small when different data triplets are used, although some of these differences are statistically significant. The maximum fRMSE^{TC} difference for a given data set due to the use of different data triplets is ~ 0.1, much smaller than the $0.2 - 0.5$ differences reported from Figure 6, and close to the precision of soil moisture data.

The dependence of the fRMSE^{TC} estimates on the data triplet used in the triple colocation is also consistent with the expected representativity differences between the four data sets. For ASCAT and AMSR-E, the fRMSE^{TC} estimates are significantly lower when both ASCAT and AMSR-E are included in the triplet (left two triplets in Figure 7) than when only one of the remote sensors is included. Likewise, for both Catchment and the SCAN/SNOTEL data, the fRMSE^{TC} is lower, sometimes significantly, when only one of the remote sensors is included (right two triplets) than when both are included in the triplet. This tendency to favor the remote sensors when both are included in the data triplet, and to favor the other two data sets when only one remote sensor is included, suggests a representativity difference between the two remote sensors on one hand, and the Catchment and SCAN/SNOTEL data on the other hand (see Appendix A).

If the results from Figure 7 are generalized across the domain presented in Section 4.1, then the representativity differences reported above will have had little impact on the results reported here, most obviously because the fRMSE^{TC} differences reported in Section 4.1 are much larger than the < 0.1 differences obtained here. Also, the representativity differences discussed above will not influence the ASCAT and AMSR-E fRMSE^{TC} in Section 4.1, so long as the fRMSE^{TC} is interpreted as being relative to a soil moisture truth defined to resolve the same features as the remote sensors. However, the Catchment fRMSE^{TC} calculated in Section 4.1 (but not shown) will have included a small representativity error (of ~ 0.15), associated with the representativity differences between the modeled soil moisture and the soil moisture truth defined by the remote sensors.

The above result is dependent on the triple colocation having been based on soil moisture anomalies from the mean seasonal cycle, rather than on anomalies from a single long-term mean or on raw data. Repeating Figure 7 with the triple colocation based on anomalies from the long-term mean over the full data record (as opposed to anomalies from a seasonally varying climatology) resulted in fRMSE^{TC} for a given data set that consistently differed by more than 0.5 depending on which data triplet was used. This confirms the previous statement of Miralles et al. (2010) that the triple colocation assumptions are better satisfied for soil moisture by using anomalies defined as deviations from the seasonal cycle.

Finally, Figure 7 also highlights that RMSE estimates from triple colocation are far more accurate than a RMSD based on comparison to only one other data set. The latter method is most often based on observations from individual in situ soil moisture sensors, yet in Figure 7 the SCAN/SNOTEL RMSE (when estimating coarse scale soil moisture) are as large as the AS-CAT and AMSR-E RMSEs. Hence, the RMSD between either remote sensor and the SCAN/SNOTEL data will significantly over-estimate the RMSE in the remotely sensed data. To address this, Miralles et al. (2010) estimate an in situ-based RMSE for data set X (RMSE$_X^{IS}$), by correcting the RMSD between X and the in situ data (RMSD$_X^{IS}$) with a triple colocation estimate of the RMSE of the in situ data (RMSE$_I^{TC}$), using RMSE$_X^{IS} \approx \sqrt{(RMSD_X^{IS})^2 - (RMSE_I^{TC})^2}$. While this method is useful for highlighting the contribution of the in situ errors to the RMSD$_I^X$, it is equivalent to simply estimating the RMSE for θ_X using triple colocation. With reference to the triple colocation equations (equations 6- 8), the corrected RMSD can be

written:

$$RMSE_X^{IS} \approx \sqrt{(RMSD_X^{IS})^2 - (RMSE_I^{TC})^2} \tag{11}$$

$$= \sqrt{< (\theta_X - \theta_I)^2 > - < (\theta_X - \theta_I)(\theta_X - \theta_Y) >} \tag{12}$$

$$= \sqrt{< (\theta_X - \theta_I)(\theta_Y - \theta_I) >} \tag{13}$$

$$= RMSE_X^{TC} \tag{14}$$

The calibration constants have been neglected above for clarity. However, this results does not change if the calibration constants are included, except for the introduction of an inconsistency between the calibration constants used in the $RMSE^{TC}$ and $RMSD^{IS}$ calculations, since the latter is based on only two data sets and will be biased (Stoffelen, 1998).

