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/$ – see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
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.