Remote Sensing of Environment 137 (2013) 288–298

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

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,e], Robert Parinussa [c], Wolfgang Wagner [e]

[a] NASA Goddard Space Flight Center, Global Modeling and Assimilation Office, Code 610.1, Greenbelt, 20771 MD, USA
[b] GESTAR, Universities Space Research Association, Columbia, MD, 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

ARTICLE INFO

Article history:
Received 13 March 2013
Received in revised form 21 June 2013
Accepted 22 June 2013
Available online 1 August 2013

Keywords:
Microwave soil moisture
Remotely sensed soil moisture validation
Triple colocation
Error propagation

ABSTRACT

Root Mean Square Errors (RMSEs) 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 local time series standard deviation (fRMSE). For both sensors, the fRMSE^{TC} 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 in general both data sets have good skill over low to moderate vegetation cover. Additionally, they have similar accuracy even when considered by land cover class, although the AMSR-E fRMSEs show a stronger signal of the vegetation cover.

1. Introduction

Soil moisture is an important control over hydrological and meteorological processes, 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 (Bartalis et al., 2007; Wagner, Lemoine, & Rott, 1999) and the Advanced Microwave Scanning Radiometer — Earth Observing System (AMSR-E) (Njoku, 1999; Owe, de Jeu, & Walker, 2001). Additionally, several remote sensors have recently been designed specifically 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, scheduled for launch in 2014 (Entekhabi, Njoku, et al., 2010).

The performance of new remotely sensed soil moisture data sets is bench-marked against predetermined root mean square error (RMSE) target accuracies (Entekhabi, Njoku, et al., 2010; Kerr et al., 2001) 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, Koster, Dong, & Berg, 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 (Reichle et al., 2004; Vinnikov, Robock, Qiu, & Entin, 1999) or differences in the soil moisture processes resolved by each soil moisture estimate

* Corresponding author. Tel.: +1 301 614 6161.
E-mail address: clara.draper@nasa.gov (C. Draper).

0034-4257 © 2013 The Authors. Published by Elsevier Inc. Open access under CC BY-NC-SA license.
http://dx.doi.org/10.1016/j.rse.2013.06.013

(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, Schlenz, Loew, & Mauser, 2012), or from networks of sparse in situ soil moisture sensors (Draper, Walker, Steinle, de Jeu, & Holmes, 2009; Reichle et al., 2007; Wagner et al., 1999). 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 RMSEs in remotely sensed soil moisture over continental 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. RMSEs are estimated for two remotely sensed soil moisture products: the Surface Degree of Saturation (SDS) retrieved from active microwave ASCAT observations (Bartalis et al., 2007; Wagner et al., 1999), and the X-band passive microwave AMSR-E soil moisture retrieved with the Land Parameter Retrieval Model (LPRM; de Jeu and Owe (2003); Owe et al. (2001)). While neither of these missions were designed to sense soil moisture, both have been providing useful soil moisture observations (Draper, Reichle, De Lannoy, & Liu, 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 (Scipal, Holmes, de Jeu, Naeimi, & Wagner, 2008; Stoffelen, 1998), which combines three independent estimates of a state variable to calculate the errors in each assuming an additive error model. The second method is error propagation through the models used to retrieve soil moisture from the microwave observations, as developed by Naeimi, Scipal, Bartalis, Hasenauer, and Wagner (2009) for the ASCAT SDS and Parinussa, Meesters, et al. (2011) for the AMSR-E LPRM retrievals. The error estimates are investigated over a continental scale domain, between 25 and 50°N in North America.

Due to the systematic differences between large scale soil moisture estimates, different soil moisture data sets describe different climates as measured by their mean and variance. 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 often higher-order central moments) are typically eliminated by rescaling all data sets to have statistics consistent with an arbitrarily selected 'reference' data set (Reichle & Koster, 2004; Scipal, Drusch, & Wagner, 2008). Over large domains, soil moisture RMSEs estimated by comparing different data sets must then be based on rescaled data sets, and so are presented relative to the climatology of the reference data set (e.g., dall'Amico et al. (2012); Dorigo et al. (2010); Draper et al. (2009); Scipal, Holmes, et al. (2008)). Hence, before investigating the triple colocation and error propagation RMSE estimates, the consequences of this rescaling are examined in terms of the information contained in the resulting RMSE estimates.

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 over large domains. Also, the assumptions underlying triple colocation are tested in Section 4.2, by examining the dependence of the estimated RMSE on the three 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 (VUT), 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 observations 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 relates to soil moisture over a ~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-bands) 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 (de Jeu & Owe, 2003; Owe et al., 2001). 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 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, Bovis, Macpherson, and Jones (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 AMSR-E 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.

Fig. 1 shows a map of the land cover classes for the regions where remotely 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 remain 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 these locations are usually screened from the soil moisture data sets using ancillary vegetation data (e.g., Draper et al. (2012)). However, in this study these locations have been retained to explicitly test whether the error estimation methods can detect the larger errors expected over dense vegetation.

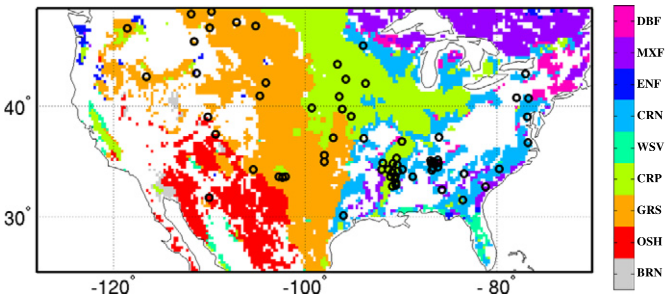

Fig. 1. MODIS land cover classes (Friedl et al., 2002), plotted where remotely sensed soil moisture data are available after quality control. The land cover classes are deciduous broadleaf (DBF), mixed forest (MXF), evergreen needleleaf (ENF), crop/natural mix (CRN), woody savanna (WSV), cropland (CRP), grassland (GRS), open shrub (OSH), and barren (BRN). Black circles indicate the location of the SCAN/SNOTEL sites used in Section 4.2.

