SMOS-IC: An alternative SMOS soil moisture and vegetation optical depth product

Roberto Fernandez-Moran 1,2, Amen Al-Yaari 1, Arnaud Mialon 3, Ali Mahmoodi 3, Ahmad Al Bitar 3, Gabrielle De Lannoy 4, Ernesto Lopez-Baeza 2, Yann Kerr 3, Jean-Pierre Wigneron 1

1 INRA, UR1391 ISPA, F-33140, Villenave d'Ornon, Centre INRA Bordeaux Aquitaine, France; amen.al-yaari@inra.fr (A. A); jean-pierre.wigneron@inra.fr (J.-P. W)
2 University of Valencia. Faculty of Physics. Dept. of Earth Physics & Thermodynamics. Climatology from Satellites Group. 46100 Valencia; Ernesto.Lopez@uv.es (E. L.)
3 CESBIO, CNES/CNRS/IRD/UPS, UMR 5126, Toulouse, France; arnaud.mialon@cesbio.cnrs.fr (A. M.); mahmoodi.ca@gmail.com (A. M.); ahmad.albitar@cesbio.cnrs.fr (A. A.); yann.kerr@cesbio.cnrs.fr (Y. K.)
4 KU Leuven, Department of Earth and Environmental Sciences, Heverlee, B-3001, Belgium; (gabrielle.delannoy@kuleuven.be (G. D. L.)

Abstract: The main goal of the Soil Moisture and Ocean Salinity (SMOS) mission over land surfaces is the production of global maps of soil moisture (SM) and vegetation optical depth (τ) based on multi-angular brightness temperature (TB) measurements at L-band. The operational SMOS Level 2 and Level 3 soil moisture algorithms account for different surface effects, such as vegetation opacity and soil roughness at 4 km resolution, in order to produce global retrievals of SM and τ. In this study, we present an alternative SMOS product which was developed by INRA (Institut National de la Recherche Agronomique) and CESBIO (Centre d’Etudes Spatiales de la BIOsphère). This SMOS-INRA-CESBIO (SMOS-IC) product provides daily SM and τ at the global scale and differs from the operational SMOS Level 3 (SMOSL3) product in the treatment of retrievals over heterogeneous pixels. Specifically, SMOS-IC is much simpler and does not account for corrections associated to the antenna pattern and the complex SMOS viewing angle geometry. It considers pixels as homogeneous to avoid uncertainties and errors linked to inconsistent auxiliary data sets which are used to characterize the pixel heterogeneity in the SMOS L3 algorithm. SMOS-IC also differs from the current SMOSL3 product (Version 300, V300) in the values of the effective vegetation scattering albedo (ω) and soil roughness parameters. An inter-comparison is presented in this study based on the use of ECMWF (European Center for Medium range Weather Forecasting) SM outputs and NDVI (Normalized Difference Vegetation Index) from MODIS (Moderate-Resolution Imaging Spectroradiometer). A 6 year (2010-2015) inter-comparison of the SMOS products SMOS-IC and SMOSL3 SM (V300) with ECMWF SM yielded higher correlations and lower ubRMSD (unbiased root mean square difference) for SMOS-IC over most of the pixels. In terms of τ, SMOS-IC τ was found to be better correlated to MODIS NDVI in most regions of the globe, with the exception of the Amazonian basin and of the northern mid-latitudes.

Keywords: SMOS; L-band; Level 3; ECMWF; SMOS-IC; soil moisture; vegetation optical depth; MODIS; NDVI

1. Introduction

The estimation of surface soil moisture (SM) at global scale is a key objective for the recent L-band (1.4 GHz) microwave missions SMOS (Soil Moisture and Ocean Salinity) (Kerr et al., 2012 [1]) and SMAP (Soil Moisture Active Passive) (Entekhabi et al., 2010 [2]). Measurements of soil moisture are needed for applications related to the study of climate change or agriculture (droughts, floods, etc.) and hydrological processes (Brocca et al., 2010 [3]) such as precipitation, infiltration, runoff and...
evaporation. Moreover, SM is considered as an Essential Climate Variable and it is included in the Climate Change Initiative (CCI) project (Hollmann et al., 2013 [4]).

The soil moisture of the first 2-3 cm soil layer is highly related to the soil emissivity at L-band through the soil dielectric constant. SMOS uses an interferometric radiometer which delivers multi-angular brightness temperature measurements at L-band (1.4 GHz). Currently, various products are derived from the SMOS data at Level 2 (Kerr et al., 2012 [1]) and at Level 3 (Al Bitar et al., 2017 [5]), such as the SMOSL3 Brightness Temperature (SMOSL3 TB) and the SMOSL3 SM and τ products, with a 625 km² sampling. The SMOS SM retrieval algorithm, which is common to both SMOS Level 2 (L2) and Level 3 (L3) products, has been continuously improved since the launch of the satellite in 2009 (Kerr et al., 2001 [6]; Mialon et al., 2015 [7]; Al Bitar et al., 2017 [5]). It has been validated against several datasets from various space-borne sensors (Al-Yaari et al., 2014 [8]; Al-Yaari et al., 2015 [9], Kerr et al., 2016 [10]). All the different versions of the L2 and L3 products, are based on the inversion of the L-band Microwave Emission of the Biosphere (L-MEB) radiative transfer model (Wigneron et al., 2007 [11]), thus retrieving two main parameters, namely soil moisture and vegetation optical depth at nadir (τ).