5. Discussion

The root mean square errors in soil moisture anomaly time series from AMSR-E and ASCAT have been estimated across a continental scale domain in North America using two methods: (i) triple colocation with Catchment model soil moisture as the third data set, and (ii) error propagation through the respective soil moisture retrieval models. These methods have been investigated to determine their utility for evaluating remotely sensed soil moisture over large domains, including for the specification of observation error variances needed for data assimilation.

In the absence of an consensus soil moisture climatology over large domains, presenting a RMSE in soil moisture units requires that it be specified

relative to a selected reference data set. The absolute value and spatial patterns of a RMSE in soil moisture units will then depend on the selected reference data set, and specifically on its standard deviation (Figure 2). In this study this dependence is reduced by presenting the RMSE for each data set as the fraction of the standard deviation of that data set (fRMSE).

Comparing the triple colocation and error propagation fRMSE over the continental scale domain indicates that both methods can accurately detect the large scale variability in soil moisture errors. In Figure 3 the regions with relatively high and low $fRMSE^{TC}$ and $fRMSE^{EP}$ agree very well, and in Figure 6 the variability in the mean $fRMSE^{TC}$ and $fRMSE^{EP}$ over each land cover class also agrees with expectations.

The error propagation methods are designed to determine the spatial and temporal variation of the errors within a given data set, and while not used here, the unique ability to produce time series of the errors may be the most useful feature of error propagation. The magnitude of the RMSE output from the error propagation depends on the magnitude of the uncertainties specified for the retrieval model parameters, and these uncertainties are not well understood. Hence, the magnitude of the error propagation output is not necessarily expected to be correct. In this study, the ASCAT $fRMSE^{EP}$ appear to be approximately correct, while the AMSR-E $fRMSE^{EP}$ were unrealistically large. For AMSR-E, the LPRM model parameter uncertainties used in the error propagation were conservatively estimated to be quite large (Parinussa et al., 2011b), and reducing the uncertainties specified for example in the roughness or single scattering albedo would reduce the error propagation output to more realistic values. More generally, to have any confidence

in the magnitude of the error propagation output will require an improved understanding of the uncertainty in the retrieval model parameters. This could be achieved during the calibration of the retrieval model parameters, by using methods that generate both parameter values and the uncertainty in those parameters (e.g., De Lannoy et al. (2013)).

The errors in the $fRMSE^{EP}$ themselves can also be useful for identifying shortcomings of the retrieval models. For AMSR-E LPRM, the $fRMSE^{EP}$ were relatively low over sparsely vegetated regions, and relatively high over densely vegetated regions, with little graduation between these two modes (Figures 3 and 6). In contrast, the AMSR-E $fRMSE^{TC}$ gradually increased with increasing vegetation density, resulting in the expected strong correlation with LAI in Figure 6. This error propagation behavior can be traced to a limitation of the tau-omega model used by the LPRM. The tau-omega model parametrizes the attenuation of the soil moisture signal by vegetation using an exponential function of vegetation optical depth (equation 2, Parinussa et al. (2011b)), resulting in an exponential increase in the tau-omega error propagation output with increasing vegetation optical depth (Figure 2, Parinussa et al. (2011b)). The sudden and steep increase in the LPRM $fRMSE^{EP}$ with increasing vegetation in Figures 3 and 6 of this study suggests that the tau-omega model is over-estimating this non-linear sensitivity to vegetation attenuation. This suggests a potential to improve the LPRM, and other retrieval algorithms using the tau-omega model.

Likewise, for ASCAT the $fRMSE^{TC}$ and $fRMSE^{EP}$ disagree over the eastern US, where the $fRMSE^{TC}$ is much higher than the $fRMSE^{EP}$ (by > 0.2). The cause of this discrepancy is unknown, however the combination of higher

fRMSE^{TC} and lower fRMSE^{EP} suggests errors in the ASCAT SDS estimates in this region associated with a physical process that is not properly accounted for in the SDS retrieval model.

