H-SAF Visiting Scientist Program

HSAF_CDOP2_VS12_02

Root-zone soil moisture index complementary validation at global scale based on triple collocation method. Comparison with State-Of-The-Art global scale root-zone soil moisture products.

Thierry Pellarin(1), Patricia de Rosnay(2), Clement Albergel(2), Saleh Abdalla(2),

Ahmad al Bitar(3)

(1) LTHE, Grenoble, France

(2) ECMWF, Reading, UK

(3) CESBIO, Toulouse, France

Contents

1) Introduction

The Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) is currently in Continuous Development and Operational Phase-2 (CDOP-2, 2012-2017). A major objective of the H-SAF is to provide high quality operational global scale soil moisture product both in surface and in the root zone. In this framework ECMWF produces the ASCAT Root Zone Soil Moisture profile Index (H14) based on ASCAT data assimilation in a dedicated Land Data Assimilation chain. Since July 2012, H14 is produced operationally by the H-SAF.

During the H-SAF CDOP phase H14 was under development. A Visiting Scientist (VS) study, conducted by Claire Gruhier from CESBIO, allowed to validate and to compare for the reference year 2010, three surface soil moisture products: the ASCAT CAF product, the demonstration data set of H14 and the under development SMOS level2 surface soil moisture. Results were published by Gruhier et al. (2011) and Albergel et al. (2012), showing that, compared to ground station measurements the H14 demonstration dataset outperforms the ASCAT and the SMOS level2 surface soil moisture products.

However this study was mostly limited to the surface soil moisture because a limited number of ground stations actually provide soil moisture measurements in the root zone. And when they do measure soil moisture at depth, the vertical sampling is very sparse, limited at most to one or two measurements in the root zone. So, the vertical representativeness of current ground measurement networks for soil moisture makes it difficult to rely on ground data to evaluate global root zone soil moisture products. Furthermore, since the surface soil moisture products validation study was conducted in 2011, the three products retrieval algorithms have been continuously improved. H14 production chain was considerably improved. For H14 the ASCAT bias correction approach was revised (de Rosnay et al., 2011) and H14 became operational in July 2012. SMOS retrieval algorithm was much improved as well as the SMOS signal and RFI filtering, and the ASCAT CAF product was also substantially improved in operations from August 2011. For SMOS a first version of the SMOS level-4 product is available since summer 2013. Other root zone soil moisture products are available nowadays, either based on re-analysis such as the Modern-Era Retrospective Analysis for Research and Applications for Land (MERRA-Land) soil moisture product (Reichle et al., 2011), or the Geoland Soil Wetness Index product (GL-SWI) based on a Sequential Kalman filtering (Lacaze et al., 2012).

The objective of the proposed study is to compare and to quantify the accuracy of four root zone soil moisture products provided by the Copernicus Global Land service (GL-SWI Soil Wetness Index based on ASCAT surface soil moisture estimates), SMOS level-4 root zone soil moisture, MERRA-Land re-analysis and the H-SAF H14 root zone soil moisture profile index. The first part of the study will be devoted to compare the four root-zone soil moisture products with in situ soil moisture measurements in US and West Africa. In the second part, a triple collocation method will be used to assess the accuracy of the four root-zone soil moisture products. Although several studies already used triple collocation in the past few years to evaluate surface soil moisture product accuracy (e.g. Scipal et al., 2008), they all focused on surface soil moisture. In this study triple collocation permits for the first time to quantify root zone soil moisture product accuracy at the global scale.

2) Data

Four root-zone soil moisture products and two ground-based measurement networks were used in this study and are described on Table I (H14, GL-SWI, MERRA-Land, SMOS and in situ measurements). Figure 1 illustrates the annual mean soil moisture for 2012 for the four global scale root-zone soil moisture products. Global maps of correlation and RMS between the four different products are shown in Annex B.

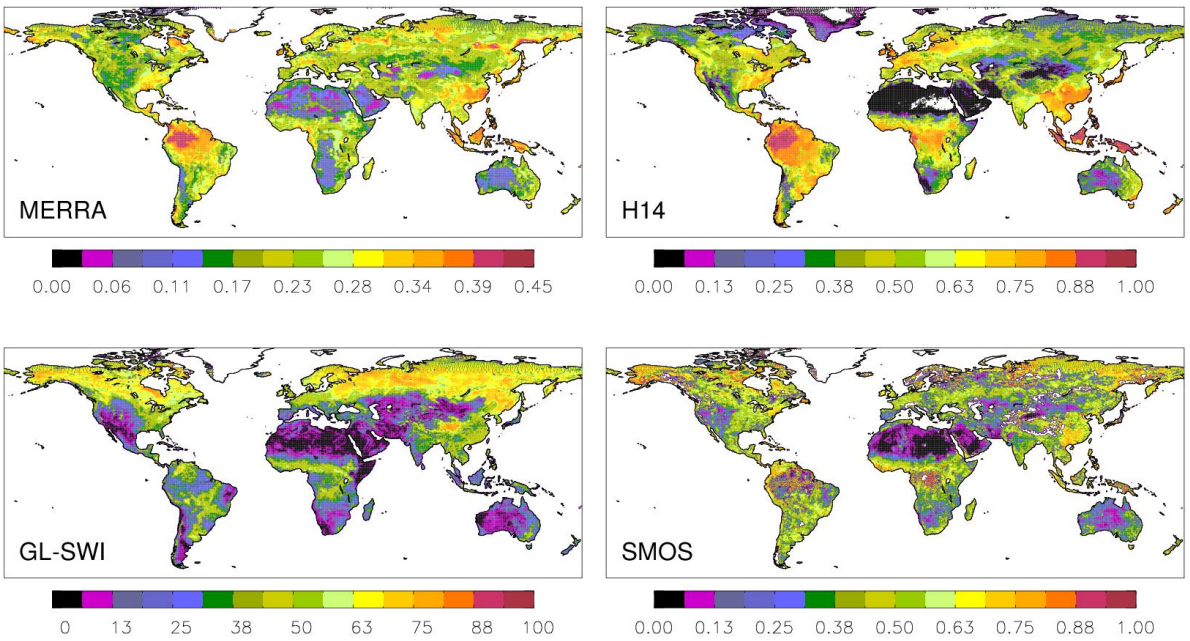


Figure 1: Annual mean root-zone soil moisture maps for MERRA, H14, GL-SWI and SMOS.

2.1 In situ measurements of soil moisture

This study makes use of in situ soil moisture measurements obtained through the International Soil Moisture Network (ISMN, http://www.ipf.tuwien.ac.at/insitu/, Dorigo et al. 2011, 2012), a data hosting centre where globally-available ground-based soil moisture measurements are collected, harmonized and made available to users. Data from 2 networks are considered for 2012: USCRN over the United States (90 stations) and AMMA in West Africa (2 stations). Table I gives a full list of reference for each network. For the specific 2012 year, 92 stations are available. All the considered networks used in this study also measure temperature, it permits to remove observations potentially affected by frozen condition.

Soil Moisture data set	**Type**	**Soil layer depth used in this study(cm)**	**Period**	**Spatial resolution**	**Number of stations**
MERRA-Land ***NASA-GMAO*** **http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset.pl**	Model product	0-100	2012	1/2º - Lat 2/3 º - Lon	Global product
SMOS level 4 root zone soil moisture product ***CESBIO***	Satellite product	20-100	2012	1/4°	Global product
GL-SWI ***Copernicus Global Land service*** **http://land.copernicus.eu/global/products/swi**	Satellite product	Tau parameter is 15days	2012	10 km	Global product
H14 ***ECMWF*** **http://hsaf.meteoam.it/soil-moisture.php**	Analysis product	0-7, 7-28 and 0-100	2012	25 km	Global product
USCRN (USA) **http://www.ncdc.noaa.gov/crn/**	In situ observations	5, 10, 20, 50 and 100	2012	Local scale	90 stations
AMMA (West Africa) **http://database.amma-international.org/**	In situ observations	5, 10, 20, 30, 60,100	2012	Local scale	2 stations

Table I: Description of the soil moisture data set used in this study

2.2 MERRA-Land product

MERRA-Land stems of the MERRA reanalysis generated by the NASA Global Modeling and Assimilation Office (GMAO) using the Goddard Earth Observing System (GEOS) version 5.2.0 (Rienecker et al. 2011; http://gmao.gsfc.nasa.gov/research/merra/). MERRA

incorporates information from in situ and remote sensing observations of the atmosphere, including many modern satellite observations. All these observations are assimilated into the GEOS-5 Atmospheric General Circulation Model. MERRA covers the period from 1979 onwards and continues to be updated a few weeks in arrears, but it does not include a land surface analysis. Estimates of surface meteorological and land surface fields for MERRA are available at hourly time steps and at 1/2º and 2/3º resolution in latitude and longitude.

A supplementary and improved product of land surface hydrological fields called MERRA-Land (Reichle et al. 2011) was generated by rerunning a revised version of the land component of the MERRA system. Compared to MERRA, MERRA-Land benefits from corrections to the precipitation forcing with the National Oceanic and Atmospheric Administration Climate Prediction Center Unified daily precipitation product and from revised parameter values in the rainfall interception model. Root zone soil moisture from MERRA-Land is used (0-100cm).

2.3 SMOS L4 Root zone soil moisture index

The SMOS root zone soil moisture index is produced by the CESBIO SMOS team (Al Bitar et al. 2013). The main inputs to this product are the SMOS L3 surface soil moisture products from CATDS. Ancillary data like ECMWF climate data and MODIS NDVI are used also to compute the transpiration in deeper layers. The root zone soil moisture is obtained using a double bucket model. The model is based on a modified exponential filter for the first layer and a physical model for the second layer. The first layer delivers an hourly soil moisture product. This product is then used in the second layer model to compute the deeper layer soil moisture. The model of the second layer is a linear formulation of the 1D Richards equations that govern water flow in unsaturated porous media. This model takes into account the transpiration of the vegetation using the FAO method, the deep layer infiltration using a gravity flow condition. In this study the second layer root zone soil moisture index is used and basically represents a 20-100 cm soil layer. The product is presented at a 1/4° regular grid in latitude and longitude.

2.4 GL-SWI product

The Global Land Soil Water Index (GL-SWI) product has been developed in the framework of geoland2 project. It is generated and distributed by the Copernicus Global Land service. It provides global daily information on soil moisture conditions in different soil layers. The SWI product is derived from EUMETSAT ASCAT SSM data at 25 km resolution by using a methodology developed at TU Wien (Wagner et al., 1999), and adapted by Albergel et al. (2008), based on a simple exponential filter. It requires a single parameter; Tau a surrogate parameter for all the processes affecting the temporal dynamics of soil moisture, such as the thickness of the soil layer, soil hydraulic properties, evaporation, run-off and vertical gradient of soil properties (texture, density). Tau is a characteristic time length; the time scale of soil moisture variation, in units of day.