2.2. Catchment model soil moisture

Soil moisture simulations from NASA's Catchment land surface model (Koster, Suarez, Ducharne, Stieglitz, & Kumar, 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 the NASA Modern-Era Retrospective analysis for Research and Applications (MERRA)-Land model version and meteorological forcing (Reichle, 2012; Reichle et al., 2011). 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, Cosh, & Jackson, 2007) sites shown in Fig. 1. At each of these sites a daily 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 (Fig. 1), with an average of 261 coincident observations at each site.

3. Methods

3.1. Triple colocation

Triple colocation is used here to estimate the errors (in terms of the RMSE) in the soil moisture anomaly time series observed by ASCAT (θ_A), AMSR-E (LPRM) (θ_L), and the Catchment model (θ_C). The soil moisture anomaly time series were defined as the deviations of the raw data from their multi-year, seasonally varying climatology. For each data set, 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.

Triple colocation was developed by Stoffelen (1998) to calibrate and evaluate noisy data sets with respect to each other (in the absence of a single error-free data source). The method applied here follows that of Stoffelen (1998). At each grid cell, the anomaly soil moisture time series for each data set are assumed to consist of the (unknown) 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, and where the subscripts A, L, and C identify the ASCAT, AMSR-E (LPRM), and Catchment data. Bias terms were not included in Eqs. (1)–(3), since (zero-mean) anomaly time series have been used.

Eqs. (1)–(3) are underdetermined, and so one data set is selected as the reference and the remaining two are calibrated to be consistent with this reference in terms of the time series variability. For example, if θ_A is the reference, α is set to one, and estimates of the remaining calibration constants, $\widehat{\lambda}$ and $\widehat{\gamma}$, can then be obtained from:

$$\frac{\langle\theta_L\theta_C\rangle}{\langle\theta_A\theta_C\rangle} = \frac{\lambda\langle\theta^2 + \theta\epsilon_L + \theta\epsilon_C + \epsilon_L\epsilon_C\rangle}{\langle\theta^2 + \theta\epsilon_A + \theta\epsilon_C + \epsilon_A\epsilon_C\rangle} \approx \widehat{\lambda} \tag{4}$$

$$\frac{\langle\theta_L\theta_C\rangle}{\langle\theta_A\theta_L\rangle} = \frac{\gamma\langle\theta^2 + \theta\epsilon_L + \theta\epsilon_C + \epsilon_L\epsilon_C\rangle}{\langle\theta^2 + \theta\epsilon_A + \theta\epsilon_L + \epsilon_A\epsilon_L\rangle} \approx \widehat{\gamma} \tag{5}$$

where $\langle\cdot\rangle$ represents the long-term mean. The final approximations in Eqs. (4) and (5) apply only if the errors in each data set are not cross-correlated with each other, nor with the true state variable (e.g., $\langle\epsilon_L\epsilon_C\rangle = 0$, $\langle\theta\epsilon_L\rangle = 0$ and so on).

The square root of the estimated $\langle\epsilon^2\rangle$ are the triple colocation RMSE estimates ($RMSE^{TC}$), and can be obtained from:

$$\left\langle\left(\theta_A - \frac{\theta_L}{\widehat{\lambda}}\right)\cdot\left(\theta_A - \frac{\theta_C}{\widehat{\gamma}}\right)\right\rangle = \langle\epsilon_A^2\rangle - \langle\epsilon_A\epsilon_L\rangle - \langle\epsilon_A\epsilon_C\rangle + \langle\epsilon_L\epsilon_C\rangle \approx \langle\widehat{\epsilon_A}^2\rangle \tag{6}$$

$$\left\langle\left(\frac{\theta_L}{\widehat{\lambda}} - \theta_A\right)\cdot\left(\frac{\theta_L}{\widehat{\lambda}} - \frac{\theta_C}{\widehat{\gamma}}\right)\right\rangle = \langle\epsilon_L^2\rangle - \langle\epsilon_A\epsilon_L\rangle - \langle\epsilon_L\epsilon_C\rangle + \langle\epsilon_A\epsilon_C\rangle \approx \langle\widehat{\epsilon_L}^2\rangle \tag{7}$$

$$\left\langle\left(\frac{\theta_C}{\widehat{\gamma}} - \theta_A\right)\cdot\left(\frac{\theta_C}{\widehat{\gamma}} - \frac{\theta_L}{\widehat{\lambda}}\right)\right\rangle = \langle\epsilon_C^2\rangle - \langle\epsilon_A\epsilon_C\rangle - \langle\epsilon_L\epsilon_C\rangle + \langle\epsilon_A\epsilon_L\rangle \approx \langle\widehat{\epsilon_C}^2\rangle. \tag{8}$$

Again the final approximations apply only if the errors for each data set are mutually uncorrelated.

Recall that the above equations were derived after choosing θ_A as the reference data set (i.e., $\alpha = 1$). To highlight this, the reference data set will be indicated in parentheses. For example, the triple colocation RMSE of the AMSR-E (L) soil moisture anomalies with ASCAT (A) as the reference data set is written $RMSE_L^{TC}(A)$. The results can be converted to another reference data set by multiplication with the appropriate calibration constant, or by repeating the calculation from the beginning with an alternative calibration constant set to one.

In summary, the triple colocation method relies on the following assumptions: i) Eqs. (1)–(3) provide an adequate error model, and ii) the errors in each data set are not cross-correlated with each other, nor with the truth. For soil moisture, the errors in different data sets can be cross-correlated, for example due to the use of the same ancillary data sets, or the same (imperfect) physics in the remote sensing retrieval algorithms or land surface models. The three data sets used here were carefully selected to minimize the chance of this occurring, however their errors could still be cross-correlated due to the common impact of problematic geophysical conditions (e.g., errors in both satellite estimates due to the intermittent presence of standing water). Additionally, Eqs. (1)–(3) are based on the assumption that all three estimates relate to the same geophysical variable, which may be problematic for soil moisture, given the representativity differences that exist between many soil moisture data sets (e.g., differences in the depth represented by each). As demonstrated in Appendix A, significant representativity differences between the data sets used in triple colocation will bias the resulting RMSE estimates in favor of the two most similarly defined data sets. Consequently, in Section 4.2 the triple colocation assumptions will be indirectly tested by examining the dependence of the results on the three data sets used.

