Spatial distribution and possible sources of SMOS errors at the global scale

Delphine Leroux, Yann H. Kerr, Philippe Richaume, Rémy Fieuzal

AUTHOR QUERY FORM

ELSEVIER	**Journal: RSE** **Article Number: 8582**	**Please e-mail or fax your responses and any corrections to:** **E-mail: Corrections.ESCH@elsevier.spitech.com** **Fax: +1 619 699 6721**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	Please confirm that given names and surnames have been identified correctly.
Q2	Please check the telephone number of the corresponding author, and correct if necessary.
Q3	Highlights should consist of only 85 characters per bullet point, including spaces. However, the Highlights provided for this item exceed the maximum requirement; thus, they were not captured. Kindly provide replacement Highlights that conform to the requirement for us to proceed. For more information, please see the Guide for Authors.
Q4	Please note that the citation "section IV.A.2" was changed to "Section 4.1.2" and was linked to its corresponding section title. Please check and amend if necessary.
Q5	Please note that the data "of this parameter is not accurate, this could have a large impact on SMOS retrievals since more than a half of the SMOS error variance is explained by the texture variance. As a consequence, very accurate texture and land cover maps are needed at the global scale." was deleted. Please check if appropriate.
Q6	Please check the contribution title.
Q7	Please provide an update for reference "Leroux et al., submitted for publication".
	Please check this box if you have no corrections to make to the PDF file. ☐

Thank you for your assistance.

Remote Sensing of Environment xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

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

Spatial distribution and possible sources of SMOS errors at the global scale

Delphine J. Leroux [a,b,*], Yann H. Kerr [a], Philippe Richaume [a], Remy Fieuzal [a]

[a] CESBIO, Centre d'Etudes Spatiales de la Biosphere, Toulouse, France
[b] Telespazio, Toulouse, France

ARTICLE INFO

Article history:
Received 7 March 2012
Received in revised form 14 February 2013
Accepted 16 February 2013
Available online xxxx

Keywords:
Triple collocation
SMOS
Error structure
Soil moisture
Multiple linear regression
Analysis of variance

ABSTRACT

SMOS (Soil Moisture and Ocean Salinity) data have now been available for over two years and, as part of the validation process, comparing this new dataset to already existing global datasets of soil moisture is possible. In this study, SMOS soil moisture product was evaluated globally by using the triple collocation method. This statistical method is based on the comparison of three datasets and produces global error maps by statistically inter-comparing their variations. Only the variable part of the errors are considered here, the bias errors are not treated by triple collocation. This method was applied to the following datasets: SMOS Level 2 product, two soil moisture products derived from AMSR-E (Advanced Microwave Scanning Radiometer)–LPRM (Land Parameter Retrieval Model) and NSIDC (National Snow and Ice Data Center), ASCAT (Advanced Scatterometer) and ECMWF (European Center for Medium range Weather Forecasting). The resulting errors are not absolute since they depend on the choice of the datasets. However this study showed that the spatial structure of the SMOS was independent of the combination and pointed out the same areas where SMOS performed well and where it did not. This global SMOS error map was then linked to other global parameters such as soil texture, RFI (Radio Frequency Interference) occurrence probabilities and land cover in order to identify their influences in the SMOS error. Globally the presence of forest in the field of view of the radiometer seemed to have the greatest influence on SMOS error (56.8%) whereas RFI represented 1.7% according to the analysis of variance from a multiple linear regression model. These percentages were not identical for all the continents and some discrepancies in the proportion of the influence were highlighted: soil texture was the main influence over Europe whereas RFI had the largest influence over Asia.

1. Introduction

Soil moisture is one of the most important variables regarding seasonal climate prediction as it plays a major role in the mass and energy transfers between the soil and the atmosphere. Several studies show the importance of the soil moisture for climate change studies (Douville & Chauvin, 2000), surface-atmosphere interactions (Koster et al., 2004), weather forecast (Drusch, 2007) or agriculture applications (Shin et al., 2006). Since August 2010, soil moisture is considered as an Essential Climate Variable (ECV) by the World Meteorological Organization (World Meteorological Organization et al., 2010).

Recently, satellite missions specially designed for monitoring soil moisture have been implemented (Soil Moisture and Ocean Salinity (SMOS), Kerr et al. (2010)) and proposed (Soil Moisture Active Passive (SMAP), Entekhabi et al. (2010)). SMOS was successfully launched by the European Space Agency in November 2009 and SMAP is scheduled to launch in October 2014 by the National Aeronautics and Space Administration. Both satellite instruments are designed to acquire data at the most suitable frequency for soil moisture retrieval (1.4 GHz, Kerr et al. (2001)). SMOS provides a global map of soil moisture every three days at a nominal spatial resolution of 43 km with an accuracy goal of 0.04 m^3/m^3.

Several approaches were developed to retrieve soil moisture using higher frequencies that have been the only options until now. These include the Scanning Multichannel Microwave Radiometer (SMMR, 1978–1987, Owe et al. (2001)), the Special Sensor Microwave/Imager (SSM/I, 1987-, Owe et al. (2001)), the Advanced Microwave Scanning Radiometer-Earth observation system (AMSR-E, 2002–2011, Owe et al. (2001), Njoku et al. (2003)), WindSat (2003, Li et al. (2010)), the Advanced Scatterometer (ASCAT, 1991-, Naeimi et al. (2009), Wagner et al. (1999)). Although their lowest frequencies (5–20 GHz) are not the most suitable for soil moisture retrievals (higher sensitivity to vegetation and atmosphere), they remain a valuable time series for the period of 1978 until now.

Currently, numerous studies are underway on the validation of SMOS soil moisture product with in situ measurements and estimates of other sensors and models. Al Bitar et al. (2012) used the Soil Climate Analysis Network (SCAN, Schaefer et al. (2007)) and the Snowpack Telemetry (SNOTEL) sites in North America to compare SMOS soil moisture retrievals and ground measurements. This study showed that SMOS soil moisture had a very good dynamic response but tended to underestimate the values. However the new versions of the SMOS

* Corresponding author at: CESBIO, Centre d'Etudes Spatiales de la Biosphere, Toulouse, France. Tel.: +33 674687451.
E-mail address: delphine.leroux@cesbio.cnes.fr (D.J. Leroux).

0034-4257/$ – see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.rse.2013.02.017

product (V4 & V5) significantly improved the general results. Jackson et al. (2012) studied SMOS soil moisture and vegetation optical depth over four watersheds in the U.S. They concluded that SMOS almost met the accuracy requirement with a RMSE of 0.043 m^3/m^3 in the morning and 0.047 m^3/m^3 in the afternoon whereas the vegetation optical depth retrievals were not reliable yet for use in vegetation analyses. Leroux et al. (submitted for publication) compared SMOS data with other satellite and model output products over the same four watersheds for the year 2010. It showed that SMOS soil moisture data were closer to the ground measurements than the other datasets and even though SMOS correlation coefficient was not the best, the bias was extremely small.

All the validation studies were performed over a few points and global conclusions cannot be drawn from single point comparisons. Moreover in situ measurements are not available at the global scale. As a second step in the validation process of SMOS soil moisture product, it is necessary to compare SMOS data to other satellite and model output products at the global scale to identify the region where the datasets differ or agree. To perform such global inter-comparison studies, statistical methods are needed. Triple collocation is a statistical method that compares three datasets and provides relative error maps as results. It has been widely used in the past in various environmental fields: sea surface winds (Caires & Sterl, 2003; Guilfen et al., 2001; Stoffelen, 1998), sea wave height (Caires & Sterl, 2003; Janssen et al., 2007), and soil moisture (Dorigo et al., 2010; Loew & Schlenz, 2011; Miralles et al., 2010; Scipal et al., 2008, 2010). More recently, the triple collocation method has been applied in the South West region of France to SMOS soil moisture retrievals with active microwave sensor retrievals and model simulations (Parrens et al., 2011). From the 40 to 72 common dates in 2010 that have been used in the triple collocation, it was found that the soil moisture retrievals from the active sensor (ASCAT) gave better results for this particular region with a relative error of 0.031 m^3/m^3 whereas SMOS had a relative error of 0.045 m^3/m^3. However the triple collocation only treats the variable part of the errors of the datasets since it compares their variances.

The two goals of this study are to evaluate the relative accuracy of SMOS soil moisture product compared to other global soil moisture products and to link this relative accuracy to physical parameters. For this purpose, triple collocation was applied to satellite products SMOS, AMSR-E (LPRM and NSIDC products), ASCAT and the ECMWF model product in 2010. For the second objective of this paper, ANOVA (Analysis of Variance) and CART (Classification And Regression Tree) analyses have been realized.

The major motivation to perform a classification of the SMOS relative errors is to provide to SMOS users a relative error estimation depending on the specifications of their region of interest. Given a set of parameter values (soil texture, land cover, etc.), it is thus possible to estimate the relative SMOS error by going through the classification tree. It also allows the user to know what the relative performance of SMOS is over their region compared to the soil moisture data sets. Another motivation would be at the algorithm level. The classification is realized at the global and continental scale and it is then possible to highlight the most influent parameters on the SMOS relative error, which represents valuable information for the SMOS Level 2 soil moisture team for future improvements.