For triple colocation there is no evidence that the absolute values of the fRMSE^{TC} are not accurate, with the caveat that the errors are relative to the soil moisture anomaly truth defined by the three data sets used. However, the dependence of the fRMSE^{TC} on the triplet of data sets used was tested at 57 SCAN/SNOTEL sites, revealing surprising robustness to representativity differences between the data sets used, particularly given the substantial representativity differences expected between the point-based in situ and coarse scale soil moisture estimates. There were only small systematic differences (below the typical precision of soil moisture) between the fRMSE^{TC} for a given data set, depending on the triplet of data sets used. While the representativity differences between the various soil moisture anomaly data sets were of little consequence here, caution is still recommended when selecting the data sets to be used in soil moisture triple colocation. Additionally, this result requires that the triple colocation be based on soil moisture anomalies from the seasonal mean.

In contrast to error propagation, triple colocation provides a consistent method for estimating the RMSE of different remotely sensed soil moisture data sets. In Section 4.1, the triple colocation results showed that in general ASCAT and AMSR-E have similar accuracy over a range of land cover conditions, although the AMSR-E errors have a stronger dependence on vegetation cover. The exceptions, both of which have previously been reported in the literature (Wagner et al., 2003; Scipal et al., 2008b; Dorigo et al.,

2010), are the high ASCAT fRMSE^{TC} in arid regions and the low ASCAT fRMSE^{TC} over the croplands in the Mid-West of the US. Note that slightly better AMSR-E accuracy is expected in most other regions of the globe where C-band observations can be used, since only X-band data were used here due to radio frequency interference contaminating the C-band observations over North America.

6. Conclusions and Recommendations

The above findings have implications for the evaluation of remotely sensed soil moisture data. Currently, novel soil moisture data sets retrieved from remote sensors are validated against predetermined target accuracies specified in soil moisture units (relative to the true soil moisture). This validation is typically based on a limited number of well-observed pixels. In Figures 2 and 3 there is substantial spatial variation in the soil moisture RMSE and fRMSE, highlighting that an evaluation based on a limited number of locations will not necessarily be representative of larger domain. Hence, the above validation approach should be complemented with distributed methods that can estimate the soil moisture RMSE globally.

Both triple colocation and error propagation can accurately detect regions of relatively high and low fRMSE. While the definition of the RMSE produced by triple colocation (unbiased RMSE of anomalies from the seasonal cycle) and error propagation (errors associated with model input and parameters only) differs from that currently defined by remote sensing target accuracies, these methods could still be useful for identifying regions where the accuracy from well-observed pixels can be confidently extrapolated, and where the ac-

curacy might differ (particularly where it is unexpectedly reduced). For most applications, triple colocation is more useful, since in addition to predicting the spatial variability in the errors, it can accurately detect the magnitude of the RMSE.

However, it is unclear how current target accuracies (in soil moisture units) should be interpreted in a truly global evaluation. Most obviously, without knowledge of the true global soil moisture climatology, an assessment in soil moisture units requires selecting a reference, and this arbitrary decision determines the magnitude of the resulting errors. Also, as pointed out by Entekhabi et al. (2010), a uniform (or maximum) RMSE in soil moisture units is difficult to interpret over a large domain, since the same value can indicate very good skill in a region of high variability and be trivially satisfied in a region with low variability. Alternatively, interpreting a target accuracy as the mean RMSE over a large domain is also problematic, since the choice of reference data set affects the relative performance of different data sets (e.g., Table 1). Hence, extending the evaluation (or specification of target accuracies) of remotely sensed soil moisture to a near-global domain will require the use of alternative metrics, such as the fRMSE.

Finally, for data assimilation observation error variances are often specified to be constant across the assimilation domain, in the soil moisture units of either the model or the observations (in the latter case, the error variance is then scaled to be consistent with the model climatology in the same manner as the observations are). Again, the specification of a constant soil moisture RMSE over a large domain is not sensible, and at a minimum it would be better to specify a constant fRMSE. An even better solution would

be to introduce the spatial variability in the fRMSE, by using mean values for each land cover class from either the triple colocation or error propagation methods. Ideally, the temporal variability from the error propagation could also be used, after appropriate rescaling to correct the absolute values.