The SMOS τ is a measure of the attenuation of microwave radiation by the vegetation canopy at L-band. Vegetation is commonly studied at optical or infrared frequencies. However, the longer wavelength of L-band sensors allows penetration of the radiation within the canopy. Thus, τ can be related to different vegetation features such as forest height (Rahmoune et al., 2013 [12], 2014 [13]), vegetation structure (Schwank et al., 2005 [14], 2012 [15]), water content (Jackson and Schmugge, 1991 [16], Mo et al., 1982 [17], Wigneron et al., 1995 [18]; Grant et al., 2012 [19]), sapflow (Schneebeli et al., 2011 [20]) and leaf fall (Guglielmetti et al., 2008 [21]; Patton et al., 2012 [22]). Furthermore, some vegetation indices can also be related to τ such as the Leaf Area Index (LAI) (Wigneron et al., 2007 [11]) and the normalized difference vegetation index (NDVI) (Grant et al., 2016 [23]). Note that some studies have also demonstrated the notable influence of soil roughness on the retrieved values of the τ parameter at both local and regional scales (Patton et al., 2012 [22]; Fernandez-Moran et al., 2015 [24]; Parrens et al., 2017 [25]).

The L-MEB model has been progressively refined and improved (Wigneron et al., 2011 [26], in press [27]). The SMOS L2 and L3 algorithms are based on a bottom-up approach where the TB contributions of 4 x 4 km land cover surfaces are convoluted using the antenna pattern to upscale the TB simulations to the sensor resolution. The use of such a bottom-up approach to retrieve SM and τ presents two main drawbacks. First this approach is impacted by the uncertainties associated with the higher resolution auxiliary files, like the land cover maps, which are used to characterize the pixel heterogeneity. Second, the approach is more time consuming as the exact antenna patterns have to be applied for each view angle.

In this study an alternative SMOS product is presented, hereinafter referred to as SMOS-IC. This product is based on a simplified approach developed by INRA (Institut National de la Recherche Agronomique) and CESBIO (Centre d’Etudes Spatiales de la Bioosphère) and differs from the operational SMOS Level 2 and Level 3 products in three main ways:

I. The SMOS-IC algorithm does not take into consideration pixel land use and assumes the pixel to be homogeneous as suggested by Wigneron et al. 2012 [28]. The SM and τ retrieval is performed over the whole pixel rather than over the fraction designated as either low vegetation or forest. Note that this approach is similar to the one considered in the development of the AMSR-E and SMAP SM algorithms (O’Neill et al., 2012 [25]). By simplifying the retrieval approach, the SMOS-IC product becomes independent of the ECMWF soil moisture information currently used as auxiliary information to estimate TB in the subordinate pixel fractions of heterogeneous pixels in the operational SMOS L2 and L3 algorithms (Kerr et al., 2012 [1]).

II. SMOS-IC uses as input SMOS Level 3 fixed angle bins Brightness Temperature (TB) data at the top of the atmosphere and contains different flags allowing to filter SM retrievals

DOI: 10.20944/preprints201703.0145.v1
accounting for the quality of the input TB data and for the TB angular range in the LMEB inversion. SMOS-IC does not make use of the computationally expensive corrections based on angular antenna patterns to account for pixel heterogeneity as in the L2 and L3 retrieval algorithms.

III. New values of the effective vegetation scattering albedo (ω) and soil roughness parameters (HR, NRV, and NRH) are considered in the SMOS-IC product. This change is based on the results of Fernandez-Moran et al. (2016) [29] who calibrated the L-MEB vegetation and soil parameters for different land cover types based on the International Geosphere-Biosphere Programme (IGBP) classes, as well as the findings of Parrens et al. (2016) [30] who computed a global map of the soil roughness HR values. The calibration of Fernandez-Moran et al. (2016) [31] was obtained by selecting the values of the parameters (HR, NRV, NRH, and ω) which optimized the SMOS SM retrievals, with respect to the in situ SM values measured over numerous sites obtained from ISMN. The parameter values resulting from this new calibration differ from those used in the current SMOS L2 and L3 products. Values currently used in the SMOS L2 and L3 algorithms (Kerr et al., 2012 [1]) were those decided before launch from literature. Over forested areas, values were updated but not over low vegetation. Consequently, in Version 620 of the L2 (and Version 300 for L3) algorithm, ω is still assumed to be zero over low vegetation canopies and ω ~ 0.06 – 0.08 over forests. Similarly, HR is equal to 0.3 for forests and HR = 0.1 for the rest of the cover types, while QR is set to zero whereas NRH and NRV are respectively set to 2 and 0 at global scale.

IV. In some cases, the Level 2 and Level 3 algorithms use values of LAI derived from MODIS [32], to initialize the value of optical depth in the inversion algorithm (Kerr et al., 2012 [1]). In SMOS-IC, this is not implemented, and the initialization of optical depth in the inversion algorithm is based on a very simple approach (given in the following) and is completely independent of the MODIS data.

An evaluation and calibration of SMOS-IC at local scale was performed in Fernandez-Moran et al. (2016) [29]. The present study aims at presenting SMOS-IC and illustrating the main features of the SMOS-IC SM and τ products at global scale, in comparison to the current SMOSL3 product. To achieve this, the SMOS-IC and SMOSL3 SM products were compared against the ECMWF SM product for ease of comparison. Furthermore, NDVI (Rouse et al., 1974 [33]) from the Moderate-Resolution Imaging Spectroradiometer (MODIS) was used as a vegetation index to analyze the seasonal changes in the τ products from both SMOS-IC and SMOSL3. The NDVI index which is derived from optical observations cannot be directly compared to the τ product, which is derived from microwave observations. However, the NDVI index is a good indicator of the vegetation density and it can be used to interpret the seasonal changes in the SMOS τ product at large scale as found by Grant et al. (2016) [23], but with some caveats: saturation effects at high levels of vegetation density, sensitivity to the effects of snow and soil reflectivity (Qi et al., 1994 [34]), etc. It may be noted that NDVI is the proxy used for estimating τ in the current operational algorithm of the SMAP mission (O’Neill et al., 2012 [35]).