2.5 H14 product

H14 is an offline analysis produced at ECMWF in the framework of the EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management "(HSAF)" project (more information at http://hsaf.meteoam.it/index.php). In its current version, it uses ECMWF Integrated Forecast System (IFS) version 38r2 (more information at http://www.ecmwf.int/research/ifsdocs/). It is a global product with a spatial resolution of about 25 km (T799) available daily at 00:00. H14 production relies on an advanced surface data assimilation system used to optimally combine conventional observation with satellite measurements; and Extended Kalman Filter as described in Drusch et al. (2009); de Rosnay et al. (2013). In the H14 production chain, the EKF soil moisture analysis uses observations from the SYNOP network i.e. two-metre air temperature and two-metre air relative humidity as well as ASCAT surface soil moisture as input. The land surface model used is HTESSEL (Van den Hurk and Viterbo, 2003, Balsamo et al., 2009), a multilayer model where the soil is discretized in four layers of 0.07, 0.28, 1.00 and 2.89 m depths (from top to bottom). Albergel et al., 2012 presented a first evaluation of H14 using in situ measurements. While they have focused the upper layer of soil (0-7cm), this study extends the validation to deeper soil horizons down to the first meter of soil.

3) Ground-based comparison

3.1 Methodology

The nearest neighbour approach was retained to match grid point location of soil moisture from H14, GL-SWI, MERRA-Land and SMOS with that of ground measurements. Datasets potentially affected by frozen condition were masked using a soil temperature threshold of 4 degrees ºC. GL-SWI estimates represent a relative measure of the soil moisture given between 0 and 100, H14 and SMOS are an index [0-1] while the MERRA-Land product is expressed in m^3m^{-3}. To enable a fair comparison, each product was normalised using their own minimum and maximum values over 2012. The GL-SWI product is available with different Tau values. In a preliminary evaluation, it was found (not shown) that a value of 15 days best represents the observations over the first meter of soil. This value is used in this study.

The ability of the four products to reproduce root zone soil moisture variability is assessed using ground measurement based on the correlation coefficient (R) and the root mean square differences (RMSD). Pixels with non-significant R values (p-value larger than 0.05) are

excluded from the computation of the average metric. As in Albergel et al. (2013), two additional metrics; the normalised standard deviation (SDV) and the centred unbiased RMSD (E) were computed (not shown). The rational for computing E and SDV is that along with R they can be displayed on a single two-dimensional diagram (Taylor diagram) and this helps with the interpretation of the results (Taylor, 2001).

3.2 Results

For all the stations used in this study, a first visual quality check was performed. When suspicious data were observed (e.g. non-physical jumps in the time series), they were discarded. The results are presented in Table II.

For this period, and for the 4 products, 90 stations have R values that are different from zero at the 5% significance level (p-value<0.05) for the three products, their scores are averaged per network. Comparisons between in-situ measurements and the three products show good temporal correlations. On the US domain, H14 and MERRA-Land products obtain the highest correlation coefficients (R=0.78 and 0.77 respectively), whereas SMOS and GL-SWI obtain R=0.73 and R=0.64 respectively. The RMSD (unit-less) represents the relative error of the soil moisture dynamical range. The average RMSD values of US network are 0.189, 0.228, 0.237 and 0.202 for MERRA-Land, SMOS, GL-SWI and H14 respectively. On the AMMA domain, SMOS and H14 products obtain the highest correlation coefficients (R=0.93 and 0.88 respectively), whereas MERRA-Land and GL-SWI obtain R=0.72 and R=0.83 respectively. Regarding the RMS scores, H14 and SMOS products obtain similar results (RMS=0.151) and MERRA-Land and GL-SWI obtain slightly greater RMS equal to 0.259 and 0.201 respectively.

Figure 2 shows a Taylor diagram illustrating the averaged statistics from the comparisons. These results underline the good range of correlations with most values being above 0.60 for the three products. The black dashed line in the Taylor diagrams represents a SDV value of one, as SDV is the ratio between the standard deviations of a product and in situ measurement; a symbol located below this line indicates that the product has less variability than the in situ measurements (SDV < 1). In most cases, root zone soil moisture from H14 and MERRA-Land better correlate with in situ measurements that one from GL-SWI.

		R				RMSD [-]			
2012	N stations	MERRA-Land	SMOS	GL-SWI	H14	MERRA-Land	SMOS	GL-SWI	H14
USCRN	90	0.77	0.73	0.64	0.78	0.189	0.228	0.237	0.202
AMMA	2	0.72	0.93	0.83	0.88	0.259	0.151	0.201	0.151

Table II: Correlation and RMSD between ground observations over two networks (USCRN and AMMA) and the four root-zone soil moisture products

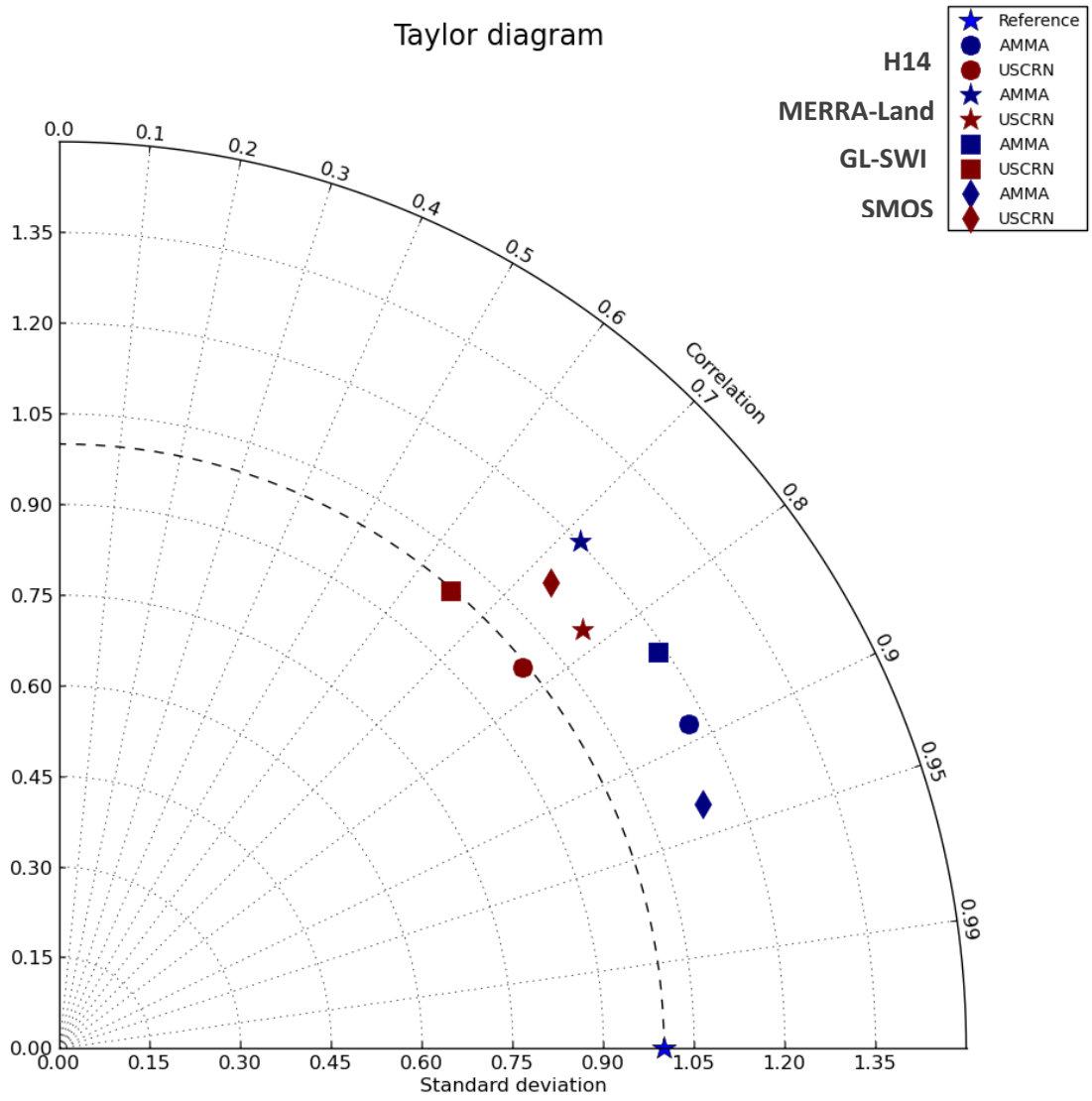


Figure 2: Taylor diagram illustrating the averaged statistics from the comparisons between the 4 products and in-situ data. Note that the USCRN symbols represent an average of 90 stations, whereas AMMA symbols represent an average of 2 stations.

4) Triple collocation

4.1 Triple collocation method

The triple collocation (TC) method was originally developed to estimates the errors associated with three time-series of a given variable without knowing the true time-series. The first application of this method was applied to sea wind estimates derived from buoys, model and satellite data (Stoffelen, 1998) then applied to other parameters including the

significant wave height (Caires and Sterl, 2003; Janssen et al., 2007). Triple collocation is a powerful method to estimate the absolute random error in standard deviation sense in each of the three collocated datasets, provided the errors (and obviously not the time-series) are not correlated.

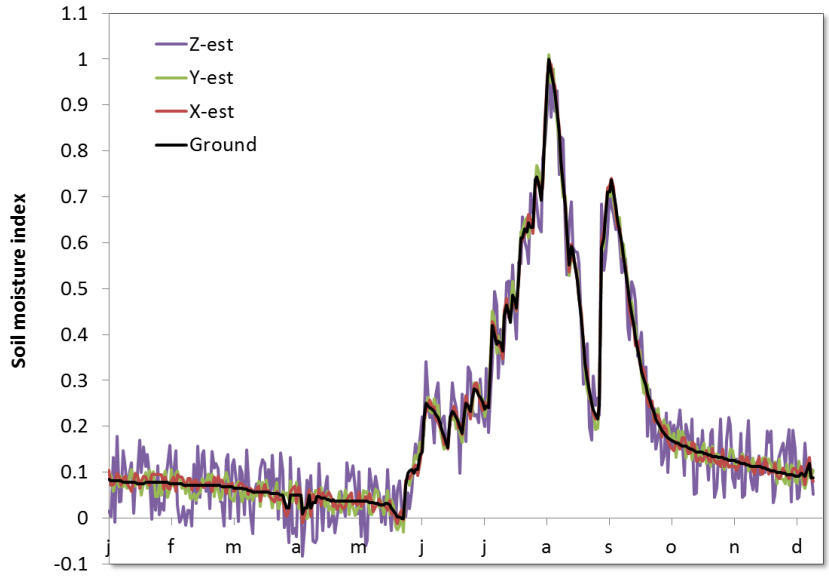
Scipal et al. (2008), applied the triple collocation to quantify uncertainties of surface soil moisture data, using data from two different remote sensing soil moisture products and ERA-40 reanalysis data. A range of studies were then conducted using the TC methodology to estimate point-to-footprint soil moisture sampling error (Miralles et al., 2010; Loew and Schlenz, 2011), to evaluate different surface soil moisture products (Hain et al., 2011; Draper et al., 2013), as a rescaling technique prior to data assimilation (Yilmaz et al. 2013). More recently, it was also used to estimate the spatial distribution of SMOS surface soil moisture errors at the global scale (Leroux et al., 2013). In the present study, the triple collocation method is applied for the first time to assess root-zone soil moisture products at the global scale.