7. Acknowledgements

Appendix A. Dependence of triple colocation on features resolved by the data triplet

For soil moisture data sets, representativity differences between different data sets are common (beyond differences in the central moments), for example due to differences in the spatial or temporal support or in the soil moisture processes resolved by different data sets. As outlined by Stoffelen (1998) the truth defined by triple colocation, against which the RMSE are estimated, includes only the features resolved by all three data sets. It is demonstrated here that where there are representativity differences between the three data sets, in that they do not all resolve the same features, the triple colocation RMSE will favor the two most similar data sets.

In the instance where one data set differs from the other two data sets in that it resolves additional variability that is not present in the other two data sets (for example variability at a finer spatial scale), the additional features

32

will be attributed to errors in that data set, increasing its triple colocation error estimate. However, in the instance where one data set differs from the other two data sets in that it lacks a source of additional variability that is present in the other two data sets, the triple colocation RMSE estimates still favor the two more similar data sets. For example, consider the triple colocation of data sets X1, X2, and Y, where X1 and X2 both resolve an additional source of variability not resolved by Y. This additional variability is assigned to representativity errors in X1 and X2, resulting in a non-negligible correlation between the 'errors' for X1 and X2. If this is the only non-negligible covariance between the errors, then the triple colocation error estimates obtained from equations 6-8 are:

$$RMSE^{TC}(X1) = \sqrt{< \epsilon_{X1}^2 > - < \epsilon_{X1}.\epsilon_{X2} >^2} \tag{15}$$

$$RMSE^{TC}(X2) = \sqrt{< \epsilon_{X2}^2 > - < \epsilon_{X1}.\epsilon_{X2} >^2} \tag{16}$$

$$RMSE^{TC}(Y) = \sqrt{< \epsilon_{Y}^2 > + < \epsilon_{X1}.\epsilon_{X2} >^2} \tag{17}$$

The additional features resolved by X1 and X2 are subtracted from the X1 and X2 RMSE estimates, and added to the RMSE estimate for Y; the triple colocation has effectively produced RMSE relative to a truth defined to include the additional features resolved by X1 and X2.

In the above example the correlated errors between X1 and X2 will also affect the calibration constants in equations 4-5. However, this effect will be secondary to that described above since in equations 4 and 5 the error covariances appear next to the variance of the truth, while in the equations above the error covariances appear next to the error variances, against which

they will constitute a much larger fraction.

References

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa-Saldaña, J., & Anderson, C. (2007). Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophys. Res. Lett.*, *34*, L20401.

Caires, S., & Sterl, A. (2003). Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.*, *16*, 3098.

dall'Amico, J., Schlenz, F., Loew, A., & Mauser, W. (2012). First results of SMOS soil moisture validation in the Upper Danube catchment. *IEEE Trans. Geosci. Remote Sens.*, *50*, 1507–1516.

De Lannoy, G., Reichle, R., & V.Pauwels (2013). Global calibration of the GEOS-5 L-band microwave radiative transfer model over land using SMOS observations. *J. Hydrometeor.*, *in press*.

Dharssi, I., Bovis, K., Macpherson, B., & Jones, C. (2011). Operational assimilation of ASCAT surface soil wetness at the Met Office. *Hydrol. Earth Syst. Sci.*, *15*, 2729–2746.

Dorigo, W., Scipal, K., Parinussa, R., Liu, Y., Wagner, W., de Jeu, R., & Naeimi, V. (2010). Error characterisation of global active and passive microwave soil moisture datasets. *Hydrol. Earth Syst. Sc.*, *14*, 2605–2616.

Draper, C., Reichle, R., De Lannoy, G., & Liu, Q. (2012). Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.*, *39*, L04401.

Draper, C., Walker, J., Steinle, P., de Jeu R., & Holmes, T. (2009). An evaluation of AMSR-E derived soil moisture over Australia. *Remote Sens. Environ.*, *113*, 703–710.