In section 2, we present a description of both SMOS algorithms (SMOSL3 and SMOS-IC) and of the MODIS NDVI and ECMWF SM data sets. The inter-comparison of the SMOS products in terms of soil moisture and vegetation optical depth is given in section 3. The inter-comparison covers almost 6 years of data, from 2010 to 2015, excluding the commissioning phase (the first six months of 2010; Corbella et al., 2011 [36]). Discussion and conclusions are presented in section 4.

2. Materials and method
2.1 SMOSL3 brightness temperature, soil moisture and vegetation optical depth

At Level 3, there are different SMOS products (Al Bitar et al., 2017 [5]). In this study we used the SMOS L3 products which include TB, τ and SM (version 300) data produced by the CATDS (Centre d’Aval de Traitement des Données SMOS) (Al Bitar et al., 2017 [5]). These products are available in the NetCDF format and on the Equal-Area Scalable Earth (EASE) 2.0 grid (Armstrong et al., 1997 [37]) with a 625 km² sampling (Brodzik and Knowles, 2002 [38]). The SMOSL3 TB is measured at the top of the atmosphere and provided in the surface reference frame (i.e., H and V polarizations) at angles ranging from 2.5° ± 2.5° to 62.5° ± 2.5°. Ascending (~ 06:00 LST at the equator) and descending (~ 18:00 LST) orbits are processed separately. The Level 3 processor uses the same physically based forward model (L-MEB) as the ESA SMOS Level 2 processor (Kerr et al., 2012 [1], Kerr et al., 2013 [39]) for the retrieval of both SM and τ from dual polarization (H, V) and multi-angular SMOS measurements. The retrieval algorithm consists of the minimization of the differences between observed and modeled Level 1 TB (through the L-MEB forward model) in a Bayesian cost function which accounts for the observation uncertainty, and also contains a prior parameter constraint. One of the characteristics of the TB modeling is the consideration of surface heterogeneity. The total modeled TB is simulated as the sum of TB contributions from several fractions (nominal or low vegetation, forest, and others as urban, water, etc.). In most of the cases, the SM retrieval is estimated from the TB contribution which corresponds to areas with low vegetation (nominal fraction), while the TB forest contribution is computed using ancillary data such as ECMWF SM. In other cases, the retrieval is performed entirely over the forest fraction. Dynamic changes as freezing or rainfall events are considered through ancillary weather data from ECMWF.

The SMOSL3 τ and SM retrievals are provided at different temporal resolutions: daily, 3-days, 10-days, and monthly averaged (Kerr et al., 2013 [39]; Jacquette et al., 2010 [40]). The quality of the SMOSL3 product containing SM and τ data is improved by the use of multi-orbit retrievals (Al Bitar et al., 2017 [5]). The SMOS ascending (6 am LST) and descending (6 pm LST) orbits are processed separately in this product in order to better account for the diurnal effects (surface, Total Electron Content which drives Faraday rotation and sun corrections) and, in some areas, Radio Frequency Interferences (RFI) effects (Oliva et al., 2012 [41]) and sun glint impacts at L-band (Khazaïal et al., 2016 [42]).

In SMOS-IC, we used the SMOS L3 TB product as input to the inversion algorithm. This product, which includes many corrections, is very easy and convenient to use (conversely, the L2 and L3 algorithms are based on L1 C TB data).

2.2 SMOS-IC soil moisture and vegetation optical depth dataset

2.2.1 Model description

As for the L2 and L3 algorithms, in SMOS-IC, the retrieval of the soil moisture and vegetation optical depth at nadir is based on the L-MEB model inversion (Wigneron et al., 2007 [11]). The retrieval is performed over pixels which are considered as entirely homogeneous; in other words, a single representative value of each input model parameter is used for the whole pixel.

In L-MEB, the simulation of the land surface emission is based on the τ-ω radiative transfer model (Mo et al., 1982 [17]) using simplified (zero-order) radiative transfer equations. The model represents the soil as a rough surface with a vegetation layer. The modeled TB from the soil vegetation medium is calculated as the sum of the direct vegetation emission, the soil emission attenuated by the canopy and the vegetation emission reflected by the soil and attenuated by the canopy following equation (1). The atmospheric contribution is neglected.

\[
TB_p(\theta) = (1 - \omega)[1 - \gamma_p(\theta)][1 + \gamma_p(\theta)r_{GP}(\theta)]T_C + [1 - r_{GP}(\theta)]\gamma_p(\theta)T_G
\]

(1)
where $\theta$ is the incidence angle, $r_{GP}$ is the soil reflectivity, $T_c$ and $T_c$ are the soil and canopy effective temperatures, $\gamma_p$ is the vegetation transmissivity (or vegetation attenuation factor) and $\omega$ is the effective scattering albedo (polarization effects are not taken into account for this parameter).

Roughness effects are parameterized through a semi-empirical approach initially developed by Wang and Choudhury (1981) [43] and refined in more recent studies (Escorihuela et al., 2007 [44]; Lawrence et al., 2013 [45]; Parrens et al., 2016 [25]). The roughness modelling is based on four parameters ($Q_R$, $H_R$, $N_{RH}$ and $N_{RV}$). The values of $Q_R$ and $N_{RP}$ ($P = H, V$) have been calibrated in Fernandez-Moran et al. (2015, 2016) [46] [29] where optimized values of $Q_R = 0$ and $N_{RP} = -1$ ($P = H, V$) were obtained globally. Thus, the calculation of the soil reflectivity $r_{GP}$ is given by:

$$r_{GP}(\theta) = r_{GP}^* (\theta) \exp\left[-H_R/\cos(\theta)\right]$$  \hspace{1cm} (2)

where $r_{GP}^* (P = H, V)$ is the reflectivity of a plane (specular) surface, which is computed from the Fresnel equations (Ulaby, 1982 [47]) as a function of $\theta$ and of the soil dielectric constant ($\varepsilon$), expressed as a function of SM, soil clay fraction and soil effective temperature ($T_c$) using the model developed by Mironov et al. (2012) [48]. $H_R$ accounts for the decrease of $r_{GP}$ due to soil roughness effects.

