CORRELATED TRIPLE COLLOCATION TO ESTIMATE SMOS, SMAP AND ERA5-LAND SOIL MOISTURE ERRORS

M.Pablos[1,2], A. Turiel[1,2], M. Vall-llossera[2,3], A. Camps[2,3], and M. Portabella[1,2]

[1]Institut de Ciències del Mar (ICM), Consejo Superior de Investigaciones Científicas (CSIC)
Pg. Marítim de la Barceloneta 37-49, 08003, Barcelona, Spain. Email: mpablos@icm.csic.es.
[2]Barcelona Expert Center (BEC) on Remote Sensing,
Pg. Marítim de la Barceloneta 37-49, 08003, Barcelona, Spain.
[3]CommSensLab, Universitat Politècnica de Catalunya (UPC) and IEEE/CTE-UPC,
Jordi Girona 1-3 (Campus Nord), 08034, Barcelona, Spain.

ABSTRACT

The novel Correlated Triple Collocation (CTC) analysis allows to assess three different data sources of similar spatial resolutions, but with two of them being correlated. In this study, the CTC was applied to estimate the unbiased random errors of the global soil moisture (SM) data provided by two L-band satellite missions —the Soil Moisture and Ocean Salinity (SMOS) and the Soil Moisture Active Passive (SMAP)— and one numerical model —the ERA5-Land. The three existing SMOS SM products distributed by different research institutions were also analyzed.

Preliminary results revealed that errors of SMOS and SMAP SM are correlated, with correlations of ~0.5–0.6. Thus, only ERA5-Land can be considered as independent. The lowest error was obtained for SMAP (0.025 m^3m^{-3}), followed by ERA5-Land (0.036 m^3m^{-3}). Among the SMOS SM, SMOS-IC had the lowest error (0.046 m^3m^{-3}), SMOS-BEC showed an intermediate value (0.048 m^3m^{-3}), and SMOS-CATDS had the highest error (0.055 m^3m^{-3}).

Index Terms— Soil moisture, triple collocation, SMOS, SMAP, ERA5-Land.

1. INTRODUCTION

Nowadays, L-band radiometry is considered the most suitable technique for remotely measuring soil moisture (SM) at a global scale. With this objective, the Soil Moisture and Ocean Salinity (SMOS) mission, launched by the European Space Agency (ESA) in 2009, is the first satellite with an L-band radiometer on board. The second mission specifically designed to monitor SM and carrying an L-band radiometer is the Soil Moisture Active Passive (SMAP), which was launched by the National Aeronautics and Space Administration (NASA) in 2015.

Both SMOS and SMAP are currently in orbit and acquire global full-polarized observations of the Earth surface every 3 days at a spatial resolution of ~40 km. However, each satellite has a different architecture and instrument characteristics. SMOS includes a synthetic aperture radiometer feeded by 69 patch antennas distributed on its arms forming a Y-shape and, thus, provides brightness temperature (TB) observations at different incidence angles (0–60°) in ascending (6:00 h, equatorial crossing local time) and descending (18:00 h) orbits. In contrast, SMAP has a real aperture radiometer, feeded by a conically scanning mesh reflector with 6 m of diameter, and acquires TB at a constant incidence angle of 40° in descending (6:00 h) and ascending (18:00 h) passes. All these differences lead to differences in the accuracy of their SM products.

The classical validation approaches to assess a remotely sensed SM dataset are based on the comparison against collocated *in situ* SM observations. Notwithstanding, these methods are only capable of characterizing differences from *in situ* data at particular field sites. In the last decade, the Triple Collocation (TC) has been widely used to analyze three SM datasets obtained by mutually independent satellite platforms/observation techniques and at different spatial scales, such as satellite, modeled and *in situ* observations [1, 2, 3]. The TC considers an unknown ground truth and assumes the errors of different data sources are uncorrelated between them, but this assumption is not always fulfilled [2, 3]. In order to overcome this limitation, the Quadruple Collocation (QC) adds a fourth independent SM dataset to have a common reference [4], while the Extended Collocation (EC) needs to know the error cross-variance between two SM datasets *a priori* [5]. Moreover, the Extended Quadruple Collocation (E-QC) includes the both aforementioned requirements [6]. In most cases, the challenge of finding three or more independent data sources and the difficulty for knowing the error cross-variance makes it impossible to apply these techniques. As an alternative, the Correlated Triple Collocation (CTC) allows to estimate global maps of the errors of three collocated datasets with similar spatial resolution, of which two of them (the first and the second) are correlated with unknown error covariance, and the third one is totally independent of the others [7].