To explain the concept of the triple collocation method, one can considers a given root-zone soil moisture time-serie (normalized between 0 –dry- and 1 –wet-) and presented with the black curve in figure 3. The "ground" time series is assumed to be the truth. Let's now consider three synthetic root-zone soil moisture estimates *X(t), Y(t)* and *Z(t)* computed as the true root-zone soil moisture associated with three different noise levels and illustrated in figure 3 with red, green and purple time-series.

$$X(t) = Truth(t) + \varepsilon_X(t) \quad (1)$$

$$Y(t) = Truth(t) + \varepsilon_Y(t) \quad (2)$$

$$Z(t) = Truth(t) + \varepsilon_Z(t) \quad (3)$$



$\langle \varepsilon_X^2 \rangle$ =1.33^{E-4}

$\langle \varepsilon_Y^2 \rangle$ =3.83^{E-4}

$\langle \varepsilon_Z^2 \rangle$ =3.18^{E-3}

$\langle \varepsilon_X . \varepsilon_Y \rangle$ =4.36^{E-6}

$\langle \varepsilon_X . \varepsilon_Z \rangle$ =-4.69^{E-5}

$\langle \varepsilon_Y . \varepsilon_Z \rangle$ =1.23^{E-5}

Figure 3: Example of Root-zone soil moisture time-series. The reference soil moisture is presented in black and the 3 synthetic estimates are presented in red (X), green (Y) and purple (Z). Corresponding mean square errors ($\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$) as well as cross errors

($\langle \varepsilon_X . \varepsilon_Y \rangle$, $\langle \varepsilon_X . \varepsilon_Z \rangle$ and $\langle \varepsilon_Y . \varepsilon_Z \rangle$) can be calculated for the three synthetic time series (right table).

In our example, the mean square error $\langle \boldsymbol{\varepsilon_X^2} \rangle$ was equal to $1.33^{\text{E-4}}$, the mean square error $\langle \boldsymbol{\varepsilon_Y^2} \rangle$ was equal to $3.83^{\text{E-4}}$ and the mean square error $\langle \boldsymbol{\varepsilon_Z^2} \rangle$ was equal to $3.18^{\text{E-3}}$ (see figure 3). Prior to apply the triple collocation analysis, the time-series require to be rescaled to each other to match their signal signature. By substracting of eq. (1) and (2), and in a similar way (1) and (3) as well as (2) and (3), the truth can be eliminated and we obtain:

$$X(t) - Y(t) = \varepsilon_X(t) - \varepsilon_Y(t) \quad (4)$$

$$X(t) - Z(t) = \varepsilon_X(t) - \varepsilon_Z(t) \quad (5)$$

$$Y(t) - Z(t) = \varepsilon_Y(t) - \varepsilon_Z(t) \quad (6)$$

Then, by multiplying the first with the second equation, and in a similar manner the first and the third and the second and the third, and by considering the mean values of each time-series, it can be written:

$$\langle (X - Y)(X - Z) \rangle = \langle \varepsilon_X^2 \rangle - \langle \varepsilon_X . \varepsilon_Y \rangle - \langle \varepsilon_X . \varepsilon_Z \rangle + \langle \varepsilon_Y . \varepsilon_Z \rangle \quad (7)$$

$$\langle (Y - X)(Y - Z) \rangle = \langle \varepsilon_Y^2 \rangle - \langle \varepsilon_X . \varepsilon_Y \rangle - \langle \varepsilon_Y . \varepsilon_Z \rangle + \langle \varepsilon_X . \varepsilon_Z \rangle \quad (8)$$

$$\langle (X - Z)(Y - Z) \rangle = \langle \varepsilon_Z^2 \rangle - \langle \varepsilon_X . \varepsilon_Y \rangle - \langle \varepsilon_Y . \varepsilon_Z \rangle + \langle \varepsilon_X . \varepsilon_Y \rangle \quad (9)$$

The main assumption of the triple collocation method is that the three estimated time-series have independent random errors. This means that, each term ε_X, ε_Y and ε_Z, which represent errors between the truth and *X*, *Y*, *Z* at a given time *t*, can be randomly either positive or negative. It follows that the product $\varepsilon_X . \varepsilon_Y$ can be either positive or negative and thus, if they have independent errors, the mean value $\langle \varepsilon_X . \varepsilon_Y \rangle$ is expected to tend to zero whereas square terms $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$ can only be positive.

Thus, cross-term $\langle \varepsilon_X . \varepsilon_Y \rangle$, $\langle \varepsilon_X . \varepsilon_Y \rangle$ and $\langle \varepsilon_X . \varepsilon_Y \rangle$ are assumed to be small compared to $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$ and the final equation of the triple collocation method can be written as :

$$\langle (X - Y)(X - Z) \rangle = \langle \varepsilon_X^2 \rangle \quad (10)$$

$$\langle (Y - X)(Y - Z) \rangle = \langle \varepsilon_Y^2 \rangle \quad (11)$$

$$\langle (X - Z)(Y - Z) \rangle = \langle \varepsilon_Z^2 \rangle \quad (12)$$

In our example (figure 3), as the true root-zone soil moisture is known, one can easily calculate the different terms of eq. (7), (8) and (9). Results are given in figure 3 (right Table). It can be confirmed that cross-terms $\langle \varepsilon_X . \varepsilon_Y \rangle$, $\langle \varepsilon_X . \varepsilon_Z \rangle$ and $\langle \varepsilon_Y . \varepsilon_Z \rangle$ are at least one order of magnitude smaller than square terms $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$. The triple collocation method consists of computing the following terms:

$$\langle (X - Y)(X - Z) \rangle = 1.88^{\text{E-4}}$$

$$\langle (Y - X)(Y - Z) \rangle = 3.19^{\text{E-4}}$$

$$\langle (X-Z)(Y-Z) \rangle = 3.22^{\text{E-3}}$$

It can be observed that the TC method gives an accurate estimation of the errors $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$ which were respectively equal to $1.33^{\text{E-4}}$, $3.83^{\text{E-4}}$ and $3.18^{\text{E-3}}$.

4.2 Application of the triple collocation at the global scale

As said in the previous section, prior to apply the triple collocation analysis, the time-series are required to be rescaled to each other to match their signal signature. This was done using the following simple rescaling procedure that forces the min and max of each time-series to range between 0 and 1. Note that this equation was also applied in the previous section for the evaluation against ground measurements.

$$X(t) = \frac{X^0(t) - \min(X^0(t))}{\max(X^0(t)) - \min(X^0(t))}$$

The same procedure is applied to obtain *Y(t)* and *Z(t)* time-series. Figure 4 illustrates the temporal evolution of four products (MERRA, SMOS, GL-SWI and H14) over two different pixels (in Tchad and Texas) in their original system of units, as well as the four rescaled time-series.

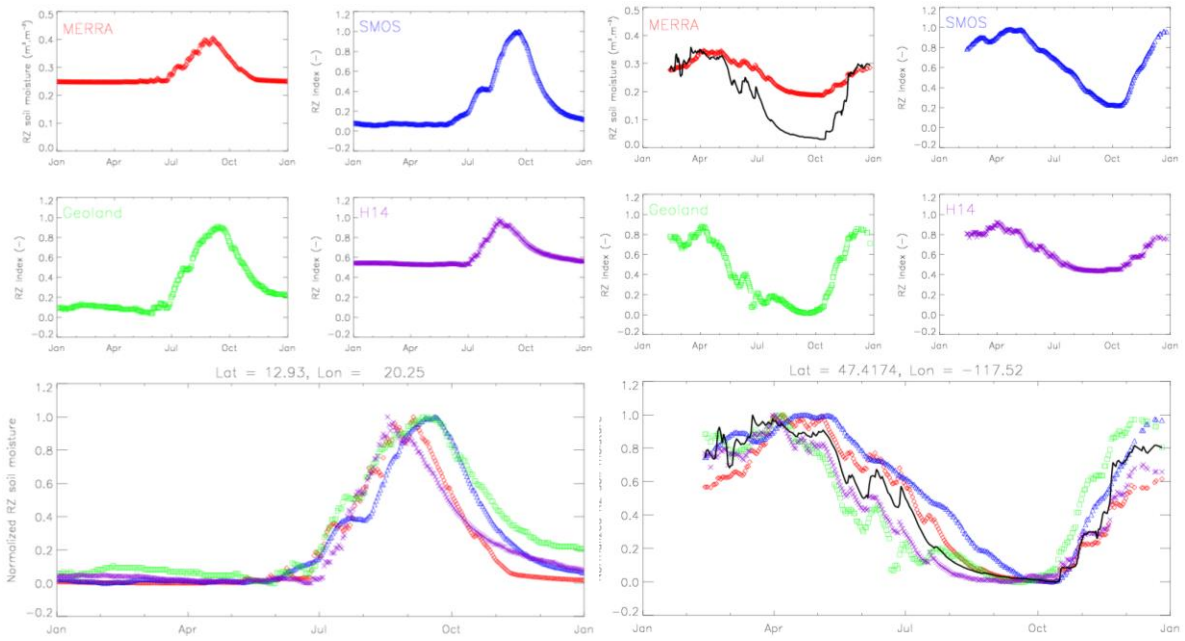


Figure 4: Temporal evolution of the four root-zone soil moisture products (MERRA, SMOS, GL-SWI and H14) in their original units, in Tchad, Africa (left graphs) and in Texas (right). The normalized products are shown on the bottom panels. Ground measurements (in m3.m-3), available over 92 pixels in US and West Africa are plotted in the Texas' graphs (top left and bottom right, black lines).

The triple collocation method was conducted at the global scale for various set of three times-series. Intuitively, it is considered that H14 and GL-SWI may not be independent since the two products are both using ASCAT soil moisture products. Thus, the triple collocation was applied on only two of the four possible triplets:

- MERRA-Land / SMOS / H14
- MERRA-Land / SMOS / GL-SWI

4.2.1 MERRA-Land / SMOS / H14 (TC#1)

Spatial distribution of calculated errors $\sqrt{\langle \varepsilon_X^2 \rangle}$, $\sqrt{\langle \varepsilon_Y^2 \rangle}$ and $\sqrt{\langle \varepsilon_Z^2 \rangle}$ is presented in figure 5. First, it can be observed that a significant amount of pixels appears in white in figure 5. On these pixels, the calculated square error $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$ is found to be negative which indicates that cross-terms $\langle \varepsilon_X . \varepsilon_Y \rangle$, $\langle \varepsilon_X . \varepsilon_Z \rangle$ and $\langle \varepsilon_Y . \varepsilon_Z \rangle$ are non-negligible. This is illustrated with the corresponding histograms (on the right of figure 5) and this behavior will be discussed in the annex section. The total amount of negative error pixels is about 13% for MERRA-Land, 6% for SMOS and 11% for H14.

When considering only pixels with positive errors, MERRA-Land product obtain the lowest error on 35% of the pixels, whereas SMOS obtain the lowest error on 29%, and H14 obtain the lowest error on 36% of the pixels. Looking in more details, it can be easily detected pixels where the three products provide low errors. If an acceptable error is arbitrary fixed to a threshold of $\langle \varepsilon^2 \rangle$=0.04 (i.e. $\sqrt{\langle \varepsilon^2 \rangle}$=0.20, blue/purple areas in figure 5), MERRA-Land obtains acceptable error on 43% of the pixels, SMOS on 38% of the pixels and H14 obtains acceptable error on 43% of the pixels.

The triple collocation method was also tested for various switched configurations of the triplet (e.g. SMOS / MERRA-Land / H14) and this has absolutely no incidence on the results. Considering the symmetric nature of equation (1,2,3), this choice does not influence the error estimation (Scipal et al., 2009).

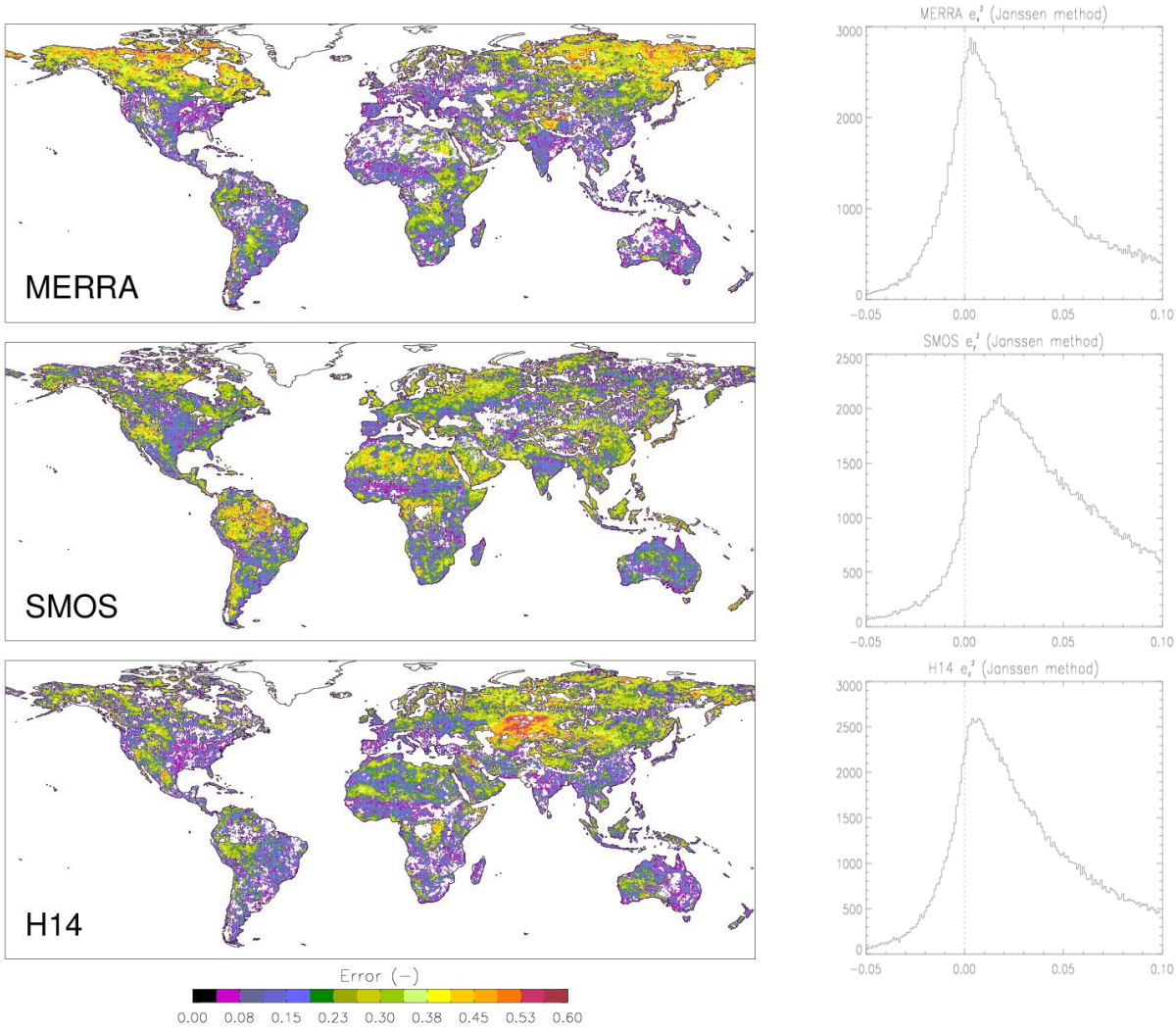


Figure 5: Spatial distribution of the calculated errors $\sqrt{\langle\varepsilon_X^2\rangle}$, $\sqrt{\langle\varepsilon_Y^2\rangle}$ and $\sqrt{\langle\varepsilon_Z^2\rangle}$ (left) and histograms of errors $\langle\varepsilon_X^2\rangle$, $\langle\varepsilon_Y^2\rangle$ and $\langle\varepsilon_Z^2\rangle$ (right). A significant part of the errors were found to be negative (about 13% for MERRA-Land, 6% for SMOS and 11% for H14) which seems to indicates that cross-terms $\langle\varepsilon_X.\varepsilon_Y\rangle$, $\langle\varepsilon_X.\varepsilon_Z\rangle$ and $\langle\varepsilon_Y.\varepsilon_Z\rangle$ are probably non-negligible.

4.2.2 MERRA-Land / SMOS / GL-SWI (TC#2)

The triple collocation method was applied to the triplet MERRA-Land / SMOS / GL-SWI and spatial distribution of calculated errors $\sqrt{\langle\varepsilon_X^2\rangle}$, $\sqrt{\langle\varepsilon_Y^2\rangle}$ and $\sqrt{\langle\varepsilon_Z^2\rangle}$ is presented in figure 6. In a similar way, there are pixels where the TC method leads to obtain negative errors which represent about 7% for MERRA-Land, 13% for SMOS and 9% for GL-SWI.

If considering only pixels with positive errors, MERRA-Land product obtain the lowest error on 31% of the pixels, whereas SMOS obtain the lowest error on 39%, and GL-SWI obtain the lowest error on 30% of the pixels. Looking in more details, an acceptable error (arbitrary fixed below a threshold of $\langle\varepsilon^2\rangle$=0.04) is obtained on 37% of the pixels with the MERRA-Land product whereas SMOS obtains good results on 44% of the pixels and GL-SWI obtains acceptable error on 36% of the pixels.

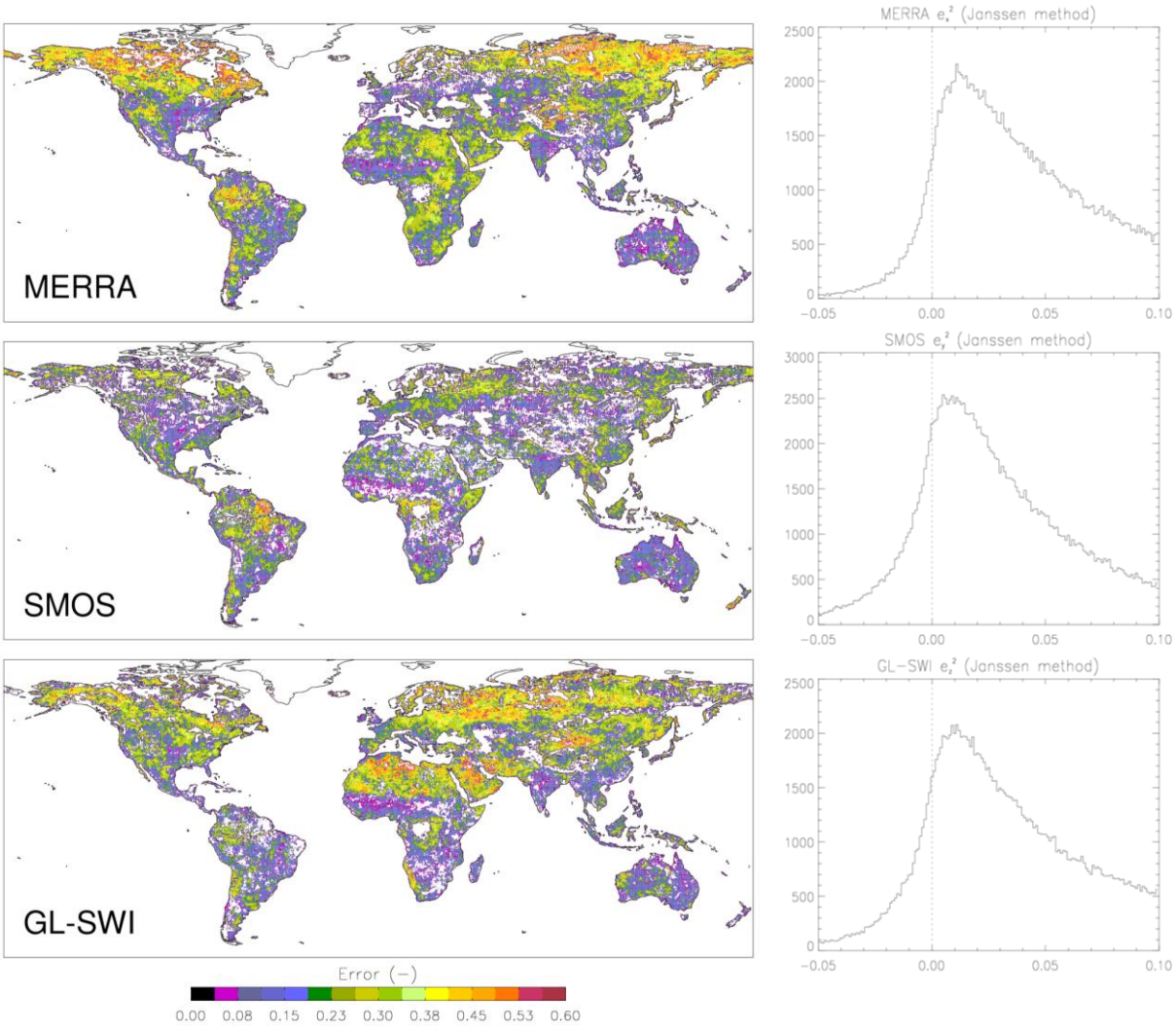


Figure 6: Spatial distribution of the calculated errors $\sqrt{\langle \varepsilon_X^2 \rangle}$, $\sqrt{\langle \varepsilon_Y^2 \rangle}$ and $\sqrt{\langle \varepsilon_Z^2 \rangle}$ (left) and histograms of errors $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$ (right). A significant part of the errors were found to be negative (about 7% for MERRA-Land, 13% for SMOS and 9% for GL-SWI).

4.2.3 General discussion

A general overview of the spatial distribution of errors (figures 5 and 6) leads to identify easily areas where there are either significant or weak root zone soil moisture errors. Large errors mainly occur in northern regions, Sahara, highly vegetated areas and mountainous regions. Low errors mainly occur in Sahel, Australia, India, Western Europe and Eastern US. On the other hand, it can be observed locations where some products exhibit strong errors when other products display low errors.

In order to understand this behavior, a selection of six 150x150 km² locations and the corresponding temporal evolution of the four products is presented in figure 7. Focusing on regions where the TC method provides low errors, it can be observed the temporal evolution of the four products in the **Niger** area (figure 7e). On this area, there is a strong agreement

between the four products which leads to low errors using the TC method. Low errors can also be observed in Western Europe and particularly in **Spain** except for the GL-SWI product. Figure 7a gives an explanation of this result showing that the GL-SWI product differs from the other three products mainly during the summer period. This is a well know issue in the ASCAT product consistent with Wagner et al., 2013. **Brazil** and **Australia** regions exhibit globally low errors for the four products but the errors are slightly more significant for the SMOS product as can be understand in figures 7d and 7f. On these areas, there is a strong correlation between the four products but the SMOS product shows a larger variability in the two selected 150x150 km² areas.

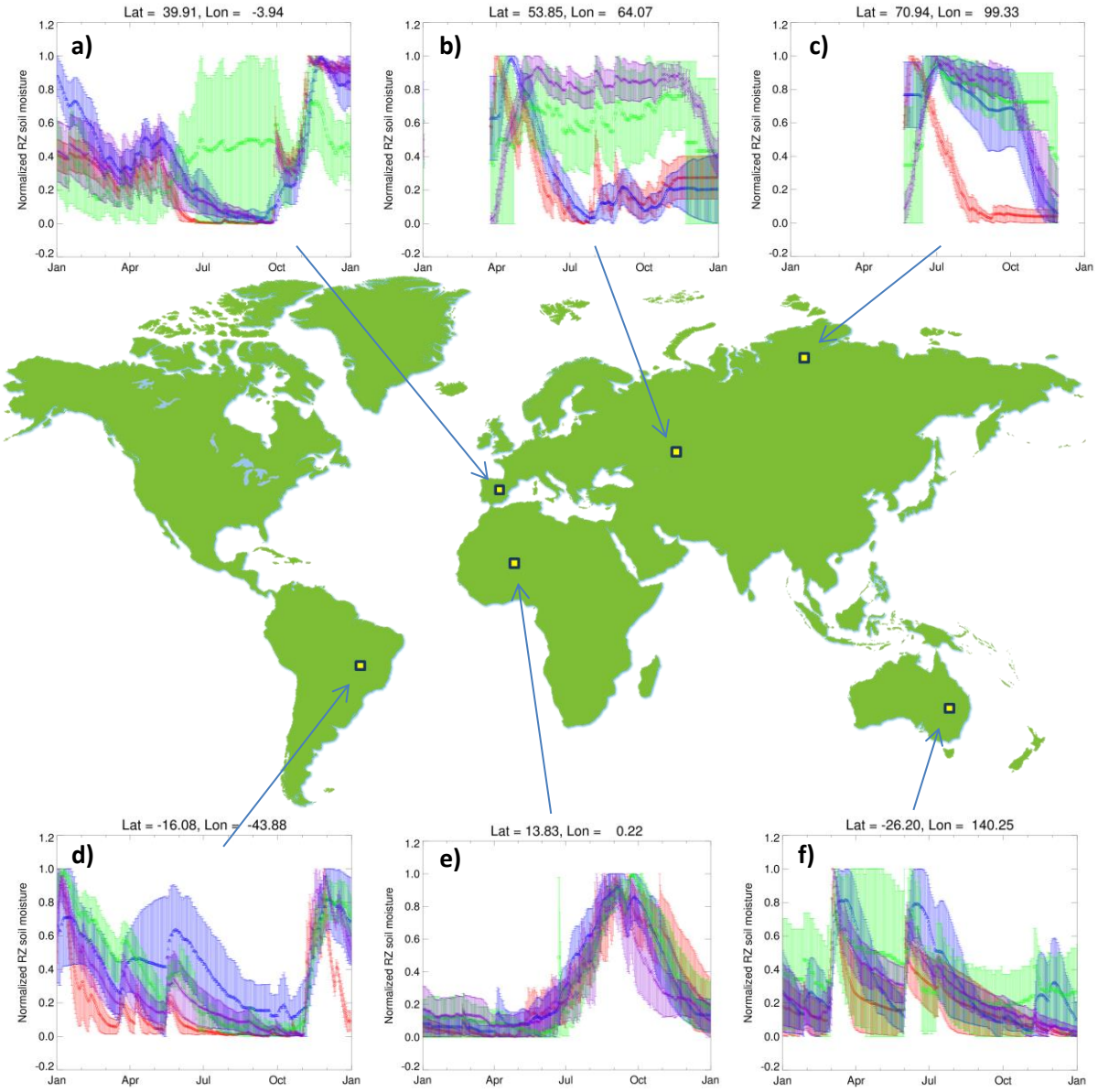


Figure 7: Examples of temporal evolution of the four root-zone soil moisture products over selected areas in Brazil, Niger, Spain, Kazakhstan, Russia and Australia. Error bars represent the spatial variability within the selected area (about 150x150 km²). Red: MERRA-Land, Blue:SMOS, Green:GLSWI, Purple: H14.

Looking at figures 5 and 6, the most significant errors can be observed in Northern regions with the MERRA-Land product. Figure 7c shows the temporal evolution of the four products in **Siberia** and shows that the MERRA-Land product strongly differs from the three other products in this area. As we do not have ground soil moisture in this region, it is not possible to identify which product is right. In the TC method, when two time-series are strongly correlated whereas the third one is uncorrelated with the two others, the triple collocation method will automatically allocate smallest errors to the two correlated time-series. This point is discussed in the Annex A section. Finally, an unexpected result was found in central Asia (**Kazakhstan**) where a strong error is obtained in the H14 product. Figure 7b illustrates that the H14 root-zone soil moisture remains at a high value during the summer period whereas the three other products (particularly SMOS and MERRA-Land) exhibit dry value during the same period. This difference was found to be related to a cold patch of soil temperature values during the winter in H14 which leads to freeze the soil up to 1 m depth during March. Here again, the TC method points out a problem with one particular product but without ground-based measurement, it is hazardous to conclude on the quality of a given product.

4.2.4 Comparison with ground-based measurements

In this last section, we examine differences between observed errors (RMS errors between the four products and the ground root-zone soil moisture measurements) and errors provided by the TC method $\sqrt{\langle \varepsilon_X^2 \rangle}$, $\sqrt{\langle \varepsilon_Y^2 \rangle}$ and $\sqrt{\langle \varepsilon_Z^2 \rangle}$. Here, we make the assumption that in-situ measurements obtained at the local scale, can be compared to 25x25 km² root-zone soil moisture products. The comparison is available over the 92 ground stations located in the US and in West Africa presented in Table I. Figure 8 shows 4 examples of the temporal evolution of rescaled root-zone soil moisture products as well as in-situ measurements. Also shown in figure 8 is the corresponding RMS errors and TC error estimates.

The first example (in Washington, US), RMS errors are relatively small (ranging between 0.09 for H14 to 0.159 for GL-SWI. The result of the TC method leads to obtains errors of nearly the same order of magnitude ranging from 0.05 (MERRA-Land, TC#1) to 0.207 (GL-SWI, TC#2). However, it can be observed that the ranking provided by the TC method is not the same as that of the ground comparison. For instance, H14 was found to be the best product compared to ground measurements whereas the TC#1 method indicates that it is MERRA-Land and the TC#2 indicates that the best product is SMOS.

The second example is located in Mississippi, US. Measured errors are larger than in the previous example ranging from 0.220 to 0.405 and the TC errors are in a similar range (0.242 to 0.451) except for the MERRA-Land product (TC#1) for which error was found to be 0.037. This behavior is difficult to explain especially as the MERRA-Land error using TC#2 was equal to 0.290. One possible explanation is the strong correlation between MERRA and H14 at this location which is not observed between MERRA and GL-SWI. Note also the negative value of $\langle \varepsilon_Z^2 \rangle$ leading to a NaN value of $\sqrt{\langle \varepsilon_Z^2 \rangle}$ in the TC#2.