Since the triple colocation was based on soil moisture anomalies (deviations from the mean seasonal cycle), the $RMSE^{TC}$ estimates represent only the errors in the soil moisture anomaly time series, or equivalently the anomalies in the soil moisture error time series. Hence, the errors in the mean seasonal cycle and the long-term mean error (bias) are not included in the $RMSE^{TC}$ estimates defined above. The choice to base the triple colocation on anomalies from the mean seasonal cycle follows Miralles, Crow, and Cosh (2010), who found anomalies to be more consistent with the triple colocation assumptions than raw soil moisture time series are. The importance of using anomalies from the mean seasonal cycle will be confirmed in Section 4.2.

Finally, many soil moisture triple colocation studies have excluded the calibration constants from Eqs. (1)–(3), and instead rescaled the data sets with the ratio of their standard deviations prior to applying the above error model (e.g., Dorigo et al. (2010); Miralles et al. (2010)). However, as discussed by Stoffelen (1998) and Yilmaz and Crow (2013), the latter approach 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%).

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, developed by Naeimi et al. (2009), are produced in parallel with the SDS data using Gaussian error propagation. For AMSR-E, Parinussa, Meesters, et al. (2011) 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 RMSE time series has been used as the error propagation RMSE estimate ($RMSE^{EP}$).

It is unclear whether the error propagation RMSEs better estimate the errors in the raw soil moisture time series, or the errors in the anomalies from the mean seasonal cycle. The error propagation 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 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 provide the errors in the anomalies from the mean seasonal cycle, consistent with the ϵ defined in Eqs. (1)–(3) for triple colocation.

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 $RMSE^{EP}$ confidence intervals cannot be estimated. In contrast, for triple colocation $RMSE^{TC}$ confidence intervals can be estimated using bootstrapping, following Caires and Sterl (2003). Bootstrapping is useful for estimating the uncertainty 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 $RMSE^{TC}$) have then been estimated directly from the bootstrapped distribution.

To estimate the confidence limits for the mean $RMSE^{TC}$ 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 assumed not to be independent, which conservatively overestimates the width of the confidence limits. 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 intervals for the mean are then calculated separately from 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° × 5° 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 ($RMSE_{5°}$) 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 $\sqrt{\frac{1}{m}\Sigma_{i=1}^{m}(RMSE_{5^\circ,i})^2}$. 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 consistent with a chosen reference data set (e.g., by using the calibration constants defined by Eqs. (1)–(3) to rescale each data set to have consistent mean and variance). 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 data sets, A and B, with equivalent means is a function of the standard deviation of each data set (σ_A and σ_B) and the correlation (R) between them:

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

The RMSD between two rescaled soil moisture data will then depend on the standard deviation of the reference data set and the correlation between the two data sets: note that the signal of the agreement between the data sets is derived from their correlation.

Hence, a RMSE calculated from rescaled data sets will have a strong dependence on the standard deviation of the reference data set. This is highlighted in Fig. 2, which 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 0.07 $m^3\ m^{-3}$ assuming a porosity of 0.5 $m^3\ m^{-3}$) for ASCAT, 0.07 $m^3\ m^{-3}$ for AMSR-E, and 0.03 $m^3\ m^{-3}$ for Catchment. The spatial patterns described by each are also very different. The magnitude of the ASCAT $RMSE^{TC}$ estimates also differs depending on which data set is used as the reference, with the mean varying between 10% SDS (0.05 $m^3\ m^{-3}$), 0.07 $m^3\ m^{-3}$, and 0.07 $m^3\ m^{-3}$ when ASCAT, AMSR-E, and Catchment are used, respectively. The spatial patterns in the ASCAT RMSE^{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.

At an individual location the ratio between the RMSEs for different data sets does not depend on the selected reference data set. However, since the relationship between the different σ maps in Fig. 2 is nonlinear, the ratio and even the ranking of the domain-averaged RMSE do depend on the reference data set. For example, Table 1 lists the mean RMSE^{TC} across the domain for each data set, presented using each data set as the reference. Most notably, with ASCAT as the reference Catchment has the highest mean RMSE^{TC}, and yet with Catchment as the reference Catchment has the lowest mean RMSE^{TC}.

This study investigates the spatial variability in remotely sensed soil moisture RMSEs. If the RMSE were presented in soil moisture units (relative to an arbitrarily selected 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:

$$fRMSE_X = RMSE_X(X)/\sigma_X(X). \tag{10}$$

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 ($\sigma_X(X)$).

By reducing the signal of the standard deviation in Eq. (9), the fRMSE statistic becomes more consistent with the common use of correlation statistics to evaluate soil moisture (e.g., de Jeu et al. (2008); Draper et al. (2012); Parinussa, Holmes, and Crow (2011); Reichle et al. (2007); Scipal, Drusch, et al. (2008)).

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	$RMSE^{TC}$ AMSR-E	$RMSE^{TC}$ CATCH
ASCAT (%)	10	18	19
AMSR-E (m^3m^{-3})	0.07	0.06	0.09
CATCH (m^3m^{-3})	0.07	0.08	0.03

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 $1/\sqrt{2}$ (~ 0.7) indicating an error variance that exceeds the variance of the true time series (since $RMSE_X(X)/\sigma_X(X) \approx \sqrt{\epsilon_X^2(X)/(\sigma_T^2(X)+\epsilon_X^2(X))}$, where σ_T is the standard deviation of the true soil moisture). Additionally, 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, as 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 at an individual location the conversion of the RMSE to fRMSE does not preserve the ratio between different error estimates, although it does preserve their