The aim of this study was to estimate the unbiased random errors of global SM measured by SMOS and SMAP missions by means of a CTC analysis. The ERA5-Land model from the European Centre for Medium-Range Weather Forecasts (ECMWF) acted as the third uncorrelated SM in the triplets. The study period (from August 2019 to February 2020, 213 days) is limited by the availability of the most recent releases of all data sources up to date. The assessed SMOS products are: i) the SMOS Level 3 (L3) SM provided by the Centre Aval de Traitement des Données SMOS (CATDS) [8, 9], hereafter SMOS-CATDS; ii) the SMOS L3 SM provided by the Barcelona Expert Center (BEC) on Remote Sensing [10], hereafter SMOS-BEC; and iii) the SMOS L3 SM provided by the Institut National de la Recherche Agronomique (INRA) and the Centre d'Etudes Spatiales de la Biosphère (CESBIO), hereafter SMOS-IC. The analyzed SMAP product corresponds to the NASA SMAP L3 SM [11]. This error characterization will give a deeper insight of the current SMOS and SMAP SM products.

2. DATA AND METHODS

2.1. SMOS-CATDS soil moisture

The global SMOS-CATDS SM is retrieved by a multi-orbit processor that allows to estimate SM together with Vegetation Optical Depth (VOD) [8, 9]. A filtering is firstly applied to remove the horizontal and vertical TB measurements out of the alias free field of view (AF-FOV) or with previously identified Radio-Frequency Interferences (RFI) and Sun glints. Similarly to the official Level 2 (L2) SM algorithm, the forward model of the CATDS processor is the L-band Microwave Emission of the Biosphere (L-MEB) radiative transfer model and the intra-pixel surface heterogeneity is considered for the retrievals.

In this research, daily ascending SMOS-CATDS L3 SM maps of the latest version (v300) at a 25-km Equal Area Scalable Earth (EASE)-2 grid were used. Additionally, pixels with a RFI probability flag value higher than 0.5 were discarded.

2.2. SMOS-BEC soil moisture

Differing from SMOS-CATDS and SMOS-IC, the SMOS-BEC SM is not retrieved, but directly obtained from ESA L2 Soil Moisture User Data Product (SMUDP) v650 after applying a quality filtering and a data binning [10]. A first filter removes the failed L2 retrievals, those with a value out of range or obtained from TB measurements prone to RFI at horizontal and/or vertical polarization. Later, a specific quality filter removes the retrievals with a Data Quality Index (DQX) higher than 0.07 m^3m^{-3}. Finally, the swath orbit-based data are binned to a daily global map using a weighted average inversely proportional to the DQX of the retrieval.

Daily ascending SMOS-BEC L3 SM maps v3.0 at an EASE-2 grid of 25 km were employed in this work.

2.3. SMOS-IC soil moisture

The SMOS-IC SM is retrieved by an alternative multi-orbit algorithm [11]. It simultaneously estimates global SM and VOD from horizontal and vertical TB measurements at incidence angles from 20 to 55°. The IC algorithm uses the same physical model than the ESA L2 and CATDS L3 processors, but simplifying the algorithm by considering homogeneous pixels and non accounting for complex antenna pattern corrections.

In this study, daily ascending SMOS-IC L3 SM maps of the new version (v2) at a 25-km EASE-2 were used. Following the post-processing recommendations for global applications, only pixels classified as moderate topography and with a Root Mean Square Error (RMSE) between modeled and measured TB lower or equal than 8 K were used.