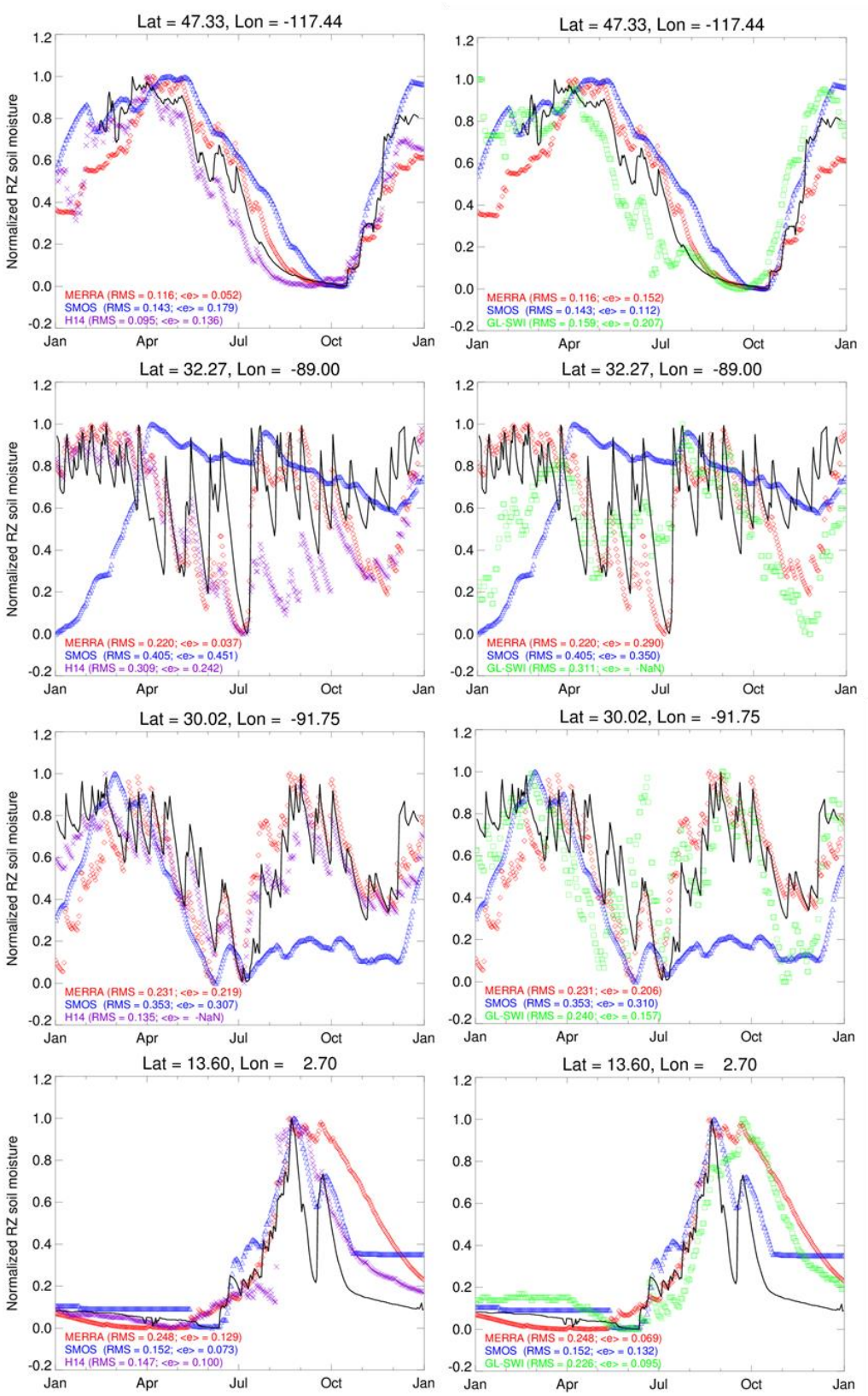


Figure 8: Four examples of temporal evolution of the normalized root-zone soil moisture products compared to in-situ soil moisture measurements. Left graphs show results obtained with the MERRA-Land / SMOS / H14 triplet (TC#1); right graphs show results obtained with the MERRA-Land / SMOS / GL-SWI triplet (TC#2). RMS errors obtained against in-situ data and TC method are indicated on each panel.

In the third example (Louisiana, US), it can be observed a very good agreement between RMS scores based on in-situ measurements and the TC method. In addition, the two triplets (TC#1 and TC#2) provide similar results. Here again, it can be noted a negative (NaN) value. An attempt of explanation for negative value occurrence is proposed in Annex A.

The fourth example is located in Niger. Observed errors are ranging between 0.147 (H14) and 0.248 (MERRA-Land) whereas the results of TC method provides errors ranging between 0.073 (MERRA-Land and SMOS) to 0.132. This underestimation of the TC error estimates can be observed in most of the pixels.

Figure 9 presents results obtained for the whole 92 pixels of the USCRN and AMMA networks. Graphics are scatterplots displaying the measured errors (based on in-situ measurements) versus the estimated errors (based on the TC method). It can be observed an overall good accuracy of the error estimated with the TC method, especially for the two satellite-based root-zone soil moisture products SMOS and GL-SWI with correlation coefficients equal to 0.72 and 0.70 respectively. Lower correlation coefficients were obtained between errors estimated (i) based on the TC method and (ii) based on in-situ measurements for the H14 and MERRA-Land error estimated (R^2=0.53 and R^2=0.17 respectively).

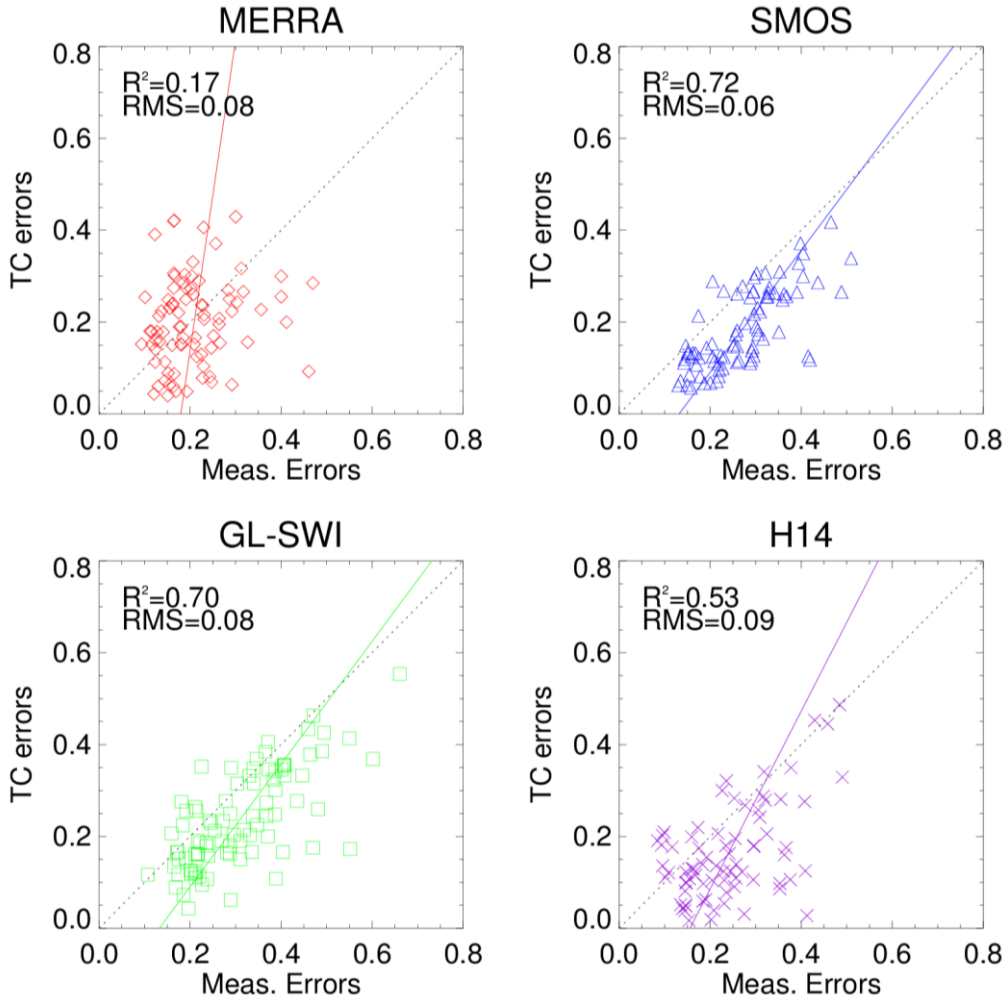


Figure 9: Comparison between observed errors (RMS based on the 92 ground stations) and errors estimated with the triple collocation (TC) method.

5) Conclusions

We have used a triple collocation method to estimate the RMS errors of four root-zone soil moisture products and we have compared results to RMS errors based on ground-based soil moisture measurements from US and West Africa networks. Results from the application of the triple collocation method show a good agreement with the ground-based RMS errors particularly for the two satellite-based root-zone soil moisture products SMOS and GL-SWI with correlation equal to 0.72 and 0.70 respectively. The TC method is less successful to estimate H14 and MERRA-Land RMS errors with correlation coefficients equal to 0.53 and 0.17 respectively.

The spatial distribution of estimated RMS errors at the global scale was investigated and has led to highlight some anomalies between the four different products. One particular anomaly was observed in central Asia with the H14 product and found to be related to a cold patch in this region.

One significant issue was related to a non-negligible amount of negative error values (up to 13% of continental surfaces) provided by the TC method. A synthetic experiment was implemented in order to understand this behavior and is presented in Annex A. It was found that negative error values mainly occur when two time series are correlated whereas the last one is slightly different of the two others. In those particular cases, the TC method leads to provide small errors for the two correlated times-series associated with significant (at least non-negligible) cross-term correlation. It follows a high probability of negative RMS errors on one of the two well-correlated time-series.

Despite this last problem, this study leads to provide global RMS errors maps of four root-zone soil moisture for the first time. Globally, the MERRA-Land product was found to have accurate estimates on 37-43% of the continental surface, SMOS obtain accurate results in about 38-44 %, GL-SWI on about 36% and H14 on about 43% of the continental surface.

Finally, it is recommended to investigate further, from the H14 production chain, the sources of H14 errors observed in the central Asia region. However, apart from this particular area, the global scale RMS maps confirm the excellent performances of the H-SAF root zone soil product for all ranges of climate and surface conditions.

ANNEX A

Explanation of the negative values of $\langle \varepsilon_X^2 \rangle$ $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$.

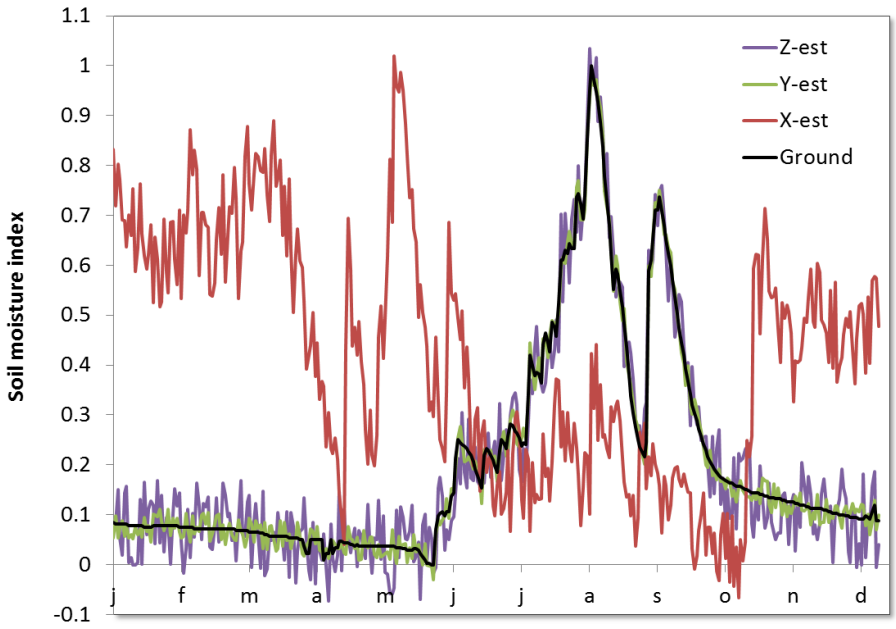
Let's consider the configuration presented in figure A1. In this example, one of the three root-zone soil moisture estimates (X) is far from the truth whereas the two other time-series (Y and Z) are close to the truth. As the truth is known, the error terms can be computed and one can obviously observe that the $\langle \varepsilon_X^2 \rangle$ error is much more important than $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$. For this example, the computation of the triple collocation equations leads to:

$$\langle (X-Y)(X-Z) \rangle = 0.207$$

$$\langle (Y-X)(Y-Z) \rangle = 4.83^{\text{E-4}}$$

$$\langle (X-Z)(Y-Z) \rangle = 3.02^{\text{E-3}}$$

Here again, the triple collocation method provides an accurate estimation of the errors $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$. However, it can be noted that the cross terms $\langle \varepsilon_X . \varepsilon_Y \rangle$, $\langle \varepsilon_X . \varepsilon_Z \rangle$ and $\langle \varepsilon_Y . \varepsilon_Z \rangle$ are now of the same order of magnitude than the square errors $\langle \varepsilon_X^2 \rangle$, $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$ which should be incompatible with the assumptions of the triple collocation method.



$\langle \varepsilon_X^2 \rangle = 0.205$

$\langle \varepsilon_Y^2 \rangle = 3.58^{\text{E-4}}$

$\langle \varepsilon_Z^2 \rangle = 3.31^{\text{E-3}}$

$\langle \varepsilon_X . \varepsilon_Y \rangle = -2.80^{\text{E-4}}$

$\langle \varepsilon_X . \varepsilon_Z \rangle = -2.41^{\text{E-3}}$

$\langle \varepsilon_Y . \varepsilon_Z \rangle = 4.91^{\text{E-5}}$

Figure A1: Example of Root-zone soil moisture time-series. The reference soil moisture is presented in black and the 3 synthetic estimates are presented in red (X), green (Y) and purple (Z).

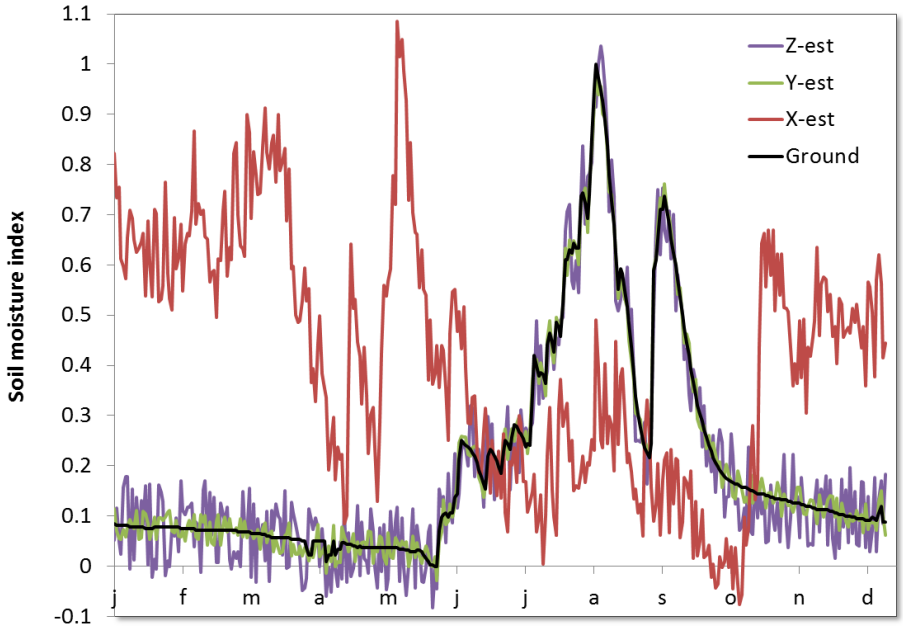
Another instructive experiment is obtained with a new random noise (of the same level) applied on the three time series (X, Y and Z). Figure A2 slightly differs from figure A1 and the computation of error (right table in figure A2) provides similar values. However, the result of the triple collocation leads to the following results:

$$\langle (X-Y)(X-Z) \rangle = 0.205$$

$$\langle (Y-X)(Y-Z) \rangle = -2.24^{E-4}$$

$$\langle (X-Z)(Y-Z) \rangle = 4.07^{E-3}$$

In this case, a negative estimation of the $\langle \varepsilon_Y^2 \rangle$ error is obtained using the TC methodology (second equation). As state in the report, negative values were obtained in the range of 6 to 13% of the continental surfaces.



$\langle \varepsilon_X^2 \rangle = 0.204$
$\langle \varepsilon_Y^2 \rangle = 4.03^{E-4}$
$\langle \varepsilon_Z^2 \rangle = 3.39^{E-3}$
$\langle \varepsilon_X . \varepsilon_Y \rangle = 2.47^{E-4}$
$\langle \varepsilon_X . \varepsilon_Z \rangle = -4.17^{E-4}$
$\langle \varepsilon_Y . \varepsilon_Z \rangle = -2.49^{E-5}$

Figure A2: same as figure A1 but with another random noise (of the same level) used to obtain the 3 synthetic root-zone soil moisture time-series. In this example, a negative value is obtained for the TC estimation of $\langle (Y-X)(Y-Z) \rangle = -2.24^{E-4} \approx \langle \varepsilon_Y^2 \rangle$.

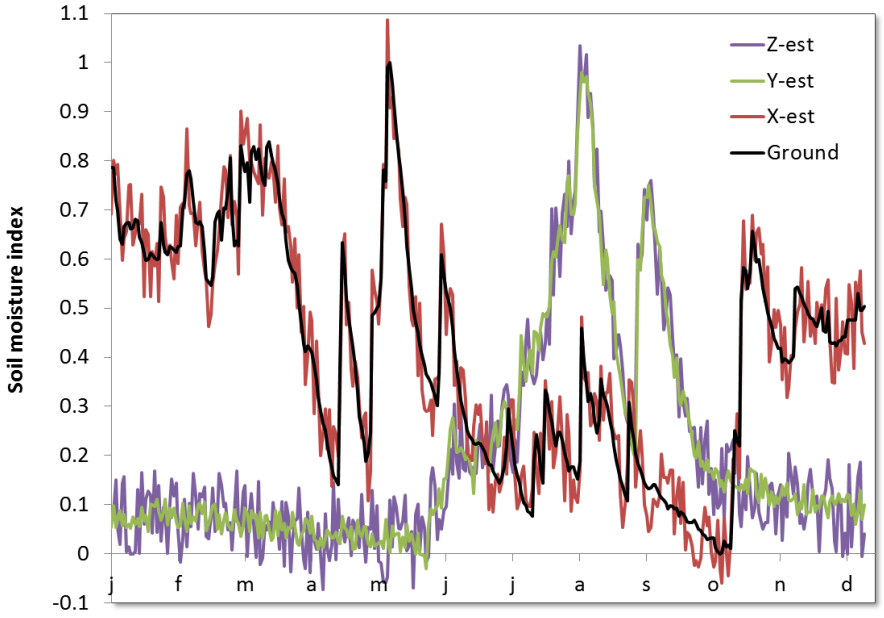
Finally, let's assume that the true soil moisture is close to the first estimates *X* as presented in figure A3. In this configuration, the mean error $\langle \varepsilon_X^2 \rangle$ is small compared to $\langle \varepsilon_Y^2 \rangle$ and $\langle \varepsilon_Z^2 \rangle$ as shown in Table right to the figure A3. In the same time, one can be noted that the cross term $\langle \varepsilon_Y . \varepsilon_Z \rangle$ is large and incompatible with TC assumptions. It follows that, as the three time-series (X), (Y) and (Z) are exactly the same as in figure A1, the computation of the triple collocation equations leads to:

$$\langle (X-Y)(X-Z) \rangle = 0.207$$

$$\langle (Y-X)(Y-Z) \rangle = 4.83^{E-4}$$

$$\langle (X-Z)(Y-Z) \rangle = 3.02^{E-3}$$

In this example, the calculation of $\langle (X-Y)(X-Z) \rangle$ is far from $\langle \varepsilon_X^2 \rangle$ due to non-negligible $\langle \varepsilon_Y . \varepsilon_Z \rangle$ term equal to 0.202. The triple collocation method leads to identify smallest error on both (*Y*) and (*Z*) time-series (-1.87^{E-4} and 3.69^{E-3} respectively), and a greater error to the (*X*) time-series (0.207). It follows that when two time-series are strongly correlated whereas the third one is uncorrelated with the two others, one of the three cross term becomes significant and the triple collocation method will automatically allocate smallest errors to the two correlated time-series. In certain cases, one of the two small errors can also be negative.



$\langle \varepsilon_X^2 \rangle$ = 3.18^{E-3}
$\langle \varepsilon_Y^2 \rangle$ = 0.201
$\langle \varepsilon_Z^2 \rangle$ = 0.207
$\langle \varepsilon_X . \varepsilon_Y \rangle$ = -8.90^{E-4}
$\langle \varepsilon_X . \varepsilon_Z \rangle$ = -1.14^{E-3}
$\langle \varepsilon_Y . \varepsilon_Z \rangle$ = 0.202

Figure A3: Exactly the same as figure A1 except that the truth is supposed to be close to the (X) time-serie.

ANNEX B, Correlation and RMS errors between the 4 different products

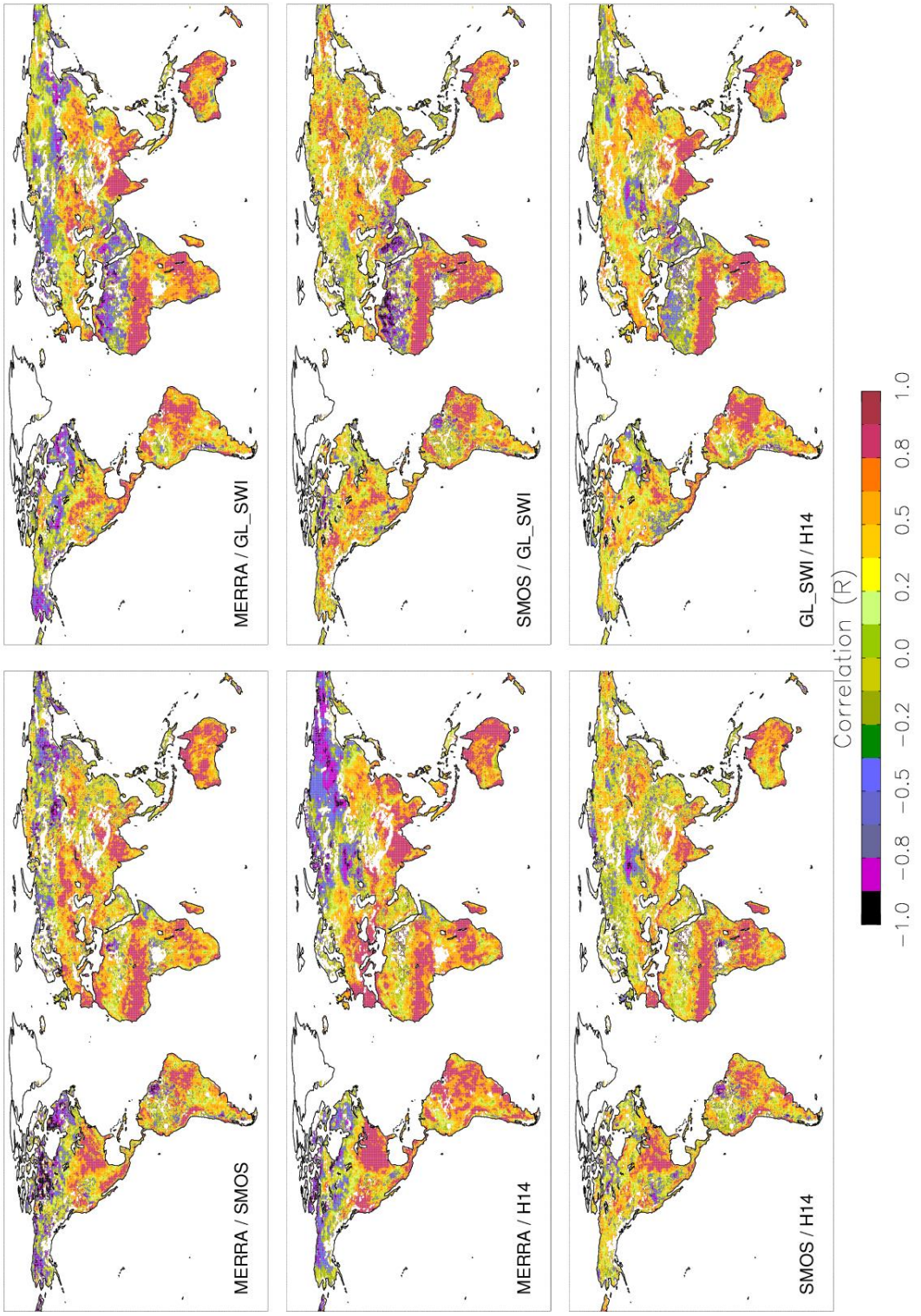


Figure B1: Cross correlation between the 4 different products

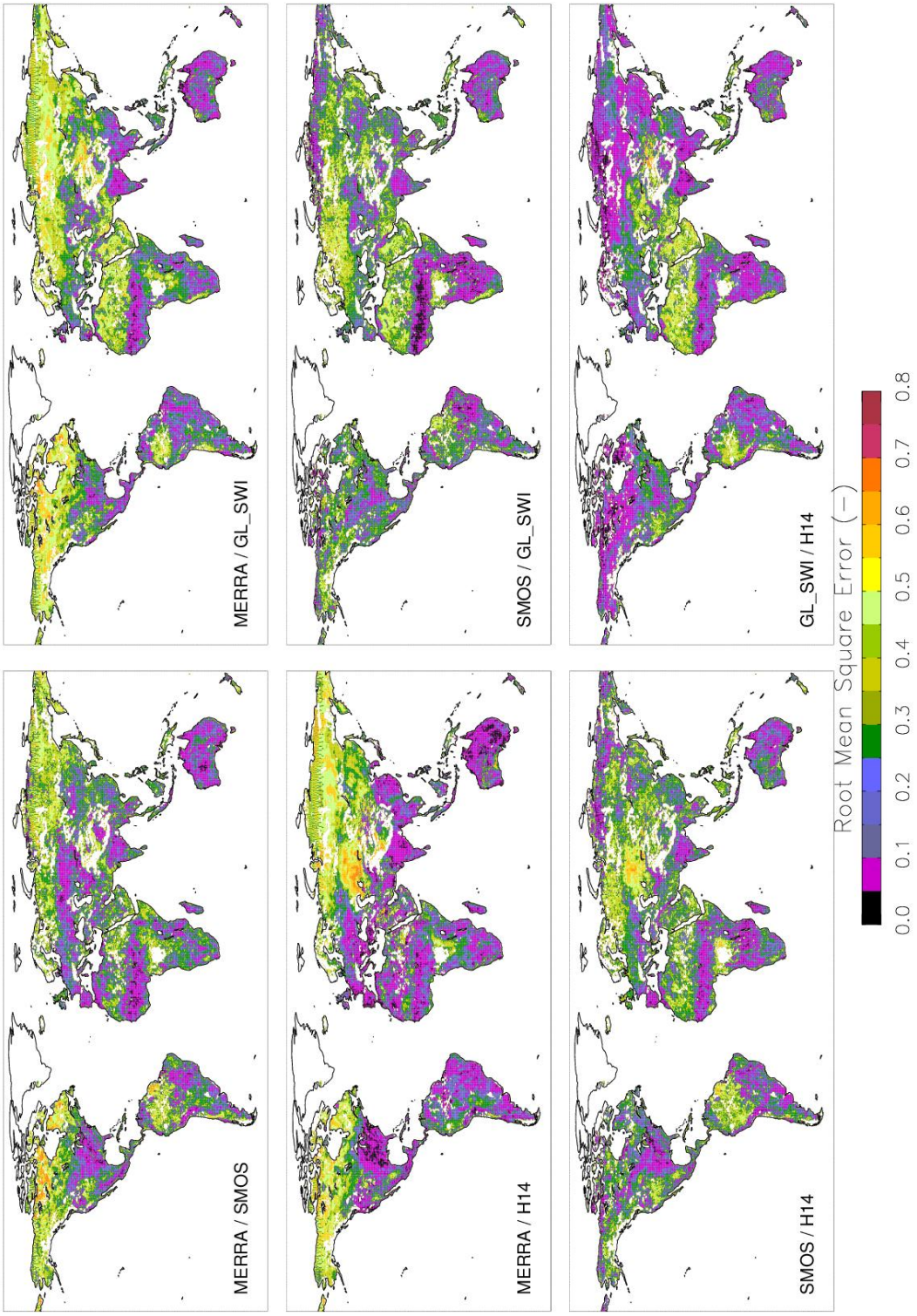


Figure B2: Root Mean Square Errors (RMSE) between the 4 different products

References

Al Bitar A. , Y. H. Kerr ,O. Merlin , F. Cabot, J.P. Wigneron, Global Drought Index From SMOS Soil Moisture, IGARSS, 22-26 July, Melbourne, 2013.

Albergel, C., de Rosnay, P., Gruhier, C., Munoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., Wagner, W., 2012: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, Remote Sensing of Environment, Vol. 118, 215-226

Albergel, C., Rudiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., Martin, E., 2008 : From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, Hydrology and Earth System Sciences, Vol. 12 (6), 1323-1337

Albergel, C., W. Dorigo, R. H. Reichle, G. Balsamo, P. de Rosnay, J. Munoz-Sabater, L. Isaksen, R. de Jeu, and W. Wagner: Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing, Journal of Hydrometeorology, 14, 1259–1277. doi:10.1175/JHM-D-12-0161.1, 2013.

Balsamo, G., Viterbo, P., Beljaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., Scipal, K., 2009: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, Journal of Hydrometeorology, Vol. 10 (3), 623-643

Caires, S. and Sterl, A., 2003: Validation of ocean wind and wave data using triple collocation, Journal of Geophysical Research-Oceans, Vol. 108 (C3)

de Rosnay, P., M. Drusch, G. Balsamo, I. L., and C. Albergel, 2011 : Extended Kalman Filter soil moisture analysis in the IFS. ECMWF Spring Newsletter, 127,12–16.

Dorigo, W., W. Wagner, R. Hohensinn, S. Hahn, C. Paulik, M. Drusch, S. Mecklenburg, P. van Oevelen, A. Robock, and T. Jackson, 2011 : The international soil moisture network: a data hosting facility for global in situ soil moisture measurements. Hydrology and Earth System Sciences, 15,1675–1698.

Dorigo, W., de Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., Fernandez-Prieto, D., 2012: Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture, Geophysical Research Letters, vol. 39

Draper C., R. Reichle, R. de Jeu, V. Naeimi, R. Parinussa, W. Wagner, 2013: Estimating root mean square errors in remotely sensed soil moisture over continental scale domains, Remote Sensing of Environment, Vol. 137, 288-298

Drusch, M., Scipal, K., de Rosnay, P., Balsamo, G., Andersson, E., Bougeault, P., Viterbo, P., 2009: Towards a Kalman Filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System, Geophysical Research Letters, Vol. 36

Gruhier C., C. Albergel, P. de Rosnay, S. Hasenauer, B. Zeiner, 2011 : Comparison between H-SAF large scale surface soil moisture, H-SAF assimilated soil moisture and SMOS level 2 soil moisture, H-SAF report, 47p

Hain, C. R., Crow, W. T., Mecikalski, J. R., Anderson, M. C., Holmes, T., 2011: An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling, Journal of Geophysical Research-Atmospheres, Vol. 116

Janssen, Paem, Abdalla, S., Hersbach, H., Bidlot, J. R., 2007: Error estimation of buoy, satellite, and model wave height data, Journal of Atmospheric and Oceanic Technology, Vol. 24 (9), 1665-1677

Leroux, D. J., Kerr, Y. H., Richaume, P., Fieuzal, R., 2013 : Spatial distribution and possible sources of SMOS errors at the global scale, Remote Sensing of Environment, Vol. 133, 240-250

Loew, A., Schlenz, F., 2011: A dynamic approach for evaluating coarse scale satellite soil moisture products, Hydrology and Earth System Sciences, Vol. 15(1), 75-90

Miralles, D. G., Crow, W. T., Cosh, M. H., 2010: Estimating Spatial Sampling Errors in Coarse-Scale Soil Moisture Estimates Derived from Point-Scale Observations, Journal of Hydrometeorology, Vol. 11 (6), 1423-1429

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., Toure, A., 2011 : Assessment and Enhancement of MERRA Land Surface Hydrology Estimates, Journal of Climate, Vol.24 (24), 6322-6338

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J. Y., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., Woollen, J., 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, Vol. 24 (14), 3624-3648

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, Vol. 35 (24)

Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, Journal of Geophysical Research-Oceans, Vol. 103 (C4), 7755-7766

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research-Atmospheres, Vol. 106 (D7), 7183-7192

van den Hurk, B. and Viterbo, P., 2003: The Torne-Kalix PILPS 2(e) experiment as a test bed for modifications to the ECMWF land surface scheme, Global and Planetary Change, Vol. 38 (1-2), 165-173

Wagner, W., Lemoine, G., Rott, H., 1999 : A method for estimating soil moisture from ERS scatterometer and soil data, Remote Sensing of Environment, Vol. 70 (2), 191-207

The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications

Wagner, W., Hahn, S, Kidd, R, Melzer, T and co-autors, Meteorologishche Zeitschrift Volume: 22 Issue: 1 Pages: 5-33

Yilmaz, M. T., Crow, W. T., Anderson, M. C., Hain, C., 2012: An objective methodology for merging satellite- and model-based soil moisture products, Water Resources Research, Vol. 48