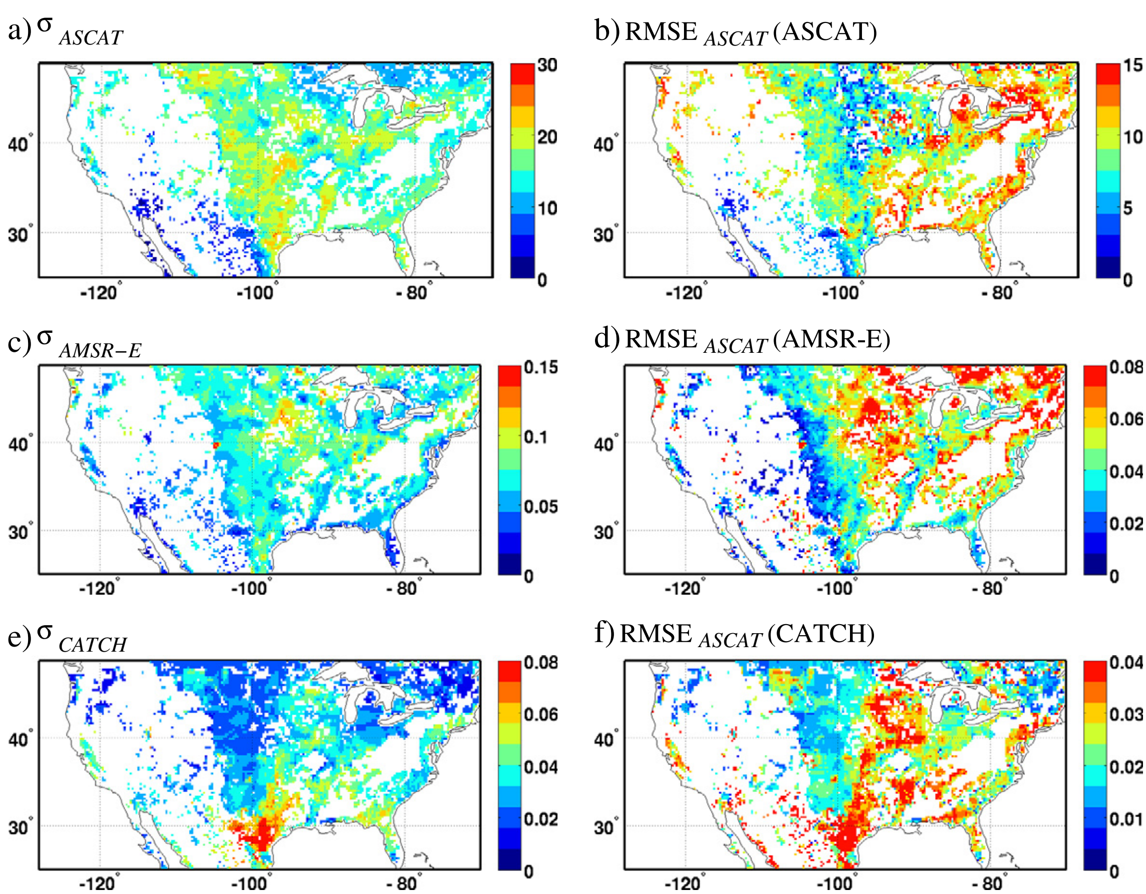

Fig. 2. Maps of (left) the standard deviation in the soil moisture anomaly time series from a) ASCAT (%), c) AMSR-E ($m^3\ m^{-3}$), and e) Catchment ($m^3\ m^{-3}$), and (right) RMSE estimates for ASCAT from triple colocation, presented using b) ASCAT (%), d) AMSR-E ($m^3\ m^{-3}$), and f) Catchment ($m^3\ m^{-3}$) as the reference data set. White indicates that triple colocation results are not available.

ranking. For example, converting $RMSE_X(Y)$ to the $fRMSE_X$ is achieved by multiplication with $\sqrt{1/(\sigma_T^2(Y)+\epsilon_X^2(Y))}$ (since for example squaring then rearranging Eq. (2) gives $\lambda(A) \approx \sqrt{\sigma_L^2(L)/(\sigma_T^2(A)+\epsilon_B^2(A))}$). Because this operation is nonlinear yet monotonic, the ratio between different errors estimates will not be conserved, although the ranking between them will be.

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 $RMSE_X(Y)/\sigma_Y(Y) \approx \sqrt{\epsilon_X^2(Y)/(\sigma_T^2(Y)+\epsilon_Y^2(Y))}$). The result is no longer self contained and can generate unexpected results, due to the dependence on $\epsilon_Y^2(Y)$.

For the remainder of this paper, the RMSE estimates are presented using 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 relevant soil moisture anomaly time series. Note that the change of units associated with reporting the errors as a fRMSE magnifies differences in 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 Fig. 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 in Section 3.4 above, the error estimates are presented here in terms of the fRMSE (Eq. 10). Fig. 3 shows maps of the ASCAT and AMSR-E fRMSE calculated from triple colocation ($fRMSE^{TC}$) and error propagation ($fRMSE^{EP}$). 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 magnitude of the error propagation output 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 magnitudes obtained for the AMSR-E $fRMSE^{EP}$ are not surprising.

Fig. 3 shows similar large scale patterns in 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 $fRMSE^{EP}$ (Fig. 3b) the eastward increase is weaker than for the other maps.

Given the uncertain magnitude of the error propagation output, the ASCAT and AMSR-E errors can only be compared based on the triple colocation results. In Fig. 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 reverse occurs.

To establish whether these differences in the $fRMSE^{TC}$ are significant, Fig. 4 shows the width of the 90% confidence interval for the $fRMSE^{TC}$ estimates (see Section 3.3), while Fig. 5 indicates regions where the (one sided) differences between the ASCAT and AMSR-E $fRMSE^{TC}$ are significant (at 5%). Fig. 4 shows that the width of the confidence intervals exceeds 0.5 for ASCAT in the Mid-West of the US and in the northeast of the domain, and also for AMSR-E over a subregion of the Mid-West (i.e., the 90% confidence interval spans >50% of the possible range). Over the rest of the domain, the typical width of the confidence intervals is between 0.1 and 0.3, with a tendency for the ASCAT and AMSR-E intervals to offset each other (i.e., one is relatively high where the other is relatively low). Fig. 5 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, with the reverse occurring in the east of the domain. Despite the large uncertainties, Figs. 4 and 5 also shows that the lower ASCAT $fRMSE^{TC}$ (compared to the AMSR-E $fRMSE^{TC}$) is significant across much of the Mid-West.

Fig. 6 shows the mean fRMSE by land cover, for each land cover class with at least 100 grid cells with triple colocation results in Fig. 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

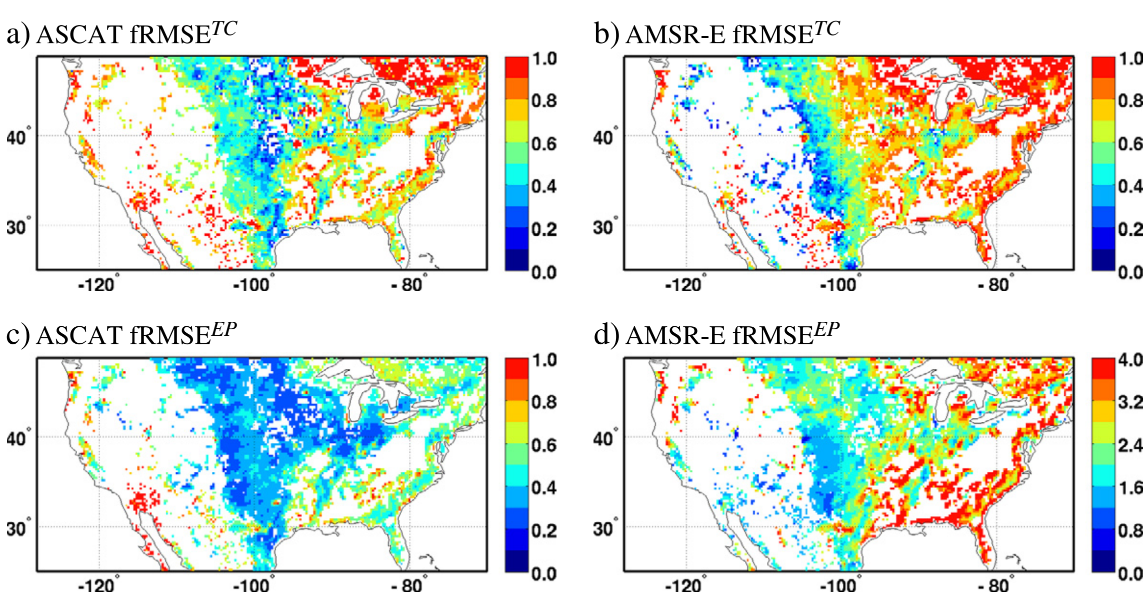

Fig. 3. fRMSE of (left) ASCAT and (right) AMSR-E, fRMSE estimated using the (upper) triple colocation and (lower) error propagation methods. White indicates that triple colocation results are not available. Note the different color scale for the AMSR-E $fRMSE^{EP}$ in subfigure d.

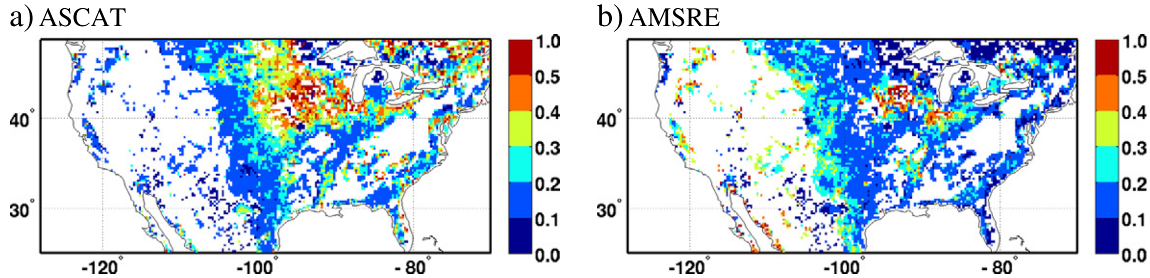

Fig. 4. Width of the 90% confidence interval for the fRMSE^{TC} of the a) ASCAT and b) AMSR-E soil moisture anomalies. White indicates that triple colocation results are not available.

class is also included in Fig. 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, there is good agreement between the variation in fRMSE^{TC} and fRMSE^{EP} between land cover types in Fig. 6, except that over the five densely vegetated categories (woody savanna, crop/natural mix, evergreen needleleaf, mixed forest, and deciduous broadleaf), the mean fRMSE^{EP} is lower than the mean fRMSE^{TC}, and is even below the $1/\sqrt{2}$ line (this is the relatively low fRMSE^{EP} in the east in Fig. 3b). The analysis was repeated without discarding ASCAT data with high error propagation errors (see Section 2.1), and this separate analysis (not shown) confirmed that this quality control step was not the cause of the low ASCAT fRMSE^{EP} in densely vegetated regions.

For AMSR-E, there are also differences in the behavior of the fRMSE^{TC} and fRMSE^{EP} across the land cover classes. For triple colocation, the variability between the mean AMSR-E fRMSE^{TC} for each land cover class closely reflects the variability in the mean LAI. However, for error propagation the mean AMSR-E fRMSE^{EP} are effectively grouped into two bins: the three land cover classes with the lowest LAI were assigned similar and relatively low mean fRMSE^{EP}, while the remaining five land cover classes were 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 Fig. 3d.

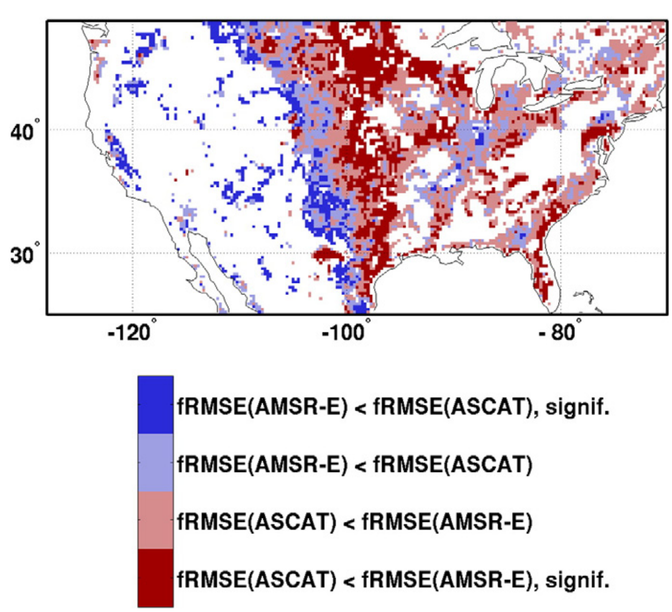

Fig. 5. Comparison of ASCAT and AMSR-E fRMSE^{TC}. Blue (red) indicates AMSR-E fRMSE^{TC} less (more) than ASCAT fRMSE^{TC}, with darker shades indicating a significant difference at 5%. White indicates that triple colocation results are not available.

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 so moderate vegetation is expected to be less detrimental to the accuracy of the ASCAT SDS than to the accuracy of the passive microwave AMSR-E soil moisture. This is consistent with Fig. 6 which shows that the relationship between the mean LAI and the mean fRMSE is much weaker for ASCAT than for AMSR-E. For ASCAT, factors other than vegetation can also be significant in determining the errors in the soil moisture retrievals. For example, open shrubs have the lowest mean LAI in Fig. 6, yet the mean ASCAT fRMSE estimates over the open shrubs are very high (close to $1/\sqrt{2}$) for both methods. The open shrub grid cells are in the arid southwest of the domain (Fig. 1), where the spatial sampling in Fig. 3 is poor (229 out of the nearly 7000 grid cells plotted). The poor performance of the ASCAT 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 ASCAT fRMSE, despite the croplands having much higher LAI. The 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 Fig. 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 cropland), at least one of the ASCAT and AMSR-E fRMSE^{TC} is significantly less than $1/\sqrt{2}$, indicating an ability to accurately detect near-surface soil moisture anomalies from satellites under low to moderate vegetation cover (which cover 63% of the domain with fRMSE values in Fig. 3). For the five 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 Figs. 2 to 5 the plotted coverage is less than that of the quality controlled data in Fig. 1, due to the triple colocation having produced negative mean square errors at 9% 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 in arid locations where the ASCAT errors are very large. This suggests that triple colocation may require a minimum skill from all three data sets, consistent with the assumption that all three data sets observe the same geophysical 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

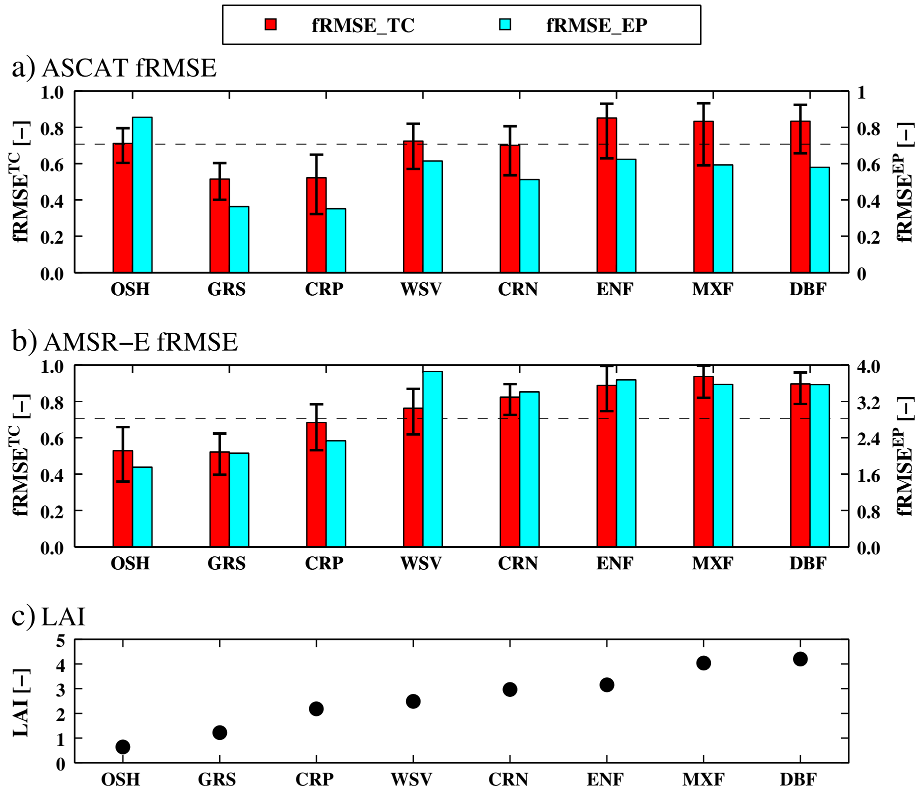

Fig. 6. Mean by land cover class for the fRMSE^{TC} and fRMSE^{EP} of a) ASCAT and b) AMSR-E, and for c) LAI. For fRMSE^{TC} the 90% confidence interval is included (uncertainty estimates are not available for fRMSE^{EP}). See the caption of Fig. 1 for the land cover class key. Note the different y-axes (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.

(Section 3.1). However, for soil moisture non-negligible cross-correlations between the errors and/or representativity differences (extending beyond systematic differences in the central moments) are common between global soil moisture data sets, potentially leading to violations of these assumptions. These assumptions are difficult to test directly, since time series of soil moisture RMSE for a given data set are not generally available. Hence, the impact on the fRMSE^{TC} of any violations to the triple colocation assumptions is examined here by testing how the fRMSE^{TC} estimates depend on the three data sets used in the triple colocation (on the assumption that violations to the assumptions will be specific to certain combinations of data sets). In situ soil moisture from the 57 SCAN/SNOTEL sites has been used as an alternate data set.

Fig. 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 (even with the very conservative method used to estimate the confidence intervals — see Section 3.3). 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 Fig. 6, and close to the typical reporting precision for soil moisture data.

The dependence of the fRMSE^{TC} estimates on the data triplet used in the triple colocation is however consistent with the expected representativity differences between the four data sets. For ASCAT and AMSR-E, the fRMSE^{TC} estimates are lower when both remote sensors are included in the triplet (left two triplets in Fig. 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 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 Fig. 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

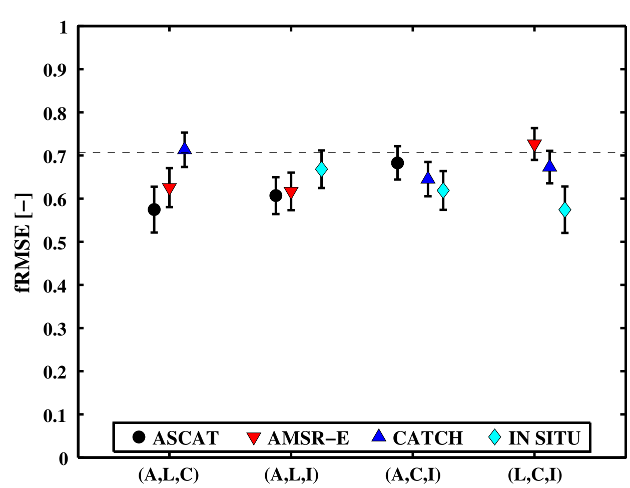

Fig. 7. Mean fRMSE and its 90% confidence interval estimated across the SCAN/SNOTEL sites, using triple colocation based on different combinations of three of (A) ASCAT, (L) AMSR-E, (C), Catchment, and (I) in situ data sets. The data triplet is indicated in the x-axis labels, while the plotted symbol/color 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.

because the $fRMSE^{TC}$ differences reported in Section 4.1 are much larger than the <0.1 differences obtained in this Section. 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 ~0.1), 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 defined as the deviations from the mean seasonal cycle, rather than as the deviations from a single long-term mean (or on raw soil moisture data). Repeating Fig. 7 with the triple colocation based on anomalies from a single long-term mean 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 that for soil moisture the triple colocation assumptions require the use of anomalies from the seasonal cycle (see also Miralles et al. (2010)).

Finally, Fig. 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 Fig. 7 the SCAN/SNOTEL fRMSE (when estimating coarse scale soil moisture) are as large as the ASCAT and AMSR-E fRMSEs. Hence, the RMSD between either remote sensor and the SCAN/SNOTEL data will significantly over-estimate the errors in the remotely sensed data. To address this, Miralles et al. (2010) correct the RMSD between modeled/remotely sensed soil moisture and in situ data using a triple colocation estimate of the RMSE of the in situ data. While this method is useful for highlighting the substantial contribution of the in situ errors to the RMSD, it is equivalent to directly estimating the modeled/remotely sensed RMSE using triple colocation (see Appendix B), and so has not been applied here.

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 near-surface 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 the observation error variances needed for data assimilation.

In the absence of a 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 magnitudes and spatial patterns of the resulting RMSE will then depend on the selected reference data set, and specifically on its standard deviation (Fig. 2). This study proposes to reduce this dependence by presenting the RMSE for each data set as the fraction of the standard deviation of that data set (fRMSE, Eq. 10).

Comparing the triple colocation and error propagation fRMSE over a continental scale domain indicates that both methods can accurately detect the large scale variability in soil moisture errors. In Fig. 3 the regions with relatively high and low $fRMSE^{TC}$ and $fRMSE^{EP}$ agree very well, and in Fig. 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 variability of the errors within a given data set, and while not used here, the unique ability to produce time series of the expected RMSE 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 $fRMSE^{EP}$ are unrealistically large. For AMSR-E, the LPRM model parameter uncertainties used in the error propagation were conservatively estimated to be quite large (Parinussa, Meesters, et al., 2011), 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 would 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., Vrugt et al. (2009)).

The uncertainty in the $fRMSE^{EP}$ suggested by comparison to $fRMSE^{TC}$ can also be useful for identifying shortcomings of the retrieval models. For AMSR-E, the $fRMSE^{EP}$ were relatively low over sparsely vegetated regions, and relatively high over densely vegetated regions, with little graduation between these two modes (Figs. 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 Fig. 6. This error propagation behavior can be traced to a limitation of the tau-omega model used by the LPRM. The tau-omega model parameterizes the attenuation of the soil moisture signal by vegetation using an exponential function of vegetation optical depth (Parinussa, Meesters, et al. (2011), their Eq. 2), resulting in an exponential increase in the tau-omega error propagation output with increasing vegetation optical depth (Parinussa, Meesters, et al. (2011), their Fig. 2). The sudden and steep increase in the AMSR-E $fRMSE^{EP}$ with increasing vegetation in Figs. 3 and 6 of this study suggests that the tau-omega model is over-estimating this nonlinear sensitivity to vegetation attenuation. This highlights a potential area of improvement to 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 magnitude of the $fRMSE^{TC}$ is not accurate, with the caveat that the errors are relative to the soil moisture anomaly truth defined by the three data sets used. When the dependence of the $fRMSE^{TC}$ on the triplet of data sets used was tested at 57 SCAN/SNOTEL sites, only small systematic differences (below the typical reporting precision of soil moisture) were found between the $fRMSE^{TC}$ for a given data set when different combinations of the three data sets were used.

This robustness to the selection of data sets used is surprising, given the substantial representativity differences expected between the point-based in situ and coarse scale satellite or model soil moisture estimates. While the representativity differences and other causes of correlated errors between the 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 mean seasonal cycle.

In contrast to error propagation, triple colocation provides a consistent method for estimating the fRMSE 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. Note that X-band AMSR-E data were used here due to radio frequency interference in the C-band observations over North America, and slightly better AMSR-E accuracy is expected in other regions where C-band observations can be used. While the ASCAT and AMSR-E $fRMSE^{TC}$ estimates were generally similar, there was some deviation in the dependence of the accuracy of each data set on the land cover. The AMSR-E $fRMSE^{TC}$ had a much stronger dependence on vegetation cover, while the ASCAT $fRMSE^{TC}$ deviated more from this dependence. Specifically, the ASCAT errors were very high in arid (open shrubs) regions, and relatively low over the croplands in the Mid-West of the US.

Similar patterns in the relative performance of passive and active soil moisture retrievals have been reported previously. Using triple colocation, Scipal, Holmes, et al. (2008) observed a much stronger relationship between the anomaly RMSE in LPRM X-band TRMM Microwave Imager (passive) retrievals than in TUW ERS (active) soil moisture data, resulting in a similar pattern of more accurate passive (active) microwave retrievals in the west (east) of North America. Also using triple colocation, Dorigo et al. (2010) found the anomaly RMSE in LPRM C-band AMSR-E data to have a stronger relationship with vegetation than the TUW ASCAT RMSEs. Additionally, using a precipitation-driven correlation metric, Crow and Zhan (2007) showed that Single Channel Algorithm (Jackson, 2003) X-band AMSR-E soil moisture errors had a stronger relationship to vegetation than the TUW ASCAT soil moisture errors, producing the same east–west division over North America. The consistency of these results, despite differences in the passive and active sensors, passive microwave retrieval algorithms, and passive wavelengths used in these studies suggests that the differing response to vegetation could be associated with differences in the nature of the active and passive observations, or with the use of the change-detection retrieval method for active microwave, rather than the radiative transfer model-based passive microwave approaches.

6. Conclusions and recommendations

The above findings have implications for the evaluation of remotely sensed soil moisture data. Currently, novel remotely sensed soil moisture data sets 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 Figs. 2 and 3 there is substantial spatial variation in the soil moisture RMSE and fRMSE, highlighting that such an evaluation based on a limited number of locations will not necessarily be representative of a larger domain. Hence, the validation efforts at well-observed pixels should be complemented with distributed methods that can estimate soil moisture uncertainty 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 (RMSE of anomalies from the mean 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 accuracy 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 fRMSE.

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, Reichle, Koster, and Crow (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 observation error variance is then scaled to be consistent with the model climatology in the same manner as the observations are). Again, as outlined by Entekhabi, Reichle, et al. (2010) 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, for example 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 estimates could also be used, after appropriate rescaling to correct the magnitude.

Acknowledgments

This research was supported by the NASA Soil Moisture Active Passive mission and by the NASA program on the Science of Terra and Aqua. Gabrielle De Lannoy and Qing Liu contributed through many helpful discussions.

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 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 Eqs. (6)–(8) are approximate by:

$$RMSE^{TC}_{X1} = \sqrt{<\epsilon^2_{X1}> - <\epsilon_{X1}.\epsilon_{X2}>^2} \tag{A.1}$$

$$RMSE^{TC}_{X2} = \sqrt{<\epsilon^2_{X2}> - <\epsilon_{X1}.\epsilon_{X2}>^2} \tag{A.2}$$

$$RMSE^{TC}_{Y} = \sqrt{<\epsilon^2_{Y}> + <\epsilon_{X1}.\epsilon_{X2}>^2}. \tag{A.3}$$

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 an 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 Eqs. (4) and (5). However, this effect will be secondary to that described above since in Eqs. (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.

Appendix B. Equivalence of Miralles et al. (2010) in situ RMSD correction to triple colocation

Miralles et al. (2010) introduce a method to 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 $\mathrm{RMSE}_X^{IS} \approx \sqrt{\left(RMSD_X^{IS}\right)^2 - \left(RMSE_I^{TC}\right)^2}$. While this method is useful for highlighting the contribution of the in situ errors to the RMSD, it is equivalent to directly estimating the RMSE of data set X using triple colocation. With reference to the triple colocation equations (Eqs. 6–8), the corrected RMSD can be written:

$$RMSE_X^{IS} \approx \sqrt{(RMSD_X^{IS})^2 - (RMSE_I^{TC})^2} \tag{B.1}$$

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

$$= \sqrt{<(\theta_X - \theta_I)(\theta_X - \theta_Y)>} \tag{B.3}$$

$$= RMSE_X^{TC}. \tag{B.4}$$

The calibration constants have been neglected above for clarity. However, this result 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).

References

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., et al. (2007). Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophysical Research Letters*, *34* L20401.

Caires, S., & Sterl, A. (2003). Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research*, *16*, 3098.

Crow, W., & Zhan, X. (2007). Continental-scale evaluation of remotely-sensed soil moisture products. *IGARS Letters*, *4*, 451–455.

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

de Jeu, R., & Owe, M. (2003). Further validation of a new methodology for surface moisture and vegetation optical depth retrieval. *International Journal of Remote Sensing*, *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. *Surveys in Geophysics*, *29*, 399–420.

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

Dorigo, W., Scipal, K., Parinussa, R., Liu, Y., Wagner, W., de Jeu, R., et al. (2010). Error characterisation of global active and passive microwave soil moisture datasets. *Hydrology and Earth System Sciences*, *14*, 2605–2616.

Draper, C., Reichle, R., De Lannoy, G., & Liu, Q. (2012). Assimilation of passive and active microwave soil moisture retrievals. *Geophysical Research Letters*, *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 Sensing of Environment*, *113*, 703–710.

Entekhabi, D., Njoku, E., O'Neill, P., Kellogg, K., Crow, W., Edelstein, W., et al. (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. *Journal of Hydrometeorology*, *11*, 832–840.

Friedl, M., McIver, D., Hodges, J., Zhang, X., Muchoney, D., Strahler, A., et al. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *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 Transactions on Geoscience and Remote Sensing*, *50*, 1623–1635.

Jackson, T. (2003). Measuring surface soil moisture using passive microwave remote sensing. *Hydrological Processes*, *7*, 139–152.

Jackson, T., Bindlish, R., Cosh, M., Tianjie, Z., Starks, P., Bosch, D., et al. (2012). Validation of Soil Moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the U.S. *IEEE Transactions on Geoscience and Remote Sensing*, *50*, 1530–1543.

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 Transactions on Geoscience and Remote Sensing*, *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. *Journal of 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. *Journal of Geophysical Research*, *105*, 24809–24822.

Mahfouf, J. -F. (2010). Assimilation of satellite derived soil moisture from ASCAT in a limited area NWP model. *Quarterly Journal of the Royal Meteorological Society*, *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. *Journal of Hydrometeorology*, *11*, 1423–1429.

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

Njoku, E. (1999). Retrieval of land surface parameters using passive microwave measurements at 6–18 GHz. *IEEE Transactions on Geoscience and Remote Sensing*, *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 Transactions on Geoscience and Remote Sensing*, *39*, 1643–1654.

Parinussa, R., Holmes, T., & Crow, W. (2011). The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrology and Earth System Sciences*, *15*, 3135–3151.

Parinussa, R., Meesters, A., Liu, Y., Dorigo, W., Wagner, W., & de Jeu, R. (2011). 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. (2012). *The MERRA-Land Data Product*. GMAO Office Note No. 3 (Version 1.2) (Available from http://gmao.gsfc.nasa.gov/pubs/office_notes).

Reichle, R., & Koster, R. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, *31* L19501.

Reichle, R., Koster, R., De Lannoy, G., Forman, B., Liu, Q., Mahanama, S., et al. (2011). Assessment and enhancement of MERRA land surface hydrology estimates. *Journal of Climate*, *24*, 6322–6338.

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. *Journal of Hydrometeorology*, *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). *Journal of Geophysical Research*, *112* D09108.

Schaefer, G., Cosh, M., & Jackson, T. (2007). The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *Journal of Atmospheric and Oceanic Technology*, *24*, 2073–2077.

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

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

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research*, *7755–7766*.

Vinnikov, K., Robock, A., Qiu, S., & Entin, J. (1999). Optimal design of surface networks for observation of soil moisture. *Journal of Geophysical Research*, *104*, 19743–19749.

Vrugt, J., ter Braak, C., Diks, C., Robinson, B., Hyman, J., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences & Numerical Simulation*, *10*, 271–288.

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sensing of Environment*, *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. *Journal of Geophysical Research*, *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. *Journal of Hydrometeorology*, *14*, 650–660.