2.4. SMAP soil moisture

The SMAP SM is retrieved by the single channel algorithm (SCA) as baseline [12]. The physical forward model of the SMAP passive processors is the *tau-omega* model, which is applied to the vertical TB, using a VOD estimated from a Normalized Difference Vegetation Index (NDVI) climatology, and land cover-based look up tables for the roughness and the single scattering albedo information.

Daily descending global SMAP L3 SM maps of the recently reprocessed version (v7) at EASE-2 grid of 36 km were used. Data was filtered to only preserve attempted and successful retrievals, using the retrieval quality flag. Finally, all maps were linearly interpolated to a 25-km EASE-2 to match the grid of the SMOS datasets.

2.5. ERA5-Land soil moisture

The ERA5-Land model is based on the Tiled ECMWF Scheme for Surface Exchanges over Land incorporating land surface hydrology (H-TESSEL), particularly the c45r1 of the Integrated Forecasting System (IFS), to provide hourly estimates of land variables at a native resolution of 9 km. Basically, ERA5-Land is a single replay of the land component of the ERA5 climate reanalysis. ERA5 assimilates a huge amount of data observations. However, ERA5-Land does not have data assimilation, but it uses the simulated ERA5 variables over land as atmospheric forcing [13].

In this study, daily global maps of ERA5-Land volumetric soil water layer 1 (0–7 cm) at 12 UTC, projected in a regular grid of 0.1°, were employed. As in the case of SMAP, the maps were linearly interpolated to a EASE-2 grid of 25 km.

2.6. Correlated Triple Collocation analysis

The selection of the beginning of the study period was determined by a correction performed in the CATDS L3 processor [14], and the ending by the availability of SMOS-IC data. Since both SMOS and SMAP have a revisit time of 3 days, a moving averaging window of 3 days was applied to daily SM to maximize the number of collocations. In the resulting 3-day averaged SM maps, only coincident samples in all datasets were taken into account. This pre-processing ensured the number of collocated samples was the same (regardless of the triplet analyzed), with an average of 99 samples. This number of samples is enough (N>50) because the CTC works appropriately even for a lower number [7].

The CTC provides the error standard deviation (std) of the three datasets included in the triplet as well as the error correlation (error-R) between the first and the second data sources [7]. The error-R was utilized to evaluate the degree of dependency between these two data records. The method was applied to a variety of triplets, using all possible combinations. However, only results from valid triplets, defined as those that satisfy the CTC assumptions, were analyzed.

3. RESULTS

The error correlations between two SM data sources are summarized in Table 1. As expected, all errors of SMOS are mutually correlated (error-R∼0.6–0.7). This is consistent with the fact that all of them are derived from observations acquired by the same instrument. Surprisingly, errors of SMOS-IC and SMAP are also correlated (error-R ∼0.6). A possible hypothesis for this could be the use of similar ancillary data in their respective retrieval algorithms, even though SMOS and SMAP instruments are independent and, consequently, their TB measurements too. No correlation was found between the errors of both SMOS-CATDS and SMOS-BEC and that of ERA5-Land, and between SMAP and ERA5-Land (error-R∼0.3). In view of these results, the CTC was the unique method that can be applied because only ERA5-Land SM could be considered as uncorrelated from the other data sources.

Figure 1 shows the unbiased random error std of each dataset. In all SMOS cases, the error has very similar patterns, displaying higher values in boreal and tropical forests. This is because SM retrievals are, in general, more accurate over areas with low or moderate vegetation than over those with a dense vegetation cover. Analyzing the different SMOS SM, SMOS-CATDS shows the highest error (0.055 m^3m^{-3}, in mean), SMOS-BEC has an intermediate value (0.048 m^3m^{-3}) and SMOS-IC exhibits the lowest (0.046 m^3m^{-3}). The SMAP SM error is lower than those estimated for all SMOS datasets (0.025 m^3m^{-3}). This is

Table 1. World average of the error correlations.

SM data sources		error-R
SMOS-CATDS	SMOS-BEC	0.68
SMOS-BEC	SMOS-IC	0.65
SMOS-IC	SMOS-CATDS	0.64
SMOS-CATDS	SMAP	0.50
SMOS-BEC	SMAP	0.47
SMOS-IC	SMAP	0.63
SMOS-CATDS	ERA5-Land	0.35
SMOS-BEC	ERA5-Land	0.28
SMOS-IC	ERA5-Land	0.49
SMAP	ERA5-Land	0.34

explained by the different characteristics of the instruments; the interferometric SMOS radiometer is noisier than the real aperture SMAP one. The error of ERA5-Land is reasonable low (0.036 m^3m^{-3}), but arid regions, such as Sahara desert and Arabian Peninsula, display a higher error.

4. CONCLUSIONS

The SMOS, SMAP and ERA5-Land SM were analyzed from August 2019 to February 2020 by means of the novel Correlated Triple Collocation (CTC). This method has the advantage of not requiring three completely independent datasets, which is sometimes difficult to achieve. Instead, the CTC is able to be applied to triplets composed of two correlated and one uncorrelated datasets that resolve similar spatial scales. This is of special interest because SMOS and SMAP SM revealed to be error-correlated between them, and only ERA5-Land SM was considered as uncorrelated in the preliminary assessment.

Both SMOS and SMAP SM showed higher errors over densely forested areas. Thus, vegetation has an important impact on the accuracy of the passive microwave retrievals. The resulting unbiased random error of SMOS SM was ∼0.050 m^3m^{-3}, while that of SMAP was 0.025 m^3m^{-3}, indicating SMAP is more accurate than SMOS. This is justified by the different characteristics of the radiometers. Among the three SMOS datasets, SMOS-IC has the lowest error, followed by SMOS-BEC, and SMOS-CATDS has the highest. ERA5-Land SM showed higher error over arid regions, with a mean of 0.036 m^3m^{-3}, which was between the SMAP and the SMOS ones.

ACKNOWLEDGMENTS: This work has been supported by the Spanish Ministry of Science and Innovation through the projects ESP2017-89463-C3-1R and ESP2017-89463-C3-2R, the ICM-CSIC Severo Ochoa Excellence Award CEX2019-000928-S, the CommSensLab-UPC Mar´ıa de Maeztu Excellence Award MDM-2016-0600, and the CSIC Interdisciplinary Thematic Platform TELEDETECT.

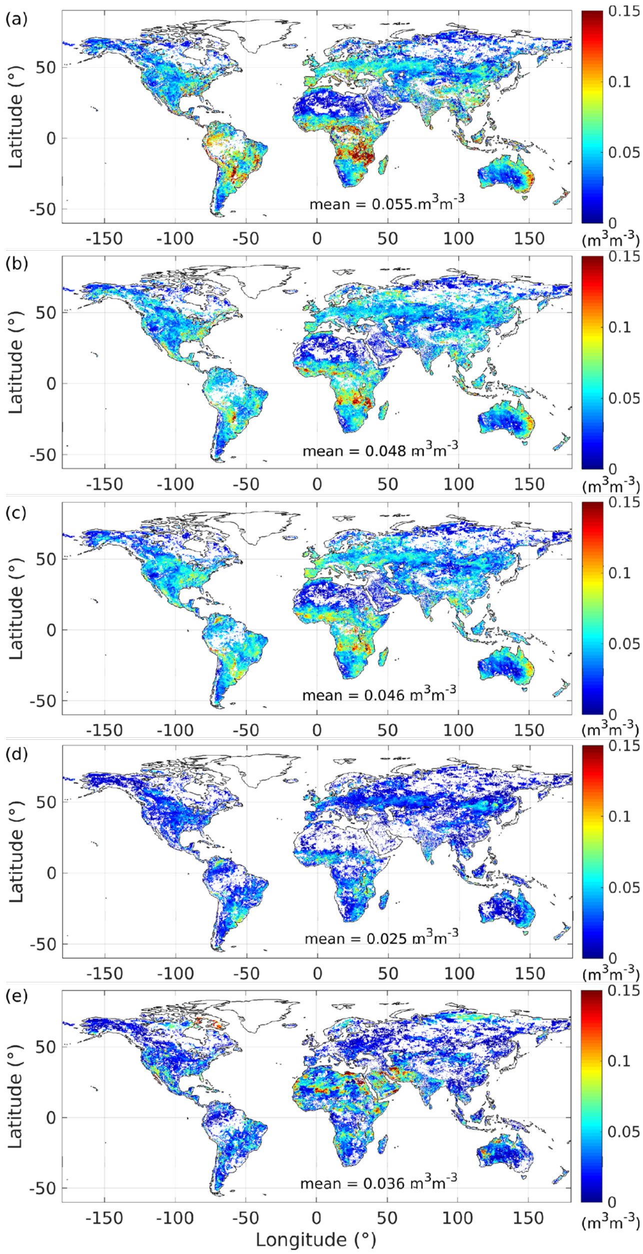

Fig. 1. SM error std estimated for SMOS-CATDS (a), SMOS-BEC (b), SMOS-IC (c), SMAP (d), and ERA5-Land (e).

5. REFERENCES

[1] Dorigo *et al.*, "Error characterisation of global active and passive microwave soil moisture datasets," *Hydrol. Earth Syst. Sc.*, vol. 14, no. 12, pp. 2605–2616, 2010.

[2] Yilmaz and Crow, "Evaluation of assumptions in soil moisture triple collocation analysis," *J. Hydrometeorol.*, vol. 15, no. 3, pp. 1293 – 1302, 2014.

[3] Gruber *et al.*, "Recent advances in (soil moisture) triple collocation analysis," *Int. J. Appl. Earth Obs.*, vol. 45, pp. 200 – 211, 2016.

[4] Pierdicca *et al.*, "Quadruple collocation analysis for soil moisture product assessment," *IEEE Geosc. and Remote Sens. Lett.*, vol. 12, no. 8, pp. 1595–1599, 2015.

[5] Gruber *et al.*, "Estimating error cross-correlations in soil moisture data sets using extended collocation analysis," *J. Geophys. Research: Atmospheres*, vol. 121, no. 3, pp. 1208–1219, 2016.

[6] Pierdicca *et al.*, "Error characterization of soil moisture satellite products: Retrieving error cross-correlation through extended quadruple collocation," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 10, no. 10, pp. 4522–4530, 2017.

[7] González-Gambau *et al.*, "Triple collocation analysis for two error-correlated datasets: Application to l-band brightness temperatures over land," *Remote Sens.*, vol. 12, no. 20, pp. 3381, 2020.

[8] Kerr *et al.*, "CATDS SMOS L3 soil moisture retrieval processor algorithm theoretical baseline document (ATBD)," Tech. Rep. SO-TN-CBSA-GS-0029, Issue 2.0, CBSA, 2013.

[9] Al Bitar *et al.*, "The global SMOS level 3 daily soil moisture and brightness temperature maps," *Earth Sys. Sci. Data*, vol. 9, no. 1, pp. 293–315, 2017.

[10] Pablos *et al.*, "BEC SMOS soil moisture products description," Tech. Rep. BEC-SMOS-PD-SM-L3v3-L4v5, version 1.0, BEC, 2020.

[11] Wigneron *et al.*, "SMOS-IC data record of soil moisture and L-VOD: historical development, applications and perspectives," *Remote Sens. Environ.*, vol. 254, pp. 112238, 2021.

[12] O'Neill *et al.*, "SMAP algorithm theoretical basis document level 2 & 3 soil moisture (passive) data products," Tech. Rep. JPL D-66480, Revision F, JPL, 2020.

[13] Muñoz-Sabater *et al.*, "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," *Earth Sys. Sci. Data*, In preparation.

[14] CATDS, "Correction of the L3 SM and VOD from CATDS," https://www.catds.fr/News/Correction-of-the-L3-SM-and-VOD-from-CATDS, Last access: 14/01/2021.