TRIPLE COLLOCATION – A NEW TOOL TO DETERMINE THE ERROR STRUCTURE OF GLOBAL SOIL MOISTURE PRODUCTS

Klaus Scipal[1], Wouter Dorigo[2], Richard deJeu[3]

1. European Space Agency, Nordwijk, the Netherlands
2. Vienna University of Technology, Vienna, Austria
3. Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

ABSTRACT

Recently Triple Collocation (TC) was adopted for soil moisture application. Results from a first application indicated that the method could be useful to estimate global error patterns. Here we test the method with new data sets. The results show that the method is robust and that it allows to derive objective error estimates.

Index Terms— soil moisture, global, validation

1. INTRODUCTION

In recent years, an increasing number of global soil moisture data sets have become available from passive and active coarse resolution satellite microwave sensors. Altogether they span a period of more than 30 years which facilitates studying long-term soil moisture behavior e.g. in relation to climate change. The use of multi-mission data also involves many scientific challenges, such as quantifying the temporal and spatial error fields for each satellite product and understanding the complex effects of the various error sources (sensor calibration, retrieval errors, model parameterization, etc.) on the observed variations.

Traditional error characterization methods rely on validation of the satellite retrievals with in situ observations. These methods suffer of three shortcomings: (1) Ground observations are restricted to a few locations worldwide and often cover only limited observation periods; and (2) Reliable error estimation is complicated by representativeness and scaling errors; and (3) These methods don't allow estimating the error of each data set but are limited to estimating difference between them.

Another tool, error propagation, is valuable to establish the uncertainty of model estimates resulting from errors in the input variables. [1] implemented a Monte Carlo based error propagation to estimate the uncertainty in soil moisture retrievals obtained from scatterometers using the TU Wien approach. [2] found an analytical solution for estimating the uncertainties of soil moisture estimates from radiometers based on the LPRM model. However, the uncertainties obtained by error propagation methods only account for random errors in the model input variables but do not tell if the model itself is correct. Therefore the uncertainties obtained for different models are difficult to compare quantitatively.

Recently, [3] introduced the triple collocation method in the field of satellite based soil moisture research. The triple collocation method allows a simultaneous estimation of the error structure and the cross-calibration of a set of at least three linearly related datasets with uncorrelated errors

2. TRIPLE COLLOCATION ERROR MODEL

2.1. Methodology

We assume we have three data sets, Θ_1, Θ_2 and Θ_3, each containing N observations of the same geophysical process. The observations can be derived from models, *in-situ* or satellite data. They are spatially and temporarily collocated and have mutually independent error structures and no systematic biases. Each observation i in these data sets differs from the hypothetical truth Θ by a residual r_i (Eq. 1, for reasons of simplicity we omitted index i).

$$\begin{aligned}\Theta_1 &= \Theta + r_1\\ \Theta_2 &= \Theta + r_2\\ \Theta_3 &= \Theta + r_3\end{aligned} \qquad (1)$$

To quantify the quality of each set of observations we need to estimate the variance σ_1^2, σ_2^2, σ_3^2, of the residuals. For this purpose we eliminate the hypothetical truth from Eq. 1 by pair wise subtraction (Eq. 2) and multiplication (Eq. 3):

$$\begin{aligned}\Theta_1 - \Theta_2 &= r_1 - r_2\\ \Theta_1 - \Theta_3 &= r_1 - r_3\\ \Theta_2 - \Theta_3 &= r_2 - r_3\end{aligned} \qquad (2)$$

$$
\begin{aligned}
(\Theta_1-\Theta_2)(\Theta_1-\Theta_3) &= r_1r_1 - r_1r_2 - r_1r_3 + r_2r_3 \\
(\Theta_1-\Theta_2)(\Theta_2-\Theta_3) &= r_1r_2 - r_2r_2 - r_1r_3 + r_2r_3 \qquad (3)\\
(\Theta_1-\Theta_3)(\Theta_2-\Theta_3) &= r_1r_2 - r_3r_2 - r_1r_3 + r_3r_3
\end{aligned}
$$

When a sufficiently large number of samples is available, we can take the average over each of the lines in Eq. 3 (indicated by angle brackets in the following). Assuming that the residuals are uncorrelated between the datasets, the covariance terms (the mean of the mixed residual products) on the right hand side of Eq.3 become 0 and we get an expression for the variance $\sigma_1^2 = \langle r_1r_1 \rangle$, $\sigma_1^2 = \langle r_2r_2 \rangle$ and $\sigma_1^2 = \langle r_3r_3 \rangle$ in which all variables are known (Eq. 4).