Under the assumption of isotropic conditions and no dependence of the vegetation optical depth on polarization, the vegetation attenuation factor $\gamma_p$ can be computed using the Beer’s law as:

$$\gamma_p = \exp[-\tau/\cos(\theta)]$$  \hspace{1cm} (3)

The retrieval of SM and $\tau$ involves the minimization of the following cost function $x$:

$$x = \frac{\sum_{i=1}^{N}(TB_p(\theta)_{mes} - TB_p(\theta))^2}{\sigma(TB)^2} + \sum_{i=1}^{2} \frac{(P_{i}^{ini} - P_i)^2}{\sigma(P_i)^2}$$  \hspace{1cm} (4)

where $N$ is the number of observations for different viewing angles ($\theta$) and both polarizations (H & V), $TB_p(\theta)_{mes}$ is the measured value over the SMOS pixels from the SMOSL3 TB product (presented in section 2.2.2), $\sigma(TB)$ is the standard deviation associated with the brightness temperature measurements (this parameter was set to the constant value of 4 K in this study), $TB_p(\theta)$ is the brightness temperature calculated using equation (1), $P_i$ ($i = 1, 2$) is the value of the retrieved parameter (SM, $\tau$); $P_{i}^{ini}$ ($i = 1, 2$) is an a priori estimate of the parameter $P_i$ and $\sigma(P_i)$ is the standard deviation associated with this estimate. A constant initial value of 0.2 m$^3$/m$^3$ was considered for SM and $\sigma(SM)$ and a value of 0.5 was considered for $\tau_{NAD}$ and 1 for $\sigma(\tau_{NAD})$.

2.2.2 Effective vegetation scattering albedo, soil roughness and soil texture parameters

One of the most important features of the SMOS-IC product is the ability to test new calibrated values of $\omega$ (Fernandez-Moran et al, 2016 [29]) and $H_R$ (Parrens et al. 2016 [30]. Table 1 presents these values for SMOS-IC and SMOSL3 V300 as a function of the IGBP land category classes. It must be noted that SMOSL3 V300 uses the ECOCLIMAP classification (Masson et al., 2003 [49]) and that in new versions of SMOSL3, IGBP land use maps could be used.

Table 1: Calibrated values of $\omega$ and $H_R$ as function of the IGBP land category classes for SMOS-IC and SMOSL3.

<table>
<thead>
<tr>
<th>Class</th>
<th>$\omega$ (SMOS-IC)</th>
<th>$\omega$ (SMOSL3 V300)</th>
<th>$H_R$ (SMOS-IC)</th>
<th>$H_R$ (SMOSL3 V300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Evergreen needle leaf forest</td>
<td>0.10</td>
<td>0.06 – 0.08*</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>IGBP Class</td>
<td>HR</td>
<td>ω</td>
<td>τ</td>
<td>met</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>2 – Evergreen broadleaf forest</td>
<td>0.10</td>
<td>0.06 – 0.08*</td>
<td>0.47</td>
<td>0.30</td>
</tr>
<tr>
<td>3 – Deciduous needle leaf forest</td>
<td>0.10</td>
<td>0.06 – 0.08*</td>
<td>0.43</td>
<td>0.30</td>
</tr>
<tr>
<td>4 – Deciduous broadleaf forest</td>
<td>0.10</td>
<td>0.06 – 0.08*</td>
<td>0.46</td>
<td>0.30</td>
</tr>
<tr>
<td>5 – Mixed forests</td>
<td>0.10</td>
<td>0.06 – 0.08*</td>
<td>0.43</td>
<td>0.30</td>
</tr>
<tr>
<td>6 – Closed shrublands</td>
<td>0.10</td>
<td>0.00</td>
<td>0.27</td>
<td>0.10</td>
</tr>
<tr>
<td>7 – Open shrublands</td>
<td>0.08</td>
<td>0.00</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td>8 – Woody savannas</td>
<td>0.12</td>
<td>0.00</td>
<td>0.35</td>
<td>0.10</td>
</tr>
<tr>
<td>9 – Savannas</td>
<td>0.10</td>
<td>0.00</td>
<td>0.23</td>
<td>0.10</td>
</tr>
<tr>
<td>10 – Grasslands</td>
<td>0.10</td>
<td>0.00</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>11 – Permanent wetland</td>
<td>0.10</td>
<td>0.00</td>
<td>0.19</td>
<td>0.10</td>
</tr>
<tr>
<td>12 – Croplands</td>
<td>0.12</td>
<td>0.00</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td>13 – Urban and built-up</td>
<td>0.10</td>
<td>0.00</td>
<td>0.21</td>
<td>0.10</td>
</tr>
<tr>
<td>14 – Cropland/Natural Vegetation Mosaic</td>
<td>0.12</td>
<td>0.00</td>
<td>0.22</td>
<td>0.10</td>
</tr>
<tr>
<td>15 – Snow and ice</td>
<td>0.10</td>
<td>0.00</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>16 – Barren and sparsely vegetated</td>
<td>0.12</td>
<td>0.00</td>
<td>0.02</td>
<td>0.10</td>
</tr>
</tbody>
</table>

*ω = 0.08 over boreal forests, ω = 0.06 over other forest types

In SMOS-IC, the retrieval of SM and τ is performed over the totality of each pixel and the input parameters Hₓ and ω are consequently constant values for the whole pixel. However, due to the heterogeneity present in all pixels, the input Hₓ and ω parameters used in the retrieval are calculated by linear weighting the Hₓ and ω contribution according to the percentage of each IGBP class within the pixel based on the values provided in Table 1. For instance, if a pixel is covered by 60% of grasslands and 40% of croplands, the effective vegetation scattering albedo considered for that pixel is calculated as follows: ω = 0.60·0.10 + 0.40·0.12 = 0.108. The assumption of linearity, which is questionable, was made here as it leads to a very simple correction, and as no other more physical and general formulation was available.