Entekhabi, D., Njoku, E., O'Neill, P., Kellogg, K.,Crow, W., Edelstein, W., Entin, J., Goodman, S., Jackson, T., Johnson, J., Kimball, J., Piepmeier, J., Koster, R., Martin, N., McDonald, K., Moghaddam, M., Moran, S., Reichle, R., Shi, J., Spencer, M., Thurman, S., Tsang, L., & Zyl, J., (2010). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE*, *98*, 704–716.

Entekhabi, D., Reichle, R., Koster, R., & Crow, W. (2010). Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeor.*, *11*, 832–840.

Friedl, M., McIver, D., Hodges, J., Zhang, X., Muchoney, D., Strahler, A., Woodcock, C., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., & Schaaf, C. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.*, *33*, 287–302.

Gherboudj, I., Magagi, R., Goita, K., Berg, A., Toth, B., & Walker, A. (2012). Validation of SMOS data over agricultural and boreal forest areas in Canada. *IEEE Trans. Geosci. Remote Sens.*, *50*, 1623–1635.

Jackson, T., Bindlish, R., Cosh, M., Tianjie, Z., Starks, P., Bosch, D., Seyfried, M., Moran, M., & Goodrich, D. (2012). Validation of Soil Moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the U.S. *IEEE Trans. Geosci. Remote Sens.*, *50*, 1530–1543.

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

de Jeu, R., Wagner, W., Holmes, T., Dolman, A., van de Giesen, N., & Friesen, J. (2008). Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surv. Geophys.*, *29*, 399–420.

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

Koster, R., Guo, Z., Yang, R., Dirmeyer, P., Mitchell, K., & Puma, M. (2009). On the nature of soil moisture in land surface models. *J. Climate*, *22*, 4322–4335.

Koster, R., Suarez, M., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model. *J. Geophys. Res.*, *105*, 24809–24822.

Mahfouf, J.-F. (2010). Assimilation of satellite derived soil moisture from ASCAT in a limited area NWP model. *Q. J. Roy. Meteorol. Soc.*, *136*, 784–798.

Miralles, D., Crow, W., & Cosh, M. (2010). Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *J. Hydrometeor.*, *11*, 14231429.

Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., & Wagner, W. (2009). An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Trans. Geosci. Remote Sens.* , *47*, 1999–2013.

Njoku, E. (1999). Retrieval of land surface parameters using passive microwave measurements at 6-18 GHz. *IEEE Trans. Geosci. Remote Sens.* , *37*, 79–93.

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

Parinussa, R., Holmes, T., & Crow, W. (2011a). The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrol. Earth Syst. Sci.*, *3135-3151*.

Parinussa, R., Meesters, A., Liu, Y., Dorigo, W., Wagner, W., & de Jeu, R. (2011b). Error estimates for near-real-time satellite soil moisture as derived from the Land Parameter Retrieval Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*, 779–783.

Reichle, R., & Koster, R. (2004). Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.*, *31*, L19501.

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

Reichle, R., Koster, R., Liu, P., Mahanama, S., Njoku, E., & Owe, M. (2007). Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res.* , *112*, D09108.

Rienecker, M., Suarez, M., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M., Schubert, S., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., & coauthors (2011). MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, *24*, 3624–3648.

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

Scipal, K., Drusch, M., & Wagner, W. (2008a). Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Adv. Water Resour.*, *31*, 1101–1112.

Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., & Wagner, W. (2008b). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophys. Res. Lett.*, *35*, L24403.

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation,. *J. Geophys. Res.*, *7755-7766*.

Vinnikov, K., Robock, A., Qiu, S., & Entin, J. (1999). Optimal design of

surface networks for observation of soil moisture. *J. Geophys. Res.*, *104*, 19743–19749.

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ.*, *70*, 191–207.

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

Wilks, D. (2006). *Statistical Methods in the Atmospheric Sciences*. Elsevier.

Yilmaz, M., & Crow, W. (2013). The optimality of potential rescaling approaches in land data assimilation. *J. Hydrometeor.*, *in press*.
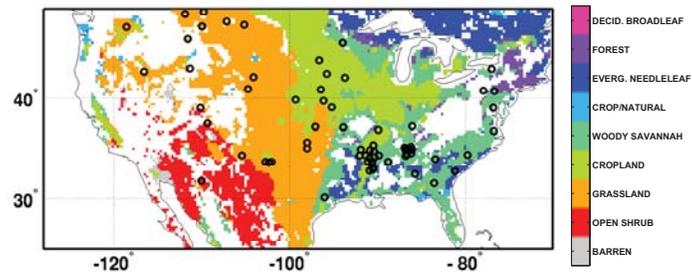
Figure 1: MODIS land cover classes (Friedl et al., 2002), plotted where remotely sensed soil moisture data are available after quality control. Black circles indicate the location of the SCAN/SNOTEL sites used in Section 4.2.
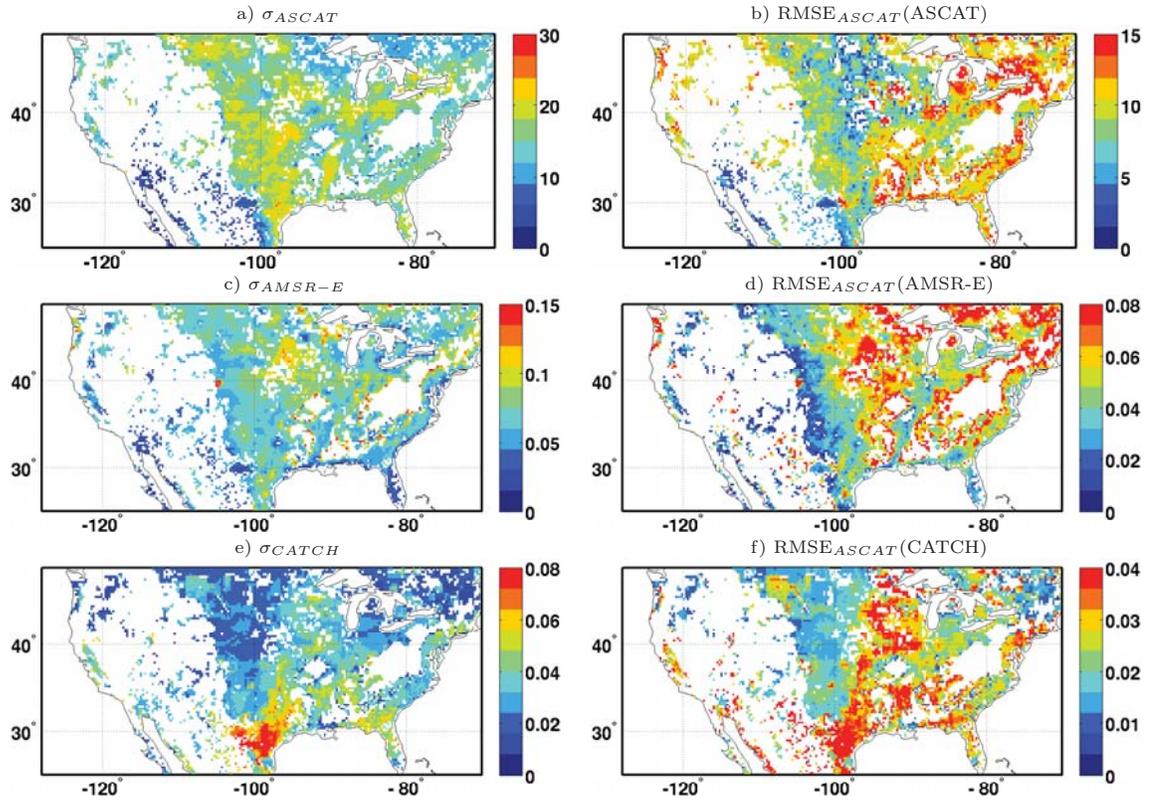
Figure 2: Maps of (left) the standard deviation in the soil moisture anomaly time series from a) ASCAT (%), c) AMSR-E ($m^3 m^{-3}$), and e) Catchment ($m^3 m^{-3}$), and (right) RMSE estimates for ASCAT from triple colocation, presented using b) ASCAT (%), d) AMSR-E ($m^3 m^{-3}$), and f) Catchment ($m^3 m^{-3}$) as the reference data set. Both are plotted only where triple colocation results are available.
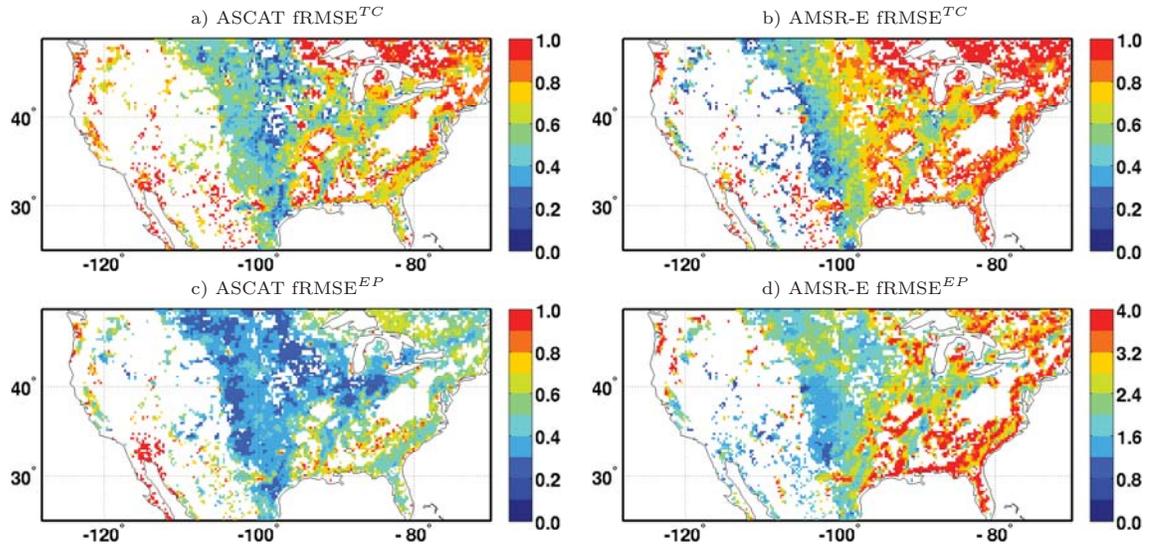
Figure 3: fRMSE of (left) ASCAT and (right) AMSR-E, fRMSE estimated using the (upper) triple colocation and (lower) error propagation methods, plotted only where triple colocation results are available. Note the different color scale for the AMSR-E fRMSE^{EP} in subfigure d.
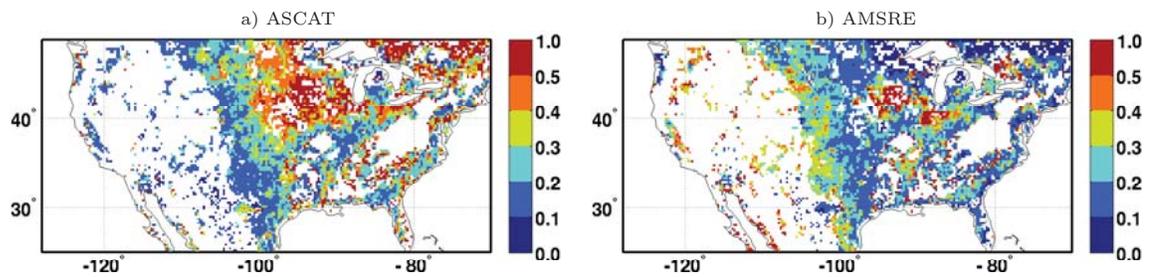

Figure 4: Width of the 90% confidence interval for the fRMSE^{TC} of the a) ASCAT and b) AMSR-E soil moisture anomalies, plotted only where triple colocation results are available.

43
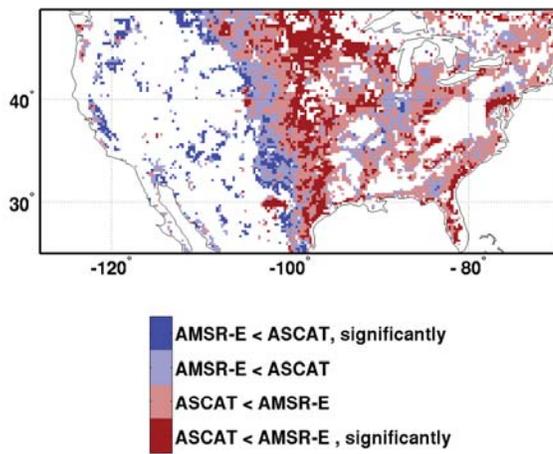
Figure 5: Comparison of ASCAT and AMSR-E fRMSE^{TC}. Blue (red) indicates AMSR-E fRMSE^{TC} less (more) than ASCAT fRMSE^{TC}, with darker shades indicating a significant difference at 5%.
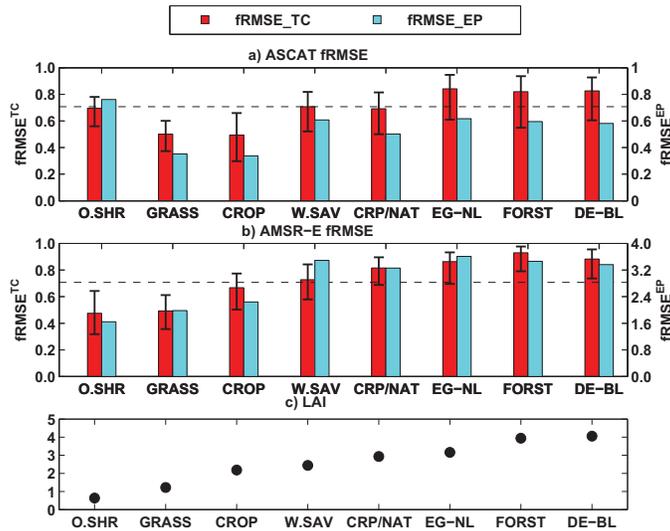
Figure 6: Mean by land cover class for the fRMSE^{TC} and fRMSE^{EP} of a) ASCAT and b) AMSR-E, and for c) LAI. For fRMSE^{TC} the 90% confidence interval is included (uncertainty estimates are not available for fRMSE^{EP}). Note the different y-axis (on right) for the AMSR-E fRMSE^{EP}. The dashed line indicates a fRMSE of $1/\sqrt{2}$, above which the signal to noise ratio is below one.
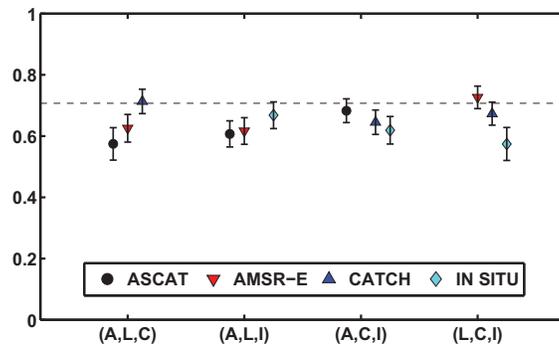
Figure 7: Mean fRMSE and its 90% confidence interval estimated across the SCAN/SNOTEL sites, using triple colocation based on different combinations of three of (A) ASCAT, (L) AMSR-E, (C), Catchment, and (I) in situ data sets. The data triplet is indicated in the x-axis labels, while the plotted symbol/color indicates the data set for which the error is estimated. The dashed line indicates a fRMSE of $1/\sqrt{2}$, above which the signal to noise ratio is below one.

Table 1: Domain-average RMSE obtained from triple colocation ($RMSE^{TC}$) for AMSR-E, ASCAT, and Catchment model soil moisture, presented using each data set in turn as the reference.

Reference	$RMSE^{TC}$		
	ASCAT	AMSR-E	CATCH
ASCAT (%)	9	15	16
AMSR-E (m^3m^{-3})	0.04	0.05	0.06
CATCH (m^3m^{-3})	0.03	0.04	0.03