$$
\begin{aligned}
\sigma_1^2 &= \langle (\Theta_1-\Theta_2)(\Theta_1-\Theta_3) \rangle \\
\sigma_2^2 &= \langle (\Theta_1-\Theta_2)(\Theta_2-\Theta_3) \rangle \qquad (4)\\
\sigma_3^2 &= \langle (\Theta_1-\Theta_3)(\Theta_2-\Theta_3) \rangle
\end{aligned}
$$

2.2. Bias Removal

It is important to note that any bias in the datasets result in a biased estimate of σ_1^2, σ_2^2 and σ_3^2 as the variance term in Eq. 3 is artificially inflated and the covariance terms may not necessarily become 0. For this reason any bias has to be removed prior to applying the TC model. In [3] an iterative linear regression considering errors in both variables approach was used to remove the systematic differences between the datasets, which is a strict solution to the problem. Here we adopt an approximation approach similar to [4]. In this approach TC is applied to anomaly data which is scaled according to [5] to provide the same mean and range.

2.3. Requirement on number of observations

To get a robust estimate of the variance $\sigma_1^2 = \langle r_1r_1 \rangle$, $\sigma_1^2 = \langle r_2r_2 \rangle$ and $\sigma_1^2 = \langle r_3r_3 \rangle$ a large number of samples is required. In practice this data is often not available. We analyzed the impact of having only a limited number of observations on the estimation of the variance. To this end we simulated three sets of observations by generating a long time series and adding random noise with a variance of 10, 20 and 30%. We then applied TC to subsets of the data, where each subset was derived by thinning the original series. The simulation indicated that at least 100 observation triplets are required for a reliable estimation of the variance. Below 100 observation limited sample size leads to systematic effects of up to 5%.

3. DATA SETS

3.1 Scatterometer data

The Advanced Scatterometer (ASCAT) on MetOp-A operates in C-band (5.6 GHz) at 25 km in VV polarization and is operational since January 2007. The backscatter measurements are converted to soil moisture estimates by applying the TU Wien model which has originally been designed for the ERS-1/2 scatterometers [1]. The TU Wien model exploits the unique sensor design and the advantages of a change detection method. To correct for the effects of plant growth and decay the model uses the vegetation sensitive signature of the multi-incidence angle observations. A soil moisture index is then retrieved relating each observation to a dry and wet backscatter reference, which results in a relative measure of surface (< 2 cm) soil moisture ranging between 0 and 1 (or 0 and 100%).

3.2 Radiometer data

Since June 2002, the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) aboard the Aqua satellite provides a nearly daily global coverage. The earth radiance is measured at six frequencies. The two frequencies considered in this study are the C-band operated at 6.9 GHz and the X-band operated at 10.7 GHz. Spatial resolutions are 73x43 km and 51x30 km for the C- and X-band, respectively. In this study we only use AMSR-E night-time observations, as it was shown that these are better suited for retrieving soil moisture than day-time observations [6]. The brightness temperatures measured by AMSR-E and SSM/I are converted to surface soil moisture applying the Land Parameter Retrieval Model (LPRM) [7]. LPRM is based on the solution of a microwave radiative transfer model and solves simultaneously for surface soil moisture, vegetation optical depth and land surface temperature without a-priori information of land surface characteristics.

3.3 Model data

3.3.1 ERA Interim

ERA Interim is a numerical weather prediction reanalysis data set containing consistent atmosphere and surface analyses for the period from 1989 to present based on the ECMWF Integrated Forecast System (IFS) model. The system runs at ~80 km horizontal resolution. In the IFS, land surface processes are described by the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL)

3.3.2 GLDAS-NOAH

The Noah model from the Global Land Data Assimilation System (GLDAS) provides soil moisture and other atmospheric and land surface variables at a 3-hour time interval for a regular global grid with a spatial resolution of 0.25°. The model is forced by a combination of NOAA/GDAS atmospheric analysis fields, spatially and temporally disaggregated NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP) fields, and observation based downward shortwave and longwave radiation fields derived using the method of the Air Force Weather Agency's AGRicultural METeorological system.