The soil texture in terms of clay content is obtained in the SMOS-IC product from the Food and Agriculture Organization map (FAO, 1998) [50]. This map is re-gridded in the same EASE 2.0 grid used by SMOSL3.

2.2.3 Quality flags

The data filtering of the SMOS-IC product was done through different scene and quality flags which are summarized in Tables 2 and 3. The scene flags indicate the presence of moderate and strong topography, frozen soil or polluted scene. TB data for pixels where the sum of the water, urban and ice fractions were higher than 10% were filtered out (considered as polluted scene). For ECMWF soil temperatures below 273.15 K, the soil was considered as frozen. The quality flags helped to filter out all cases suspected to give dubious results. Consequently, only TB values not affected by noise (RFI, Sun glint effects, etc) were selected. For this, only TB values whose standard deviations were within radiometric accuracy were kept (TB with a standard deviation exceeding 5 K plus the TB radiometric accuracy were filtered out). Moreover, only retrievals (i) made in the range of incidence angles of 20 to 55° and (ii) with a range of angular values exceeding 10° (to ensure a sufficient sampling of the angular distribution) were considered. The quality flags helped also to filter out those retrievals.
where the RMSE values between the measured (L3 TB) and the L-MEB modeled TB data were larger than 12 K as suggested in Wigneron et al. (2012) [28].

Table 2: Description of the SMOS-IC scene flags

<table>
<thead>
<tr>
<th>Scene flags</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of moderate topography</td>
<td>Same filter as SMOSL3 V300</td>
</tr>
<tr>
<td>Presence of strong topography</td>
<td>Same filter as SMOSL3 V300</td>
</tr>
<tr>
<td>Polluted scene</td>
<td>Water, urban and ice fractions (according to the IGBP classification) represent less than 10% of the pixel</td>
</tr>
<tr>
<td>Frozen scene</td>
<td>Soil temperature &lt; 273.15 K</td>
</tr>
</tbody>
</table>

Table 3: Description of the SMOS-IC quality flags

<table>
<thead>
<tr>
<th>Quality flags</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM retrieved successfully</td>
<td></td>
</tr>
<tr>
<td>SM retrieved successfully but not recommended</td>
<td>RMSE &lt; 12 K</td>
</tr>
<tr>
<td>No data after first filtering</td>
<td>TB not valid (angles out of the range 20–55º); TB standard deviation higher than accuracy + 5 K</td>
</tr>
<tr>
<td>Failed retrieval (1)</td>
<td>TB angle separation too narrow (angle difference &lt; 10º)</td>
</tr>
<tr>
<td>ECMWF soil temperatures out of range</td>
<td>“No value” or values out of range</td>
</tr>
<tr>
<td>Failed retrieval (2)</td>
<td>Values of clay content are out of range</td>
</tr>
<tr>
<td>Failed retrieval (3)</td>
<td>SM &lt; 0 m³/m³</td>
</tr>
</tbody>
</table>

2.3 ECMWF and MODIS data

The ECMWF dataset used in this study for the SM product inter-comparison was obtained from the SMOSL3 SM pre-processor. This ECMWF product has a spatial resolution of 625 km² and 1-day temporal resolution, using the same EASE 2.0 grid and interpolated in time and space to fit the SMOSL3 sampling resolutions. It is based on the ERA-Interim dataset. ERA-Interim uses a numerical weather prediction (NWP) system (IFS – Cy31r2) to produce reanalyzed data (Berrisford et al., 2011) [51].

The ECMWF soil surface (Level 1, top 0-7 cm soil layer) and soil deep temperature (Level 3, 28-100 cm) are used in the computation of the effective soil temperature for the SMOS-IC and SMOSL3 SM products following the parameterization of Wigneron et al. (2001) [52]. It is worth noting that unlike the SMOSL3 SM product, the SMOS-IC processor does not use the ECMWF SM product to compute contributions from the fixed fractions (i.e. fraction of the scene over which the SM retrieval is not performed), and is only considered for evaluation purpose in this study.

The ECMWF SM product represents the top 0-7 cm surface layer and it has been frequently compared to retrieved SM at global scale (Al-Yaari et al., 2014 [53]; Albergel et al., 2013 [54]; Leroux et al., 2014) [55]. ECMWF SM was found by Albergel et al. (2012) [56] to represent very well the SM
variability at large scales. It is also known to give erroneous values in some areas (Louvet et al., 2015 [57]; Kerr et al., 2016 [10]).

The NDVI product used in this study was obtained from the 16 day NDVI MODIS Aqua and Terra data, with a 1 km resolution. This product was re-gridded in the EASE 2.0 grid in order to make it comparable with SMOS-IC and SMOSL3 SM. Different studies have shown that \( \tau \) at microwave frequencies has high spatial correspondences with MODIS NDVI (De Jeu and Owe, 2003 [58]; Andela et al., 2013 [59]) even though both products have shown sensitivity to different aspects of the vegetation dynamics (Grant et al., 2016 [23]).

2.4 Inter-comparison

The inter-comparison was made for both SMOS-IC and SMOSL3 products by direct comparison between SM \((\text{m}^3/\text{m}^2)\) and \( \tau \), against, respectively, the ECMWF SM and MODIS NDVI products. This section explains the filtering which was applied to the latter datasets and the metrics used in the evaluation process.

2.4.1 Data filtering

In the evaluation step, only ascending SMOS SM retrievals were selected (Al-Yaari et al., 2014 [8][53]). Flags associated with SMOSL3 were used to filter both SMOSL3 and SMOS-IC.

For the SMOSL3 SM product, a quality index (DQX) estimates the retrieval quality. In this study, data with DQX > 0.06 \( \text{m}^3/\text{m}^2 \) were excluded. In parallel, the Level 3 RFI probability flag was used to filter out SM data contaminated by RFI. SM retrievals with an associated RFI probability higher than 20% and frozen areas were removed (surface temperature < 273.15 K). The SMOS-IC and SMOSL3 filtered retrievals of SM and \( \tau \) used in the study were inter-compared for the same dates. For both SMOS products (SMOSL3 and SMOS-IC), SM values out of the range 0 - 0.6 \( \text{m}^3/\text{m}^2 \) (Dorigo et al., 2013 [60]) and \( \tau \) values out of the range 0 - 2 were filtered out. We only considered pixels with temporal series of at least 15 values for the product inter-comparison.

In order to compare \( \tau \) with MODIS NDVI, the daily \( \tau \) values were re-gridded to 16-day mean values produced every 8 days following the same methodology as described in Grant et al. (2016) [23].

2.4.2 Metrics

For evaluation purposes, the following metrics were used: Pearson correlation coefficient (R), bias, root mean square difference (RMSD) and unbiased RMSD (ubRMSD). Equations for the calculation of the SM metrics are the following:

\[
R = \frac{\sum_{i=1}^{n}(SM_{EC(i)} - SM_{EC}) \left( SM_{SMOS(i)} - SM_{SMOS} \right)}{\sqrt{\sum_{i=1}^{n}(SM_{EC(i)} - SM_{EC})^2 \sum_{i=1}^{n}(SM_{SMOS(i)} - SM_{SMOS})^2}}
\]

\[
\text{bias} = (SM_{SMOS} - SM_{EC})
\]

\[
RMSD = \sqrt{(SM_{SMOS} - SM_{EC})^2}
\]

\[
ubRMSD = \sqrt{RMSD^2 - \text{bias}^2}
\]

where \( n \) is the number of SM data pairs, \( SM_{SMOS} \) is the SMOS SM product (SMOSL3 SM or SMOS-IC) and \( SM_{EC} \) is the ECMWF SM. It should be noted the use of RMSD instead of root mean square error (RMSE) as ECMWF SM contain errors and cannot be considered as the “true” ground SM value.
In this study, only significant correlations were considered by means of a p-value filtering for SM retrievals, i.e. pixels where the p-value was above 0.05 were filtered out.

In order to evaluate $\tau$, $R$ was calculated as follows:

$$R = \frac{\sum_{i=1}^{n}(NDVI_i - \bar{NDVI})(\tau_{SMOS(i)} - \bar{\tau}_{SMOS})}{\sqrt{\sum_{i=1}^{n}(NDVI_i - \bar{NDVI})^2} \sum_{i=1}^{n}(\tau_{SMOS(i)} - \bar{\tau}_{SMOS})^2}$$

where $\tau_{SMOS}$ is the vegetation optical depth at nadir ($\tau$) retrieved from the SMOSL3 or SMOS-IC product.

3. Results and discussion

3.1 Soil moisture

Figure 1 shows the values of the temporal mean SM over the globe and over the period 2010-2015 for the three SM datasets considered in this study: (a) SMOS-IC, (b) SMOSL3 SM, and (c) ECMWF. It should be kept in mind that ECMWF SM is representative of the first 0-7 cm of the soil surface (Albergel et al., 2012 [61]) and the inherent nature of the simulated soil moisture (Koster et al., 2009 [62]) is different to that measured by the SMOS satellite observations, which are sensitive to the first ~0-3 cm of the soil surface (Escorihuela et al., 2010 [63]; Njoku and Kong et al., 1977 [64]). In Figure 1, ECMWF SM must be analyzed in terms of spatial patterns rather than absolute values. Although Figure 1 (a) and (b) have many similarities, some spatial patterns showed by the ECMWF SM product are in better agreement with SMOS-IC than with SMOS L3 SM. For instance, over the Appalachian region in the Eastern US, SMOSL3 SM shows a dry area whereas SMOS-IC SM is closer to ECMWF, as these regions are known to be relatively wetter than the regions of west and midwest (Sheffied et al., 2004 [65]; Fan et al., 2004 [66]). This was partly explained by differences between ECOCLIMAP and IGBP and the use of ECMWF SM data in Mahmoodi et al., 2015 [67]. On the other hand, drier retrievals were found for SMOS-IC in the intertropical regions of Africa, for instance over the savannas and grasslands of Sahel. Over these regions SMOS-L3 SM is closer to ECMWF SM than SMOS-IC SM.
Figure 1: Temporal mean of soil moisture (m^3/m^3) during 2010-2015: (a) SMOS-IC, (b) SMOSL3 SM, and (c) ECMWF data. White values mean “no valid SMOS data”.