The datasets used in this study are presented in Section 2 and the triple collocation method is introduced in Section 3. The SMOS error map was then related to physical parameters (soil texture, land cover and RFI (Radio Frequency Interference) probability maps) in order to understand better what caused SMOS largest errors (Section 4).

2. Description of the datasets

2.1. SMOS satellite product

The Soil Moisture and Ocean Salinity (SMOS, Kerr et al. (2010b)) satellite was launched in November 2009. This is the first satellite specially dedicated to soil moisture retrieval over land with an L-band passive radiometer (1.4 GHz, Kerr et al. (2001)). SMOS provides global coverage in less than 3 days with a 43 km resolution. The satellite is polar orbiting with equator crossing times of 6 am (local solar time (LST), ascending) and 6 pm (LST, descending). It is generally assumed that at L-band the signal is mainly influenced by the soil moisture contained in the top 2.5–5 cm of the soil on average.

SMOS acquires brightness temperatures at multiple incidence angles, from 0° to 55° with full polarization mode. The angular signature is a key element of the retrieval algorithm that provides soil moisture and the vegetation optical depth through the minimization of a cost function between modeled and acquired brightness temperatures (Kerr et al., 2012; Wigneron et al., 2007). These products are known as Level 2 products (Kerr et al., 2012) and are available on the ISEA-4h9 grid (Icosahedral Snyder Equal Area, Carr et al. (1997)) whose nodes are equally spaced at 15 km. In this study, the SMOS Level 2V4 products were used.

In SMOS algorithm, many physical parameters are involved in the form of auxiliary data (Kerr et al., 2010a) and they all play a very important role since the algorithm is applied differently according to their values (Kerr et al., 2010a, 2012). One of the unique features of SMOS algorithm is in the consideration of the heterogeneity inside the field of view of the radiometer. Around each pixel, an extended grid of 123 × 123 km at a 4 km resolution is defined to quantify the heterogeneity seen by the radiometer. Each pixel of this extended grid belongs to one of the ten following land cover classes (aggregated from ECOCLIMAP land cover ecosystems, Masson et al. (2003)): low vegetation, forest, wetland, saline water, fresh water, barren, permanent ice, urban area, frost and snow. The frost and snow classes have been disregarded in this study because they evolve over time and only average values for 2010 will be compared. To account for the antenna pattern of the instrument, a weighting function is applied.

Despite the fact that the SMOS frequency band is protected from emission by the laws, there exist some interferences all over the globe. Some regions are however more affected than others, i.e., China, and Western Europe at the beginning of 2010. More and more efforts are dedicated to suppress these interference sources but the impact of the RFI on the signal and on the soil moisture retrievals cannot be ignored. The RFI occurrence probability maps that have been used in this study are the average of the 3-month occurrence maps derived from the Level 1C SMOS data (brightness temperatures) and are available in the Level 2 products. It should be noted that 15% of the pixels on Earth are so affected by RFI that no retrieval attempt has been realized, and even more if we count the number of pixels where the inversion has failed because of the RFI presence.

2.2. AMSR-E satellite products

The Advanced Microwave Scanning Radiometer — Earth Observing System (AMSR-E) was launched in June 2002 on the Aqua satellite and stopped producing data in October 2011. This radiometer acquired data at a single 55° incidence angle and at 6 different frequencies: 6.9, 10.7, 18.7, 23.8, 36.5 and 89.0 GHz, all dual polarized. The crossing times were 1:30 am (LST, descending) and 1:30 pm (LST, ascending).

There are several products available using AMSR-E data to estimate soil moisture. The soil moisture product provided by the National Snow and Ice Data Center (NSIDC) is obtained from an iterative inversion algorithm using the 10.7 GHz and 18.7 GHz channels (Njoku et al., 2003). Initially, this algorithm was developed for 6.9 GHz and 10.7 GHz but due to the presence of RFI, the 6.9 GHz frequency was not usable for monitoring environmental parameters. Land surface parameters like soil moisture and vegetation optical depth are provided on a 25 km regular grid (Njoku, 2004). The

Please cite this article as: Leroux, D.J., et al., Spatial distribution and possible sources of SMOS errors at the global scale, *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2013.02.017

Level 3 DailyLand V6 products were used in this study (referred as the AMSR-E(NSIDC) product thereafter).

The Land Parameter Retrieval Model (LPRM, Owe et al. (2001)) retrieves the soil moisture and the vegetation optical depth using a combination of the 6.9 and 10.7 GHz frequencies (10.7 GHz acquisitions were used in the areas of the world where 6.9 GHz was polluted by RFI) and 36.5 GHz to estimate the surface temperature. This algorithm is based on a microwave radiative transfer model with a prior information about soil characteristics. The Level 3 AMSR-E v03 Grid products are available on a 0.25° × 0.25° grid only for the descending orbit (referred as the AMSR-E(LPRM) product thereafter). These data have been beforehand quality-controlled and the contaminated estimates due to high topography, extreme weather conditions such as snow have been flagged and have not been considered in this study.

In order to compare these datasets properly with SMOS retrievals, the AMSR-E(LPRM) and AMSR-E(NSIDC) soil moisture data have been interpolated over the SMOS grid.

2.3. ASCAT satellite product

The Advanced Scatterometer (ASCAT) has been launched in October 2006 on the MetOp-A satellite as a follow-on to the ERS (European Remote Sensing) scatterometer which started operating in 1991. It has been acquiring data in C-band (5.3 GHz). The scatterometer is composed of six beams: three on each side of the satellite track with azimuth angles of 45°, 90° and 135° (incidence angles are in a range of 25° to 64°) and generates two swaths of 550 km each with a spatial resolution of 25 or 50 km (in this study the 25 km resolution was used). The crossing times are 9:30 pm (resp. 9:30 am) LST for the ascending (resp. descending) orbit.

Soil moisture is retrieved using the Technische Universitat Wien soil moisture algorithm (Naeimi et al., 2009; Wagner et al., 1999) which corrects for the effect of vegetation and then retrieves an index ranging from 0 (dry) to 1 (wet) accounting for the top 2 cm relative soil moisture.

ASCAT data have also been interpolated over the SMOS grid, such that all the datasets are comparable on the same grid.

2.4. ECMWF model product

The European Center for Medium range Weather Forecasting (ECMWF) provides medium range global forecasts and in this context, some environmental variables including the soil temperature, the evaporation or the soil moisture are produced.

The SMOS Level 2 processor uses a custom made climate data product from ECMWF that is used to set the initial values in the cost function solution and for fixed parameters in the algorithm to compute the different contributions of each land cover class. This product from ECMWF is considered an internal SMOS product as it is specially computed at SMOS overpasses by interpolating in space and time the ECMWF forecast products over the SMOS grid. This custom ECMWF product also has the same spatial resolution as SMOS and has been used in this study. The ECMWF soil moisture represents the top 7 cm below the surface.

3. Triple collocation

3.1. Theory

As in Stoffelen (1998) and Dorigo et al. (2010), we propose an approach where the three datasets θ_i are linearly linked to the hypothetical truth θ with a bias term r_i and a scale factor s_i. The triple collocation method consists in estimating the errors ε_i. These errors are relative to the hypothetical truth θ and are comparable with each other since they are relative to the same quantity. However they are not absolute errors. One dataset is arbitrarily chosen as the reference dataset so that $r_1 = 0$ and $s_1 = 1$. Thus, the truth θ and the first product θ_1 cannot be identical since the error term ε_1 remains:

$$\begin{cases} \theta &= \theta_1 + \varepsilon_1 \\ \theta &= r_2 + s_2\theta_2 + \varepsilon_2 \\ \theta &= r_3 + s_3\theta_3 + \varepsilon_3 \end{cases}. \quad (1)$$

By taking the average over the year (<.>) and assuming that the errors ε_i have a zero mean, the following expressions of the mean hypothetical truth are obtained:

$$\begin{cases} <\theta> &= <\theta_1> \\ <\theta> &= r_2 + s_2<\theta_2> \\ <\theta> &= r_3 + s_3<\theta_3> \end{cases}. \quad (2)$$

Let θ'_i be the anomaly term of the dataset i that is defined by $\theta'_i = \theta_i - <\theta_i>$. By subtracting Eqs. (1) and (2), the bias terms r_i disappear:

$$\begin{cases} \theta' &= \theta'_1 + \varepsilon_1 \\ \theta' &= s_2\theta'_2 + \varepsilon_2 \\ \theta' &= s_3\theta'_3 + \varepsilon_3 \end{cases}. \quad (3)$$

The anomalies θ'_i are assumed to be independent to the errors ε_i of the other datasets. Since $<\theta'_i>$ and ε_i are null (θ'_i is the anomaly to the mean and ε_i is a zero mean additive noise) and θ_i can be considered as a deterministic quantity, we finally have $<\theta'_i\varepsilon_i> = 0$. It also assumed that the errors are independent to each other: $<\varepsilon_i\varepsilon_j> = 0$. By taking the average and combining the lines, the scale factors and the mean square true anomaly can be derived:

$$\begin{cases} s_2 &= <\theta'_1\theta'_3> / <\theta'_2\theta_3> \\ s_3 &= <\theta'_1\theta'_2> / <\theta'_2\theta'_3> \\ <\theta'^2> &= <\theta'_1\theta'_2><\theta'_1\theta_3> / <\theta'_2\theta'_3> \end{cases}. \quad (4)$$

With Eq. (2) combined with Eq. (4), the bias terms can be computed:

$$\begin{cases} r_2 = <\theta_1> - s_2<\theta_2> \\ r_3 = <\theta_1> - s_3<\theta_3> \end{cases}. \quad (5)$$

By taking the square of Eq. (3) and its mean and by using Eqs. (4) and (5), we finally obtain the expressions of each averaged square errors ε_i that can be written as functions of known anomalies:

$$\begin{cases} <\varepsilon_1^2> &= \dfrac{<\theta'_1\theta'_2><\theta'_1\theta'_3>}{<\theta'_2\theta'_3>} - <\theta'^2_1> \\ <\varepsilon_2^2> &= \dfrac{<\theta'_1\theta'_2><\theta'_1\theta'_3>}{<\theta'_2\theta'_3>} - \dfrac{<\theta'_1\theta'_3>^2}{<\theta'_2\theta'_3>^2}<\theta'^2_2> \\ <\varepsilon_3^2> &= \dfrac{<\theta'_1\theta'_2><\theta'_1\theta'_3>}{<\theta'_2\theta'_3>} - \dfrac{<\theta'_1\theta'_2>^2}{<\theta'_2\theta'_3>^2}<\theta'^2_3> \end{cases}. \quad (6)$$

Since the errors ε_i have a zero mean, $<\varepsilon_i^2>$ can be interpreted as their variances. The dataset with the lowest error variance is considered as the most accurate among the three. The choice of which dataset is the reference (the first line in the equations) changes the values of the relative errors since the projection space changes, but it does not change the patterns of the errors, i.e., if dataset #2 gave poor results over Africa compared to the other datasets, it will still give poor results even as the reference, but with different error values. It is essential to note that these errors are strictly related to the choice of the triplet and cannot be compared to errors derived from another triplet.

Please cite this article as: Leroux, D.J., et al., Spatial distribution and possible sources of SMOS errors at the global scale, *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2013.02.017

3.2. Assumptions, requirements and methodology

Many assumptions have been made in the triple collocation method described in the previous section:

- the soil moisture products are linearly linked to the true soil moisture
- the errors are random with a zero mean and mutually independent.

The first assumption might not be true and it would require to add at least a second order in the starting Eq. (1). However, all the selected products are supposed to estimate the soil moisture so it is reasonable to assume that they are close to the truth by using a scale factor, a bias and an error term.

The errors ε_i are assumed to have a zero mean and to be mutually independent so that their covariances are null ($<\varepsilon_i\varepsilon_j> = 0$). Otherwise, we would need to add cross-correlation terms in Eq. (4) that would have to be arbitrarily estimated and would have repercussions in the final expressions of the error terms in Eq. (6). In order to avoid this case of dependency between the errors, the datasets were chosen very carefully: soil moisture products were derived from different algorithms or were based on acquisitions at different frequencies. This study does not consider triplets with the two ASMR-E products together (NSDIC and LPRM) since they were both derived from the same acquisitions. Thus, the triple collocation was applied to the following triplets: (SMOS–AMSR-E(LPRM)–ECMWF), (SMOS–AMSR-E(NSIDC)–ASCAT) and (SMOS–AMSR-E(LPRM)–ASCAT).

The triple collocation method is based on statistics and they are only reliable if the number of available samples is large enough. Scipal et al. (2010) determined that a minimum of 100 samples is required to apply the triple collocation. In this study, satellite and model data have been compared for 2010 and after combining the different orbits and swaths the minimum of 100 common dates are not satisfied. To get around this problem, we propose to collect the samples of the six closest neighbors (that are equally distant of 15 km from a central point on the ISEA grid) as if they were the samples of the central point.

By taking into account the six closest neighbors, geophysical variance is introduced. However, since the grid nodes are 15 km distant to each other, the degrees of freedom are not increased much. Nevertheless, 15 km should represent a reasonable distance to assume that the soil surface conditions are different and heterogeneous enough to contain a significant amount of information and thus increase the statistical power of the method.

Since bias would be interpreted as higher deviations, it is important to use non biased datasets when applying the triple collocation method. In this study, the triple collocation has been applied to the anomalies of the variables and not to the variables themselves. Since the anomalies have a zero mean by definition, there cannot be any systematic bias between them.

The triple collocation method can be summarized in four steps:

1. compute the anomalies for each point of the grid for the three datasets
2. compute the scaling factors with Eq. (4)
3. compute the bias with Eq. (5)
4. compute the variances of the errors with Eq. (6).

The results of the triple collocation are relative errors. It is important to keep in mind that these errors only represent the variable part of the total error for each tested dataset. Indeed the triple collocation uses the anomalies to compute these errors and the bias part are then not treated by this method.

3.3. Results

3.3.1. SMOS/AMSR-E(LPRM)/ECMWF

Fig. 1 shows the error maps of SMOS, AMSR-E(LPRM) and ECMWF. From this first triplet, SMOS did not perform better than the other two datasets in terms of variable error. SMOS had the lowest error for 17% of the points whereas LPRM was better for 44% and ECMWF for 39% (triple collocation was applied to 302,474 points in this case). The worst SMOS errors were located in East USA, North of North America, Europe, India and East Asia whereas the best performances were in West USA, North Africa, Middle East, central Asia and Australia. ECMWF gave very good results over Europe, South America, part of North America and India whereas LPRM best results covered West USA, some parts of Africa, Asia and West Australia.

3.3.2. SMOS/AMSR-E(NSIDC)/ASCAT

Fig. 2 shows the error maps of SMOS, AMSR-E(NSIDC) and ASCAT. With this triplet, SMOS performed better on 21% of the points against 35% for NSIDC and 44% for ASCAT (247,798 points globally) in terms of variable error. The total number of points taken into account for this specific triplet was lower than for the previous triplet since in this case three satellite datasets were implicated whereas the model ECMWF was part of the previous triplet (there was always a ECMWF soil moisture value for each SMOS or AMSR-E soil moisture retrieved value).

With this triplet again, SMOS performed better over North America and Australia, NSIDC gave better results in North Africa and Middle East whereas ASCAT was better in North Africa and central Asia. The weakness of the NSIDC product is its non-dynamic retrievals Gruhier et al. (2008), i.e., the NSIDC soil moisture time series are relatively flat. So over arid regions where there is no precipitation nor vegetation, NSIDC performed well but over the other regions, NSIDC did not give satisfying results. ASCAT had good performances over the entire globe except over Europe, East Australia, Sahel region and the USA.

3.3.3. SMOS/AMSR-E(LPRM)/ASCAT

Fig. 3 shows the error maps of SMOS, AMSR-E(LPRM) and ASCAT. SMOS performed better than the other datasets over 21% of the points against 34% for LPRM and 45% for ASCAT in terms of variable error and for a total of 210,368 points. Less points were available than with the previous triplet even if only the AMSR-E product was changed. LPRM and NSIDC are algorithmically different and LPRM retrieved in general less soil moisture than NSIDC. With this triplet again, SMOS had good performances over the same regions: North America, North Africa, Middle East, central Asia and Australia.

3.4. SMOS error distribution

The computation of SMOS global error through the triple collocation was performed with other satellite and model datasets. This study pointed out that SMOS gave very good results over North America, North Africa, Middle East, central Asia and Australia. In general, SMOS gave better results than the other products in North America, central Asia and Australia. These particular regions are known to be RFI free and might be an explanation of why SMOS performs better over these regions.

The triple collocation is a statistical method that only considers the variable part of the error and not the bias part since it uses the anomalies and not the products themselves. From previous validation studies (Leroux et al., submitted for publication), it has been shown that the SMOS soil moisture product had a very low bias compared to the other soil moisture products. But this low bias has no impact in the triple collocation results. Therefore it should be noted that even if the triple collocation results show that one product is better than another, it is only in terms of variable errors.

Even if SMOS was chosen as the scaling reference each time, the resulting error maps cannot be compared since the dates in common were not the same for all the triplets tested. However, the distribution of the SMOS error was very similar on the three maps. The regions where SMOS was good were the same for the three tests. Thus, these maps can be used to understand and find a link with other physical or algorithmic parameters such as soil texture, land use maps and RFI.

Please cite this article as: Leroux, D.J., et al., Spatial distribution and possible sources of SMOS errors at the global scale, *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2013.02.017

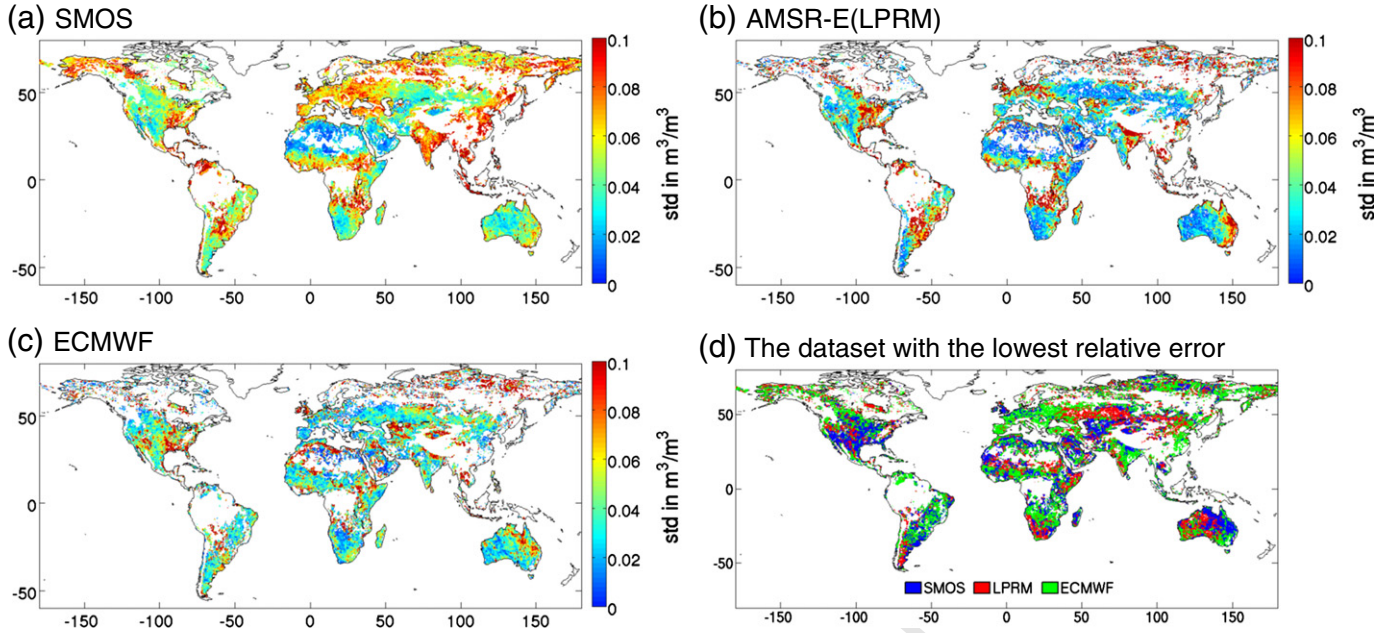

Fig. 1. Standard deviations of (a) SMOS errors, (b) AMSR-E(LPRM) errors, and (c) ECMWF errors. (d) shows the areas in which either SMOS (blue), AMSR-E(LPRM) (red) or ECMWF (green) shows the smallest error (variable part of the error). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. SMOS error analysis

In this section, possible links between SMOS errors and biophysical parameters are investigated. To identify which parameter was responsible for which proportion of the SMOS error, a multiple linear regression model with an analysis of variance (ANOVA) was realized. As a second step, a classification (CART) was performed in order to identify sets of parameters leading to either small or large SMOS errors. Another motivation to perform a CART analysis is to be able to estimate or predict SMOS error depending on the values of the investigated biophysical parameters.

The following list of parameters was investigated: percentage of sand (% sand), percentage of clay (% clay), mean RFI probability over 2010 (RFI), fraction of forest (FFO), wetlands (FWL), water bodies (FWP), salted water (FWS), barren (FEB), ice (FEI), and urban (FEU) as seen by the radiometer. The fraction of low vegetated area was not taken into account as an explanatory variable because it is a combination of the other fractions (sum of the other fractions subtracted to 1) and it would have become a constraint parameter. Moreover, the main goal is to identify the parameters that deteriorate the SMOS retrievals. Global and continental analyses were performed to identify regional behaviors and differences. This work only included standard errors from the triple

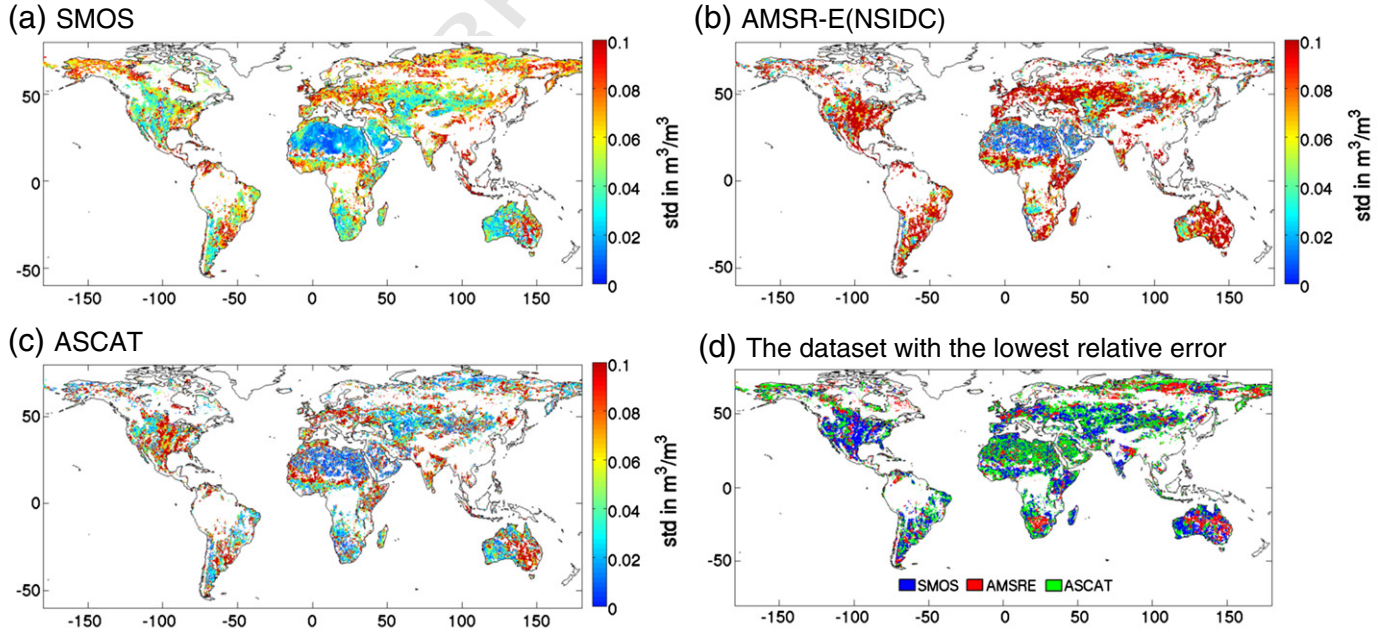

Fig. 2. Standard deviations of (a) SMOS errors, (b) AMSR-E(NSIDC) errors, and (c) ASCAT errors. (d) shows the areas in which either SMOS (blue), AMSR-E(NSIDC) (red) or ASCAT (green) shows the smallest error (variable part of the error). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Leroux, D.J., et al., Spatial distribution and possible sources of SMOS errors at the global scale, *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2013.02.017

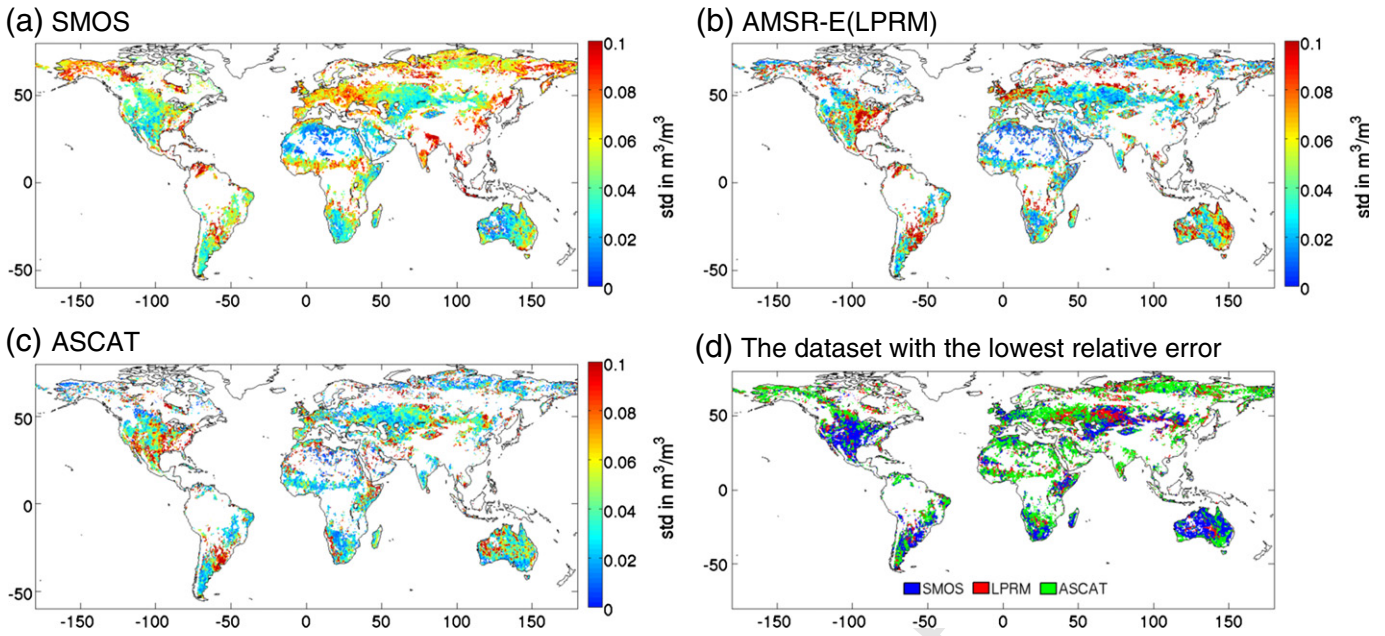

Fig. 3. Standard deviations of (a) SMOS errors, (b) AMSR-E(LPRM) errors, and (c) ASCAT errors. (d) shows the areas in which either SMOS (blue), AMSR-E(LPRM) (red) or ASCAT (green) shows the smallest error (variable part of the error). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

collocation method from 0 to 0.1 m^3/m^3 which represented 95% of the points.

The SMOS error map derived from the triple collocation with SMOS, AMSR-E(LPRM) and ECMWF was used in this part of the study. This triplet was chosen because it had the advantage to cover more points over the globe.

The mean values and the standard deviations of the SMOS error and of each parameter globally and for every continent have been computed and are shown in Fig. 4. As seen in Fig. 1, North Africa and Australia are the two continents where SMOS has the lowest errors: 0.040 m^3/m^3 and 0.043 m^3/m^3 respectively (Fig. 4). However, the errors over Australia are more homogeneous with the lowest standard deviation (0.015 m^3/m^3).

The percentage of clay is quite stable over all the continents (around 24%). South America is the region where this parameter is the most heterogeneous with a standard deviation of 10% whereas East Asia is the most homogeneous with a bit more than 5%. The percentage of sand is more heterogeneous than the clay. South Africa and Australia are the regions with the highest mean percentages of sand (55.9% and 55.6% respectively) whereas East Asia is the most homogeneous region with a standard deviation of only 8.7%. The radio frequency interferences do not affect equally all the regions. Whereas Australia, America and South Africa are almost not concerned, Europe, North Africa and Asia are highly contaminated. The fraction of forest, as estimated by the ECOCLIMAP land cover (and supposed to be seen by the radiometer), is very heterogeneous, especially over America and South Africa. As expected for desert regions, there is not a lot of forest in North Africa and Australia. The other fractions do not represent a large proportion of what is modeled by the ECOCLIMAP land cover. Though, it can be noted that there are more wetlands in Europe and East Asia; more water bodies in North America; more barren soils in North Africa; and slightly more cities are modeled by ECOCLIMAP in Europe and North America.

4.1. Multiple linear regression and analysis of variance (ANOVA)

The multiple linear regression is a statistical method that studies the relation between a variable Y and several explanatory variables $X_1, X_2, \cdots, X_n$. This method only accounts for linear relationship of the following form:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n + \varepsilon \quad (7)$$

$$\begin{aligned} <\varepsilon^2_{smos}> = {} & \alpha_o + \alpha_{clay}X_{clay} + \alpha_{sand}X_{sand} + \alpha_{RFI}X_{RFI} + \alpha_{FFO}X_{FFO} \\ & + \alpha_{FWL}X_{FWL} + \alpha_{FWP}X_{FWP} + \alpha_{FWS}X_{FWS} + \alpha_{FEB}X_{FEB} \\ & + \alpha_{FEI}X_{FEI} + \alpha_{FEU}X_{FEU}. \end{aligned} \quad (8)$$

If the explanatory variables are not in the same unit as the variable Y, which is the case in this study, it is absolutely necessary to normalize each variable. All the α parameters are then computed with the least square method.

In order to only keep the relevant explanatory variables, the variables explaining less than 1% of the SMOS error variance according to the ANOVA were removed from the regression model. The correlation coefficients of the multiple linear regression model are indicated in Table 1 with all the variables.

ANOVA is a statistical method used to study the modification of the mean of a variable according to the influence of one or several explanatory variables. The proportion of the influence of each variable on SMOS error was computed from the linear regression model (Fig 5, Table 2). A parameter having a negative influence means that the greater value this parameter has, the greater the error is, i.e., the error evolves with this particular parameter. As a contrary, a positive influence is: the greater the value of the parameter is, the lower the error is.

4.1.1. Global results

At the global scale, the multiple linear regression and the ANOVA showed that more than half (57%) of the variation of SMOS error was due to the variation of the forest fraction, with a negative influence, i.e., the more forest the radiometers see, the larger the error is. The second most important explanatory variable is the percentage of sand, representing 22% of the total variation with a negative influence. The third influent variable is the fraction of wetlands (7% negatively). Along with the large proportion of the influence of these three variables (86%) on SMOS error, it also means that if one of these variables was not

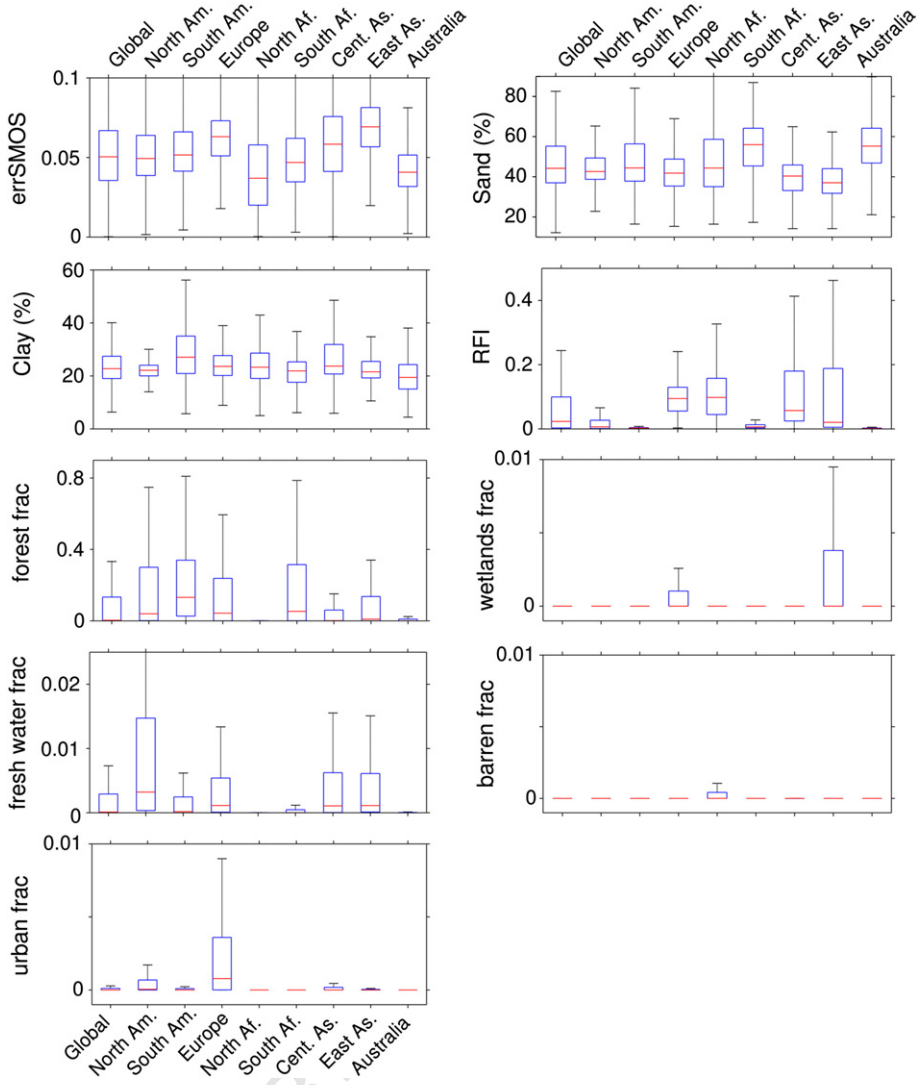

Fig. 4. Mean and standard deviation of each parameter (except the salted water and the ice fractions that were null or very close to 0 for all the continents), at the global scale and per continent. The SMOS error is the result of the triple collocation from the SMOS–AMSR-E(LPRM)–ECMWF triplet.

correctly estimated, this would have had a big effect on SMOS final error.

RFI did not have a big influence at the global scale (less than 2%). RFI influence can be very high but more at the regional scale than at the global scale.

The only positive influence was the fraction of barren soil (FEB), i.e., the more barren there is, the lower the error is. All these results were expected since fractions of water, wetlands, urban, RFI or forest disturb the signal and make the soil moisture retrieval more challenging.

4.1.2. Continental results

At the continental scale, there are some discrepancies. The fraction of forest (FFO), which was the most influent factor at the global scale, was still the most influent variable for most of the continents but did not represent more than 10% for Europe and East Asia whereas it represented at least 30% for the other continents. The RFI influence was also different, its largest influence being in Asia, Europe and North Africa.

Over North America, the fraction of forest represented 67% of the variation of SMOS error. This can be seen in Figs. 1, 2 and 3, SMOS errors were higher in the northern part where there are more trees. The second and third influent variables were the fraction of urban and the percentage of sand with 8% and 7% respectively. Over this continent, the unexpected result was the sign of the influence of the percentage of sand, which was positive. This positive influence of the sand was also found for Europe and South America.

For South America, the percentage of sand and the fraction of forest had almost the same proportion of influence, 39% and 43% respectively. The soil texture was playing a major role in the SMOS error for this continent. Soil texture also had a large influence in Europe (46%), South Africa (46%), East Asia (39%) and Australia (54%). However, the soil texture influence was not always represented the same way: mostly by the clay with positive influence for Europe (43%) and East Asia (30%), by the clay with a negative influence for South Africa (38%), by the sand negatively for Australia (53%) and by the sand positively for South America (39%) whereas the global soil texture influence was represented by the sand negatively (22%).

Table 1
Correlation coefficients of the multiple linear regression for global and continents. R_{tot} represents the statistics when all the explanatory variables are considered.

	Global	North America	South America	Europe	North Africa
R_{tot}	0.455	0.425	0.470	0.540	0.387
		South Africa	Central Asia	East Asia	Australia
R_{tot}		0.606	0.523	0.451	0.610

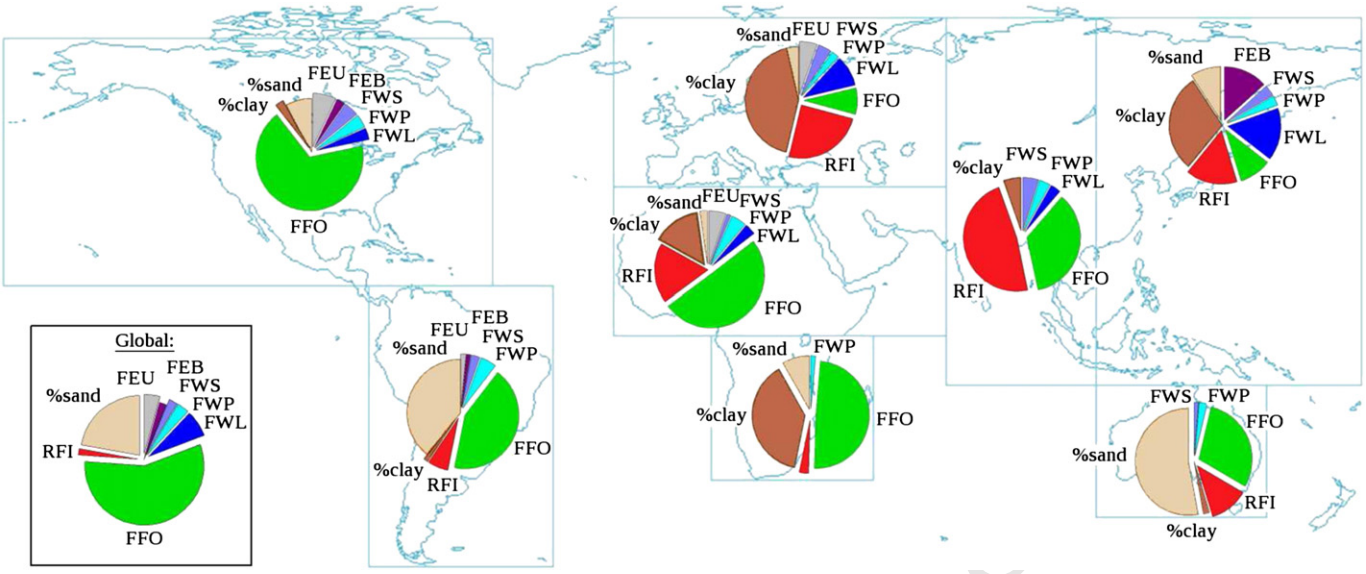

Fig. 5. Proportion of the influence of each variable on SMOS error for each continent and at the global scale: percentage of sand and clay (% sand, % clay), mean RFI probability in 2010 (RFI), fraction of forest (FFO), fraction of wetlands (FWL), fraction of water bodies (FWP), fraction of salted water (FWS), fraction of barren (FEB) and fraction of urban (FEU). A detached slice represents a negative influence and an attached slice a positive influence.

The RFI influenced the SMOS error more significantly in central Asia (47%), Europe (24%), North Africa (18%) and East Asia (16%). Regarding East Asia, the statistics about RFI might have been compromised because there were less points in this area due to the very strong RFI and no soil moisture value was retrieved. But over the other three cited continents, RFI were not negligible, especially for central Asia where it was even the first influent factor.

Nevertheless, the influence of the forest fraction over Australia can be surprising. This continent does not have a lot of forests but SMOS error values were very low, except near the coasts where there are some forests. That is why the forest fraction was playing such a large influence in Australia.

4.2. Classification and regression trees (CART)

The main goal of the classification process is to summarize and predict a variable by a set of explanatory variables. With the values of the different explanatory parameters, it will be possible to estimate the SMOS error by going through the resulted classification tree defined by the CART.

The classification was performed recursively by investigating each variable and each possible threshold value to create the most homogeneous classes. Let x_i be a variable and s_j a value of this variable, then a partitioning where $x_i < s_j$ and $x_i \geq s_j$ splits the dataset into two disjoint sets or partitions. Partitions are created until a certain level of homogeneity is obtained.

The homogeneity is represented by the Gini index. Let k be the classes ($k = 1,\cdots,C$; where C is the total number of classes), then the Gini index of the partition A is defined as follows:

$$I(A) = 1 - \sum_{k=1}^{C} p_k^2 \quad (9)$$

where p_k are the fractions of the observations belonging to the class k in the partition A. $I = 0$ represents the perfect homogeneity where only one class is present in the partition, $I = (C-1)/C$ represents pure heterogeneity where all the classes appear equally in the partition.

If the sum of the Gini indexes of the two possible partitions is less than the Gini index of the partition to be split, then the homogeneity has been improved and the partitioning is accepted and realized.

In order to avoid to split too much and to finally only capture the noise (over-learning), the resulting tree is pruned. The pruning level is controlled by the complexity parameter (cp). Any split that does not decrease the overall lack of fit by a factor of cp is not attempted, i.e. the overall R^2 must increase by cp at each step. Thus, computing time is saved by pruning off splits or partitions that are considered as not worthwhile. In this study the complexity parameter cp was arbitrarily set to 0.01 for global and continental studies.

Table 2
Proportion of the SMOS error variance explained by each explanatory variable: percentage of sand and clay (% sand, % clay), mean RFI probability in 2010 (RFI), fraction of forest (FFO), fraction of wetlands (FWL), fraction of waterbodies (FWP), fraction of salted water (FWS), fraction of barren (FEB) and fraction of urban (FEU).

	% sand	% clay	RFI	FFO	FWL	FWP	FWS	FEB	FEI	FEU
Global	22.2	0.1	1.7	56.8	7.2	3.3	2.4	2.0	0.1	4.2
North Am.	7.8	2.6	0.6	66.9	3.5	4.0	4.5	2.5	0.6	7.0
South Am.	39.3	1.4	6.2	42.8	0.1	4.8	2.5	1.2	0.1	1.6
Europe	3.5	42.7	24.4	7.9	9.3	2.4	4.2	0.2	0.2	5.2
North Af.	2.3	14.7	18.1	50.2	3.4	4.7	1.2	0.1	–	5.3
South Af.	8.1	37.9	2.8	48.2	0.6	1.4	0.3	0.5	–	0.2
Cent. Asia	0.8	5.3	47.0	34.7	3.1	3.4	4.7	0.3	0.2	0.5
East Asia	9.2	29.7	15.6	9.7	15.8	2.9	3.4	13.2	0.0	0.5
Australia	52.5	1.9	11.8	29.5	0.2	2.4	1.3	0.3	–	0.1

4.2.1. Global results

A regression tree was calculated for the entire world (Fig. 6) and ended with six leaves. These leaves represent the end of several decisions leading to a mean value for the SMOS error. Five decisions were made and were considered as the most discriminant decisions to split the entire dataset by the CART algorithm. The decisions resulting in the lowest SMOS error (0.031 m^3/m^3) were: FFO = 0 followed by FWP = 0. Over the entire globe, the characteristics of very low SMOS error value were no forest and no water body. These two parameters with their threshold values differentiated the best global dataset. As a contrary, the decisions leading to the worst SMOS error value were: FFO > 0 and RFI ≥ 0.07. The combination of these two decisions leads in average to a very high value of SMOS error (0.070 m^3/m^3).

The first split was also very discriminant in terms of average SMOS error since if the statement FFO = 0 was true then the mean SMOS

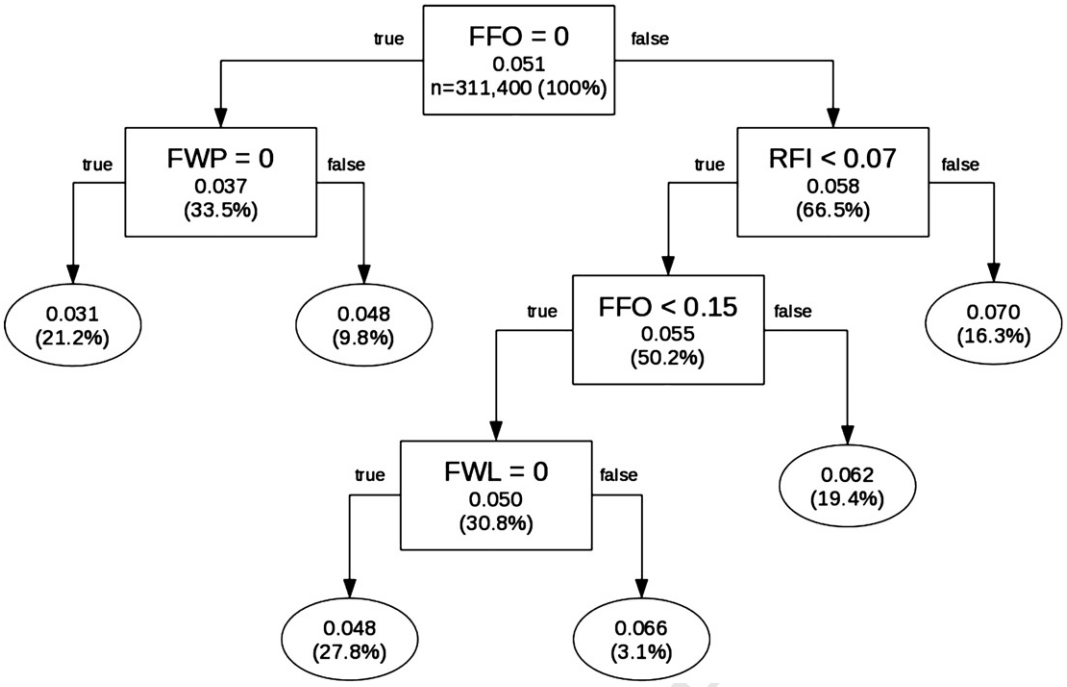

Fig. 6. Regression tree of the SMOS error at the global scale. In the rectangles are written the splitting condition, the mean value of the SMOS error before applying this decision and the number of points that are concerned. If the condition is fulfilled then the branch *true* of the tree is followed. The end of each branch is concluded by a leaf represented by a circle with the mean SMOS error value and the number of points.

error value was 0.037 m^3/m^3 whereas if it was false, the mean value was 0.058 m^3/m^3. The decision true concerned one third of the points whereas the decision *false* two thirds. If the first statement was *false*, then the next decision was about the RFI value. The threshold proposed at this stage of the tree was 0.07, i.e., the number of acquisitions for a point that has been polluted by RFI in 2010 is 7%. If this threshold was not respected, then it resulted in a very high SMOS error value (0.070 m^3/m^3 in average). If the probability was below 0.07 then the next decision was again about the forest fraction: FFO < 0.15. This second time, the threshold value was higher (15% of the scene being covered by forest). If this statement was true, then a last decision needed to be made about the fraction of wetlands seen by the radiometer: FWL = 0. This last decision was important as well because if this was *true*, the mean SMOS error was 0.048 m^3/m^3 whereas if it was *false*, it was 0.066 m^3/m^3, which is about 50% more.

The two analysis (ANOVA and CART) gave complementary results. The ANOVA computed the part of the SMOS error variance that was explained by each explanatory variable. The CART analysis computed which variable and with which value, the SMOS error dataset can be split so that sub-datasets can be explained differently. That also created a list of decisions and the mean SMOS error value depended on these decisions. So even if some explanatory variables did not have a large impact in the ANOVA, they can have a large influence in how to split the dataset. For example, from the ANOVA, RFI was only explaining 1.7% of the total variability whereas it was the second decision parameter that lead to the largest SMOS error from the CART analysis.

4.2.2. Continental results

Regression trees as in Fig. 6 were computed for each continent and the decisions which lead to the lowest and largest SMOS error values are summarized in Table 3. The CART was stopped the same way as in the global case with a complexity parameter (*cp*) set to 0.01 in order to make the conclusions comparable. The forest fraction (FFO) was involved in all the decisions and was in most cases the first variable that was used to split the dataset.

North Africa, South Africa and Australia were the three regions where the last branch of the CART resulted in a very low average SMOS error value. Over North Africa, 50% of the points verified the conditions FFO = 0, FEU = 0, and FWP = 0 and resulted in a SMOS average error value of 0.026 m^3/m^3; over South Africa, 17% of the points verified FFO = 0 for a mean error of 0.031 m^3/m^3 and over Australia 40% of the points verified FFO = 0, % clay < 19.1 for a mean error of 0.033 m^3/m^3.

East Asia and central Asia were the two regions where the CART identified the largest average error values. Over East Asia, the conditions RFI ≥ 0.20, % clay ≥ 22.2 were fulfilled by 21% of the points for a mean error value of 0.080 m^3/m^3 and over central Asia, 31% of the points verified the conditions RFI ≥ 0.11, FFO > 0 for a mean error value of 0.076 m^3/m^3.

The classification of South America is coherent with the ANOVA results (Section 4.1.2 and Fig. 5) since for this region, the sand had a positive influence, it is then natural to have the lowest SMOS error obtained with a percentage of sand above 40%.

5. Conclusions and perspectives

SMOS soil moisture has been available for almost three years starting January 12, 2010. The first year of data was used to evaluate the SMOS error structure at the global scale. For this purpose, the triple collocation was applied to SMOS and to two other global soil moisture datasets among AMSR-E(LPRM), LPRM(NSIDC), ASCAT and ECMWF. The error maps showed that SMOS gave better results (lower errors) than the other datasets over North America, Australia and central Asia.

Even though the error values from one triplet cannot be compared to the results from another triplet, the structure of SMOS error at the global scale exhibited similar patterns on the three error maps: high error values over East and South Asia, Europe and highly vegetated areas (Amazon, central Africa and extreme North of America); low error values over North Africa, extreme South Africa, central Asia, Australia and part of North America. By obtaining these three similar

Please cite this article as: Leroux, D.J., et al., Spatial distribution and possible sources of SMOS errors at the global scale, *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2013.02.017

Table 3
Sets of parameter values that lead to the lowest and largest SMOS error from the classification process. The parameters are written in the same order as in the tree and the mean SMOS error values are indicated in parenthesis.

	Lowest SMOS error (m^3/m^3)	%	Largest SMOS error (m^3/m^3)	%
Global	FFO = 0, FWP = 0 (0.031)	21.8	FFO > 0, RFI ≥ 0.07 (0.070)	16.3
North Am.	FFO = 0, FWP = 0 (0.037)	12.1	FFO > 0, % clay ≥ 20.4 (0.067)	17.2
South Am.	FFO = 0, RFI < 6.10^{-3}, % sand ≥ 40, FEU = 0 (0.043)	23.9	FFO > 0, % sand < 44, FWP > 0 (0.070)	11.8
Europe	FFO = 0, RFI < 0.16, FWL = 0, FEU = 0 (0.038)	9.6	FFO > 0, RFI ≥ 0.06 (0.068)	56.3
North Af.	FFO = 0, FEU = 0, FWP = 0 (0.026)	50.6	FFO > 0 (0.064)	22.1
South Af.	FFO = 0 (0.031)	16.7	FFO ≥ 0.32, % clay < 17.4 (0.068)	18.8
Cent. Asia	RFI < 0.11, FFO = 0 (0.038)	24.0	RFI ≥ 0.11, FFO > 0 (0.076)	30.6
East Asia	RFI < 0.20, FFO = 0 FWL = 0 (0.049)	10.0	RFI ≥ 0.20, % clay ≥ 22.2 (0.080)	20.6
Australia	FFO = 0, % clay < 19.1 (0.033)	39.7	FFO ≥ 0.24 (0.061)	8.6

global maps (related to three different triplets), we can assume that the triple collocation method is robust to the choice of the triplet.

The second goal of this study was to relate the SMOS error values and structures to physical and algorithmic parameters: the fractions of forest, wetland, water body, salted water, barren, ice, and urban in the radiometer field of view defined by the ECOCLIMAP land cover, the soil texture (percentages of sand and clay) and the probability of RFI (radio frequency interference) in 2010. A multiple linear regression model was computed for each continent and globally. An analysis of variance (ANOVA) determined the proportion of SMOS error variance explained by each parameter according to the multiple linear regression model.

At the global scale, the fraction of forest (FFO) explained most of the SMOS error variance (57%) followed by the texture with the percentage of sand (22%). The more forest or sand we assume the radiometer sees, the larger the SMOS error is. These proportions vary a lot depending on the continent. Over Europe, the proportion of the variance explained by the texture is around 46% followed with the RFI with 24% whereas over North America, the FFO explained 67%. A special care needs to be brought to the forest and sandy regions.

The presence of RFI is a serious issue for SMOS retrievals and even though the proportion of explained SMOS error variance was high over several continents (Asia, Europe, North Africa), the global proportion explained by RFI remained extremely low (2%). This can be explained by the fact that the points that were highly infected by RFI did not have any soil moisture retrieval and were not then considered in that study.

In order to identify the set of parameters that could lead to very high or very low SMOS errors, a classification and a regression tree (CART) was computed for each continent and globally. A CART is determined by the way to split a dataset so that the subsets are more similar than the original dataset. At the global scale, and for most of the continents, the FFO was the first parameter to split the original dataset into two subsets. Except for South America, the value of FFO for the first decision was very low (maximum 0.07) so even a small fraction of forest can lead to a very high error value. For central Asia and East Asia, the RFI was the first parameter that leads to the first split. These two regions were extremely affected by RFI.

Frozen and snow covered grounds have not been taken into account and can play a major role in the retrievals, especially in Northern latitudes. Since the SMOS algorithm uses the ECMWF values to compute the contributions of each class represented in the radiometer field of view, it is possible that ECMWF influences SMOS retrievals and that the anomalies of these two datasets are not uncorrelated. Nevertheless, the global results and error patterns of the SMOS relative error are almost the same for all the triplets of the triple collocation, so the final analysis remains the same. One point that has not been covered in this study is the impact of the uniformity or the non-uniformity of the observed area on the retrievals. It is expected though that if the scene is very heterogeneous (many different landcover classes combined), the contributions of each class need to be evaluated and thus introduce more error in the retrieved soil moisture. Finally, a reprocessing of SMOS brightness temperatures is underway and should lead to improved soil moisture retrievals.

From the results of this study, several hints for future improvements concerning SMOS soil moisture algorithm have emerged. The variation of the forest fraction explains more than half of the error variance at the global scale and improving the parameterization of the forest model is definitely needed. The second most important parameter is the soil texture. However, the role of the soil texture is not clear according to the results of this study since it has a positive influence over America or Europe and a negative influence over Africa, Asia or Australia. In the soil moisture retrieval process, the soil texture is mostly taken into account in the computation of the soil dielectric constant. Level 2 soil moisture processor used the Dobson dielectric constant model (Dobson et al., 1985) and since April 2012, the Mironov model (Mironov & Fomin, 2009) has been used in the V5.5 product. This change in the model should modify the presented results and improvements are expected especially in the sandy regions where the Mironov model is known to perform better.

Acknowledgments

The authors would like to thank the French CNES/TOSCA program and Telespazio France for their funding and support.

References

Bitar, A. A., Leroux, D., Kerr, Y., Merlin, O., Richaume, P., Sahoo, A., et al. (2012). Evaluation of SMOS soil moisture products over continental US using SCAN/SNOTEL network. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 1572–1586.
Caires, S., & Sterl, A. (2003). Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research, 108*, 43(1)–43(15).
Carr, D., Kahn, R., Sahr, K., & Olsen, T. (1997). ISEA discrete global grids. *Statistical Computing and Statistical Graphics Newsletter, 8*, 31–39.
Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., & Elrayes, M. A. (1985). Microwave dielectric behavior of wet soil. 2. Dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing, 23*, 35–46.
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 data sets. *Hydrology and Earth System Sciences, 7*, 5621–5645.
Douville, H., & Chauvin, F. (2000). Relevance of soil moisture for seasonal climate predictions: A preliminary study. *Climate Dynamics, 16*, 719–736.
Drusch, M. (2007). Initializing numerical weather predictions models with satellite derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set. *Journal of Geophysical Research, 113*.
Entekhabi, D., Njoku, E., ONeill, P., Kellogg, K., Crow, W., Edelstein, W., et al. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE, 98*, 704–716.
Gruhier, C., de Rosnay, P., Kerr, Y., Mougin, E., Ceschia, E., Calvet, J., et al. (2008). Evaluation of AMSR-E soil moisture product based on ground measurements over temperate and semi-arid regions. *Geophysical Research Letters, 35*.
Guilfen, Y., Chapron, B., & Vandemark, D. (2001). The ERS scatterometer wind measurement accuracy: Evidence of seasonal and regional biases. *Journal of Atmospheric and Oceanic Technology, 18*, 1684–1697.
Jackson, T., Bindlish, R., Cosh, M., Zhao, T., 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.

Please cite this article as: Leroux, D.J., et al., Spatial distribution and possible sources of SMOS errors at the global scale, *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2013.02.017

Janssen, P., Abdalla, S., Hersbach, H., & Bidlot, J. -R. (2007). Error estimation of buoy, satellite and model wave height data. *Journal of Atmospheric and Oceanic Technology*, *24*, 1665–1677.

Kerr, Y., Waldteufel, P., Richaume, P., Davenport, I., Ferrazzoli, P., & Wigneron, J. -P. (2010). *SMOS Level 2 processor soil moisture algorithm theoretical basis document.* Technical Report SO-TN-ESL-SM-GS-0001 CESBIO, IPSL-Service d'aeronomie, INSA-EPHYSE, Reading University, Tor Vergata University.

Kerr, Y., Waldteufel, P., Richaume, P., Wigneron, J., Ferrazzoli, P., Mahmoodi, A., et al. (2012). The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, *50*, 1384–1403.

Kerr, Y., Waldteufel, P., Wigneron, J., Delwart, S., Cabot, F., Boutin, J., et al. (2010). The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE*, *98*, 666–687.

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., Dirmeyer, P., Guo, Z., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*, 1138–1140.

Q7 Leroux, D., Kerr, Y., Bitar, A. A., Gruhier, C., Bindlish, R., Jackson, T., et al. (submitted for publication). Comparison between SMOS and other satellite and model forecast products. *IEEE Transactions on Geoscience and Remote Sensing*.

Li, L., Gaiser, P., Gao, B., Bevilacqua, R., Jackson, T., Njoku, E., et al. (2010). Windsat global soil moisture retrieval and validation. *IEEE Transactions on Geoscience and Remote Sensing*, *48*.

Loew, A., & Schlenz, F. (2011). A dynamic approach for evaluating coarse scale satellite soil moisture products. *Hydrology and Earth System Sciences*, *15*, 75–90.

Masson, V., Champeau, J. -L., Chauvin, F., Meriguet, C., & Lacaze, R. (2003). A global data base of land surface parameters at 1 km resolution in meteorological and climate models. *Journal of Climate*, *16*, 1261–1282.

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.

Mironov, V., & Fomin, S. (2009). Temperature and mineralogy dependable model for microwave dielectric spectra of moist soils. *PIERS Online*, *5*, 411–415.

Naeimi, V., Scipal, K., Bartalis, Z., & Wagner, S. H. 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. (2004). *Updated daily. AMSR-E/Aqua daily L3 surface soil moisture, interpretive parameters, QC EASE-grids. Digital Media.*

Njoku, E., Jackson, T., Lakshmi, V., Chan, T., & Nghiem, S. (2003). Soil moisture retrieval from ASMR-E. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 215–229.

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.

Parrens, M., Zakharova, E., Lafont, S., Calvet, J. -C., Kerr, Y., Wagner, W., et al. (2011). Comparing soil moisture retrievals from SMOS and ASCAT over France. *Hydrology and Earth System Sciences Discussions*, *8*, 8565–8607.

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., Dorigo, W., & de Jeu, R. (2010). Triple collocation — A new tool to determine the error structure of global soil moisture products. *Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4426–4429).

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.

Shin, D., Bellow, J., LaRow, T., Cocke, S., & OBrien, J. (2006). The role of an advanced land model in seasonal dynamical downscaling for crop model application. *Journal of Applied Meteorology and Climatology*, *45*, 686–701.

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research*, *103*, 7755–7766.

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.

Wigneron, J., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M., Richaume, P., et al. (2007). L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields. *Remote Sensing of Environment*, *107*, 639–655.

World Meteorological Organization, Intergovernmental Oceanographic Commission, United Nations Environment Programme, & International Council for Science (2010). Implementation plan for the global observing system for climate in support of the UNFCCC. *Technical Report World Climate Observing System.*

Please cite this article as: Leroux, D.J., et al., Spatial distribution and possible sources of SMOS errors at the global scale, *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2013.02.017