4. RESULTS

4.1 Error Estimation

Figure 1 shows the triple collocation errors for a combination of ASCAT, AMSR-E C-band, and ERA-Interim soil moisture estimates. The errors σ (i.e. the square-root of the values obtained from Eq. 4) are expressed in the climatology of the ERA-Interim re-analysis data set. The results of the error estimation suggest that all three data sets are characterized by a relatively low error. The mean global error is 0.017 m^3m^{-3} for the ASCAT, 0.019 m^3m^{-3} for the AMSR-E C-band observations and 0.018 m^3m^{-3} for ERA-Interim. The low error can be explained by the low dynamic range of the ERA-Interim soil moisture which has been used as reference. Generally, error estimates are lowest in arid regions such as Southern Africa, mainland Australia or Central Asia. This is explained by the very low amounts of precipitation received and hence the very low variability of soil moisture. In very dry areas (e.g. those of central Australia) errors of soil moisture derived from AMSR-E C–band are remarkably lower than soil moisture estimates derived from ASCAT and, to a smaller degree, than the modeled soil moisture of ERA-Interim. The relatively high errors obtained for scatterometer data in these areas are a well-known phenomenon and are partly caused by volume scattering effects in dry, loose sand and by systematic roughness effects. On the other hand, soil moisture derived from AMSR-E is prone to larger random errors in moderately to densely vegetated areas, like for instance found in south-eastern North America and northern Argentina. Figure 1 also shows the areas for which either ASCAT (shown in blue) or AMSR-E (red) gives the lowest errors which indicate that the active sensor provides more reliable observations in the presence of moderate to dense vegetation whereas the passive sensor provides more reliable observations in dry climates.

4.2 Influence of Frequency

Figure 2 illustrates the influence of increasing observation frequency on the error structures obtained for radiometer observations. TC was applied on a combination of ERA-Interim/ASCAT/AMSR-E X-band which can be compared to the combination ERA-Interim/ASCAT/AMSR-E C-band shown in Figure 1. On average, there is a clear increase in errors with increasing frequency, especially in areas characterized by moderate to dense vegetation cover, like in southeast Siberia.