Figure 2 displays different time series metrics derived from the direct comparison between SMOSL3 SM (a) and SMOS-IC SM (b) with ECMWF SM for 2010-2015. According to correlation (R) results, lowest R values were found in forests for both products. A lower number of negative R values
were found between the SMOS-IC and ECMWF SM products. Conversely, SMOSL3 SM yielded negative correlations with ECMWF SM over several forest regions, namely the boreal forests of Alaska, Canada and Russia, and the tropical forests of Amazon and Congo basins. Over the non-forested biomes, R values were also found to be generally higher for SMOS-IC, when compared to SMOSL3 SM. Substantial differences were found in terms of RMSD and ubRMSD: in general, lower values were obtained for the SMOS-IC product, especially over the intertropical regions of America and Africa (in terms of ubRMSD) and the boreal forests of Eurasia (in terms of RMSD). On the other hand, results do not show important differences in terms of bias between the two SMOS products: both SMOS-IC and SMOSL3 SM products are generally much drier than ECMWF SM, except over some arid and semi-arid areas (deserts in central Asia and Australia, Sahara in Northern Africa). The general negative values of the bias can be partly explained by the differences in sampling depths between the SMOS observations (~0-3 cm top soil layer) and the modeled ECMWF SM (0-7 cm top soil layer). Considering this difference in sampling depths, the observed difference in SM bias patterns in Figure 2, should be interpreted with care.
Figure 2: Pixel-based statistics during 2010-2015 computed between ECMWF SM simulations and SMOSL3 SM (left) and SMOS-IC (right) SM retrievals: (a)(b) correlation coefficient, (c)(d) RMSD, (e)(f) bias, and (g)(h) ubRMSD.
Figure 3 is focused on the results in terms of correlation and ubRMSD, considered as first order criteria. It displays a world map which shows where the best correlation coefficient (R) and ubRMSD are obtained by comparing ECMWF SM with SMOS-IC SM (red) or SMOSL3 SM (blue) in the period 2010-2015. Areas where the result differs by less than 0.02 in terms of R values between SMOSL3 SM and SMOS-IC are represented in green color. This threshold is different for the ubRMSD metric and it was set to 0.005 m³/m³. It can be seen that the red color is dominant, meaning that SMOS-IC SM is generally closer to ECMWF in terms of temporal dynamics. There are some exceptions. For instance, regions colored in blue (SMOSL3 is closer to ECMWF than SMOS-IC) can be found especially for the ubRMSD metric, in central Europe, central and Northern Asia. It should be noted here that only pixels with significant correlations i.e. p-value < 0.05 and a number of data (>15) are presented.

Figure 3: Comparison of the SMOS SM products with respect to ECMWF showing: (a) where SMOS-IC SM (red) or SMOSL3 SM (blue) leads to the best correlation coefficient, or where the difference in R < 0.02 (green) among both SMOS products; (b) where SMOS-IC SM (red) or SMOSL3 SM (blue) lead to the lowest ubRMSE or where the difference in ubRMSD < 0.005 (green).
In order to better assess the range of R and ubRMSD values, the dispersion diagrams displayed in Figure 4 show the scatter plot of both metrics for all pixels and for both SMOS products (SMOS-IC and SMOSL3 SM). In terms of correlation, the R values are generally larger for SMOS-IC. There are also a number of pixels where SMOSL3 SM yields negative correlations whereas R is positive for SMOS-IC. In terms of ubRMSD, the largest number of pixels with lower ubRMSD corresponds to the SMOS-IC SM product.

Figure 4: Scatter plot of correlation (a) and ubRMSD values (b) obtained by comparing both SMOS-IC and SMOSL3 SM to ECMWF SM.

3.2 Vegetation optical depth

The evaluation of the accuracy of the $\tau$ values retrieved from SMOS at global scale is not a simple issue due to the absence of a consensus on the reference values to be considered at large scale coming from models or in situ measurements. Some studies have been done at local scale. For instance, over croplands and grasslands, $\tau$ values at L-band vary generally between 0 and 0.6 (Saleh et al., 2006 [68], Wigneron et al., 2007 [11]). Over forests and from L-band radiometer measurements, Ferrazzoli et al. (2002) [69] found maximum values of $\tau \sim 0.9$, and Grant et al. (2008) [70] found values of $\tau \sim 0.6-0.7$ for a mature pine forest stand in les Landes forest, and $\tau \sim 1$ for a mature deciduous (beech) canopy in Switzerland.

Figure 5 shows a global map of the temporal mean of the retrieved $\tau$ values for both SMOS-IC and SMOSL3 products. Both products show $\tau$ values which are sensitive to vegetation, as the highest $\tau$ values were found for the main boreal and tropical forests. It must be noted that the $\tau$ values coming from the SMOSL3 product were larger than those obtained by the SMOS-IC product.
In order to identify possible patterns, Figure 6 shows a global map which illustrates the differences of $\tau$ between both SMOS datasets (SMOSL3 minus SMOS-IC). This result shows that the greatest differences between both $\tau$ datasets were found over forest areas, particularly tropical regions; namely Amazon and Congo River basins and Borneo and New Guinea tropical forests, where significantly larger $\tau$ values were obtained with SMOSL3.
Figure 6: mean bias: SMOSL3 $\tau$ minus SMOS-IC $\tau$ for 2010-2015.

Figure 7 shows the correlations obtained by comparing the SMOSC-IC and SMOSL3 $\tau$ datasets to MODIS NDVI. All correlations values are presented here including those not significant as done by Grant et al. (2016) [23]. It can be noted that slightly higher correlation values are generally obtained with SMOS-IC especially in the west of Mexico, the Northeastern regions of Brazil and some parts of the Sahel. Conversely higher R values were obtained in western and central Europe with SMOSL3. The lowest correlations were found generally over forests for both SMOS products; a result which can be partly related to the tendency of NDVI to saturate for high biomass and LAI values. However, higher R values were obtained with SMOS-IC for some areas of the boreal forests and the tropical forests of Africa.

A global map that shows for each pixel which $\tau$ dataset (SMOSL3 or SMOS-IC) leads to the largest correlation (R) values with MODIS NDVI is presented in Figure 8. Over northern mid-latitudes, larger correlations were generally obtained with SMOSL3. However, except for these regions, the highest R values were generally obtained with SMOS-IC while no clear patterns were found in terms of longitude. Figure 9 shows a dispersion diagram in order to assess the range of correlation values found for both SMOS $\tau$ datasets against MODIS $\tau$. The diagram generally yields positive correlations, although a non-negligible number of negative correlations can be noted for both SMOS products.
Figure 8: Comparison of SMOS-IC and SMOS-L3 $\tau$ products with respect to MODIS NDVI: higher correlation (R) values between SMOS-IC $\tau$ and MODIS NDVI (red) or between SMOS-L3 $\tau$ and MODIS NDVI (blue) and where the difference in R < 0.02 (green).

Figure 9: Scatter plot showing correlation values obtained between SMOS-IC $\tau$ and MODIS NDVI against correlation values obtained between the $\tau$ from SMOSL3 and MODIS NDVI.

4. Summary and conclusions

This study presents an alternative SMOS SM and $\tau$ product, referred to as SMOS-IC. In terms of soil moisture, the presentation is based on an inter-comparison between SMOS-IC, the official Level
3 SMOS SM product (SMOSL3, V300), and a modeled SM product (ECMWF SM). The SMOS-IC product is based on the retrieval of SM and τ over pixels treated as homogeneous by means of the L-MEB model inversion. SMOS-IC uses the multi-angular and dual-polarization SMOSL3 TB product as the main input for the L-MEB model inversion. The L-MEB model input parameters (effective vegetation scattering albedo ω and the roughness parameter Hr) are estimated as a function of IGBP land category classes which compose the pixel. These parameter values are derived from previous analyses made by Fernandez-Moran et al. (2016) [31] and global maps of the roughness Hr parameter estimated by Parrens et al. (2016) [30]. Conversely, the SMOSL3 product considers different fractions over the pixel and performs SM and τ retrievals over the main fraction of the pixel (usually low vegetation) or over forests in some cases. In the SMOSL3 retrieval algorithm, the TB value of the pixel fraction which is not considered in the retrieval (the forest fraction in general) is estimated based on auxiliary ECMWF SM data. This specific approach may lead to dry SM bias in forested regions, as noted by Wigneron et al. (2012) [28]. Currently, in the SMOSL3 V300 retrieval algorithm, the values of the vegetation and soil roughness parameters differ mainly between forest and low vegetation categories.

The SMOSL3 and SMOS-IC soil moisture retrievals were compared globally against ECMWF SM data for the period 2010-2015. This evaluation extends the work of Fernandez-Moran et al. (2016) [31] who evaluated a preliminary version of the SMOS-IC product at local scale using numerous in situ SM stations from ISMN and found higher R and lower ubRMSE with SMOS-IC than with the SMOSL3 V300 product. At global scale, both the SMOS-IC and SMOSL3 SM products were generally found to be drier than the ECMWF SM product. However, the larger soil sampling depth of the ECMWF SM (0-7 cm) with respect to SMOS SM (~0 - 3 cm), as well as the inherently different nature of simulated soil moisture (Koster et al., 2009 [62]), makes it difficult to truly assess the performance of the SMOS products in terms of bias at global scale. In terms of temporal variations, higher correlation values and lower ubRMSE values were generally found between SMOS-IC SM and ECMWF SM, than between SMOSL3 SM and ECMWF SM.

The ECMWF SM data set used in this study is not ”truth”, and a larger inter-comparison of SMOS-IC and SMOSL3 against other modeled SM products should be made in the future to confirm the very preliminary results found in this study. In terms of τ values, the SMOS-IC and SMOSL3 τ products were compared to MODIS NDVI values over 2010-2015 in terms of correlation values. The SMOS-IC τ product presents a lower range of values (~0-0.6) than the one obtained with the SMOSL3 τ product (~0-1.2). The latter range of τ values (obtained for SMOSL3) is in better agreement than SMOS-IC τ, with the ranges of retrieved τ values based on in situ L-band radiometric measurements (τ ~0.6 -1.0) performed over mature coniferous and deciduous forests in Europe. Conversely, slightly higher correlation values were obtained between SMOS-IC τ and MODIS NDVI, than between SMOSL3 τ and MODIS NDVI, except in the Amazon basin and in regions of the northern mid-latitudes.

The τ results should also be interpreted with care: the NDVI index is derived from optical sensors while the τ index is derived from L-band microwave measurements and therefore can sense deeper through the vegetation canopy. Moreover, the NDVI index is used to monitor the green vegetation, while the τ index is related to the whole vegetation water content (including stems, trunks, branches and senescent vegetation elements). So at L-band, the NDVI index (as the LAI index) is only a proxy which is used to provide an estimate of τ over rather low vegetation covers during the vegetation growth (O’Neill et al, 2012 [35]; Wigneron et al., 2007 [11]; Lawrence et al., 2014 [71]; Grant et al., 2016 [23]). A larger inter-comparison of the SMOS-IC and SMOSL3 τ products against different vegetation data sets (remotely sensed products, LAI, forest biomass) should be made in the future to confirm the results found in this study.

As for the Level 2 and 3 algorithms, based on rather complex and detailed concepts and auxiliary data sets, the simple SMOS-IC algorithm will be improved regularly and will be used to improve L2 and L3 SMOS retrieval algorithms. These different approaches are complementary and a regular inter-comparison analysis between them should be of great benefit to improve the L-MEB inversion, and ultimately the SM and τ products retrieved from the SMOS observations.
Acknowledgements

The authors would like to thank the TOSCA (Terre Océan Surfaces Continentales et Atmosphère) CNES program and the European Space Agency (ESA) for funding this study. The authors acknowledge CATDS for the SMOSL3 dataset (http://catds.ifremer.fr) and the cooperation of the different soil moisture in situ networks of the International Soil Moisture Network (ISMN) project.

Author contributions

Jean-Pierre Wigneron and Roberto Fernandez-Moran designed the SMOS-IC product with the helpful contribution of CESBIO. Arnaud Mialon and Ali Mahmoodi optimized the code, improved the data format and processed the data set at CESBIO. Amen Al-Yaari made the analysis of the IC data and produced all the figures; Yann Kerr, Gabrielle de Lannoy, Ahmad Al Bitar and Ernesto Lopez-Baeza provided scientific expertise; Roberto Fernandez-Moran and Jean-Pierre Wigneron wrote the paper.

Conflicts of interest

The authors declare no conflicts of interest.
References


55. Leroux, D. J.; Kerr, Y. H.; Bitar, A. A.; Bindlish, R.; Member, S.; Jackson, T. J.; Berthelot, B.; Portet, G. Comparison Between SMOS, VUA, ASCAT, and ECMWF soil moisture products over four watersheds in US. 2014, 52, 1562–1571.