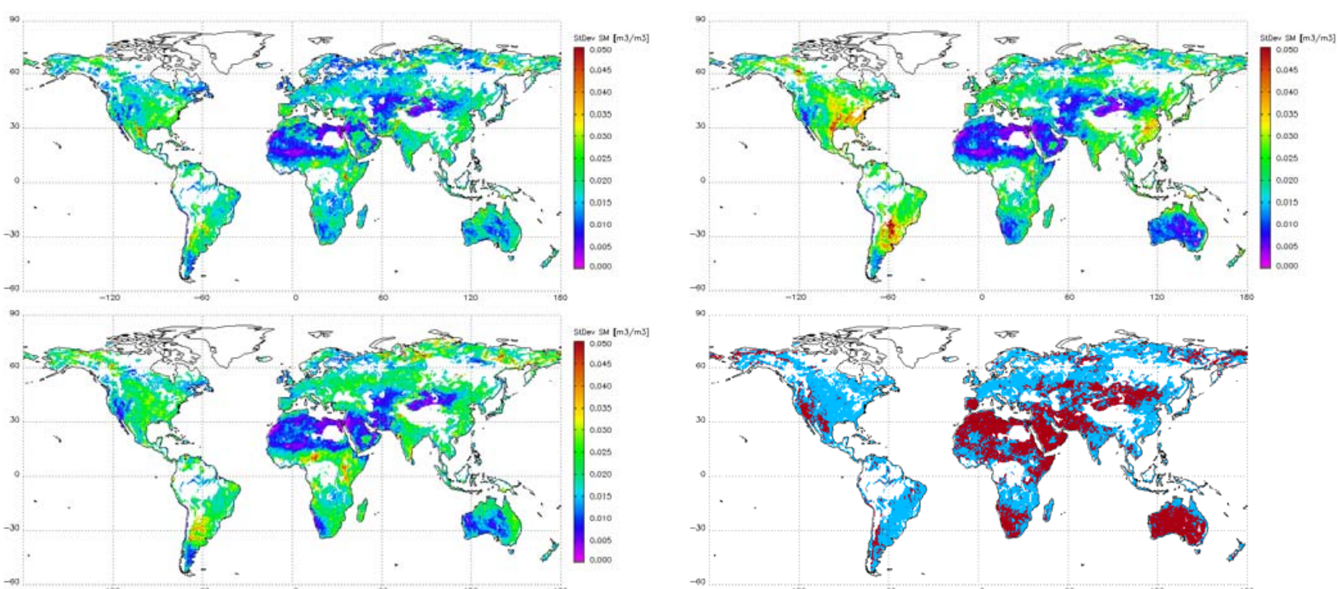

Figure1. Spatial errors σ of (top left) ASCAT, (top right) AMSR-E C-band, and (bottom left) ERA-Interim surface soil moisture estimates. Errors are expressed in the climatology of ERA-Interim. (bottom right) shows the areas in which either ASCAT (blue) or AMSR-E (red) shows the smallest error value. White areas indicate areas for which less than 100 common observations are available.

This behavior can be explained by the fact that for the corresponding decrease in wavelength the soil moisture signal emitted from the surface is increasingly absorbed by the vegetation canopy. The trends observed for the different observation bands correspond well to the trends in frequency-related uncertainties of LPRM products obtained by error propagation [2].

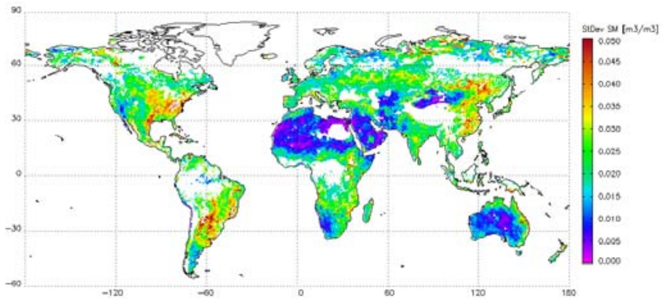

Figure 2. Spatial errors σ of AMSR-E X-band data.

4.3 Influence of different combinations

To investigate the robustness of the method, we applied TC using ERA Interim and GLDAS-NOAH data in different combinations (i.e. AMSR-E/ASCAT/ERA-Interim and AMSR-E/ASCAT/GLDAS-NOAH). Changing one of the data sets should have no impact on the estimated variance of the other two data sets. The impact on the estimated error when using different input datasets is negligible (Figure 3). Larger differences (red and blue colors in Figure 3) are found in regions with a low number of collocated observations and are of numerical rather then systematic nature.

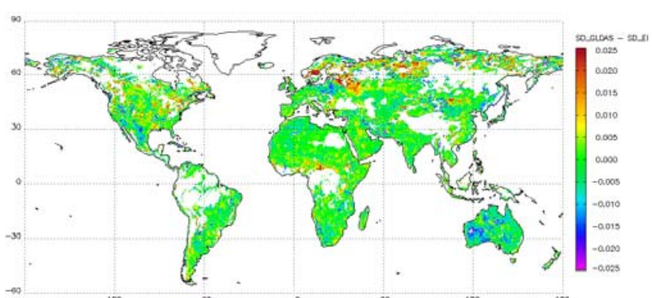

Figure 3. Difference in error σ of AMSR-E C-band data derived from the combination AMSR-E/ASCAT/ERA-Interim and AMSR-E/ASCAT/GLDAS-NOAH.

5. CONCLUSIONS

In [3] TC, a method to objectively derive errors of global observed and modelled soil moisture data sets was introduced. We tested the method using different data sets and applying it on a global scale. The results indicate that TC is a robust error estimation tool. The derived spatial error patterns are consistent with known performance issues of the different sensors. The impact of different wavelength on the performance of the retrieval is also well captured. The error estimated for each data set is insensitive towards the selection of the other data sets used in TC.

In the next steps we will quantify the impact of biases on the error estimation by comparing results using different bias correction schemes.

5. REFERENCES

[1] V. Naeimi, K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner, "An Improved Soil Moisture Retrieval Algorithm for ERS and METOP Scatterometer Observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, 2009, pp. 1999-2013.

[2] R. Parinussa, A.G. Meesters, Y.Y. Liu, W. Dorigo, W. Wagner, and R.A. De Jeu, "An analytical solution to estimate the error structure of a global soiol moisture data set," *IEEE Transactions on Geoscience and Remote Sensing, submitted.*, 2010.

[3] R.A. de Jeu, K. Scipal, T.R. Holmes, V. Naeimi, and W. Wagner, "A possible solution for the problem of estimating the error structure of global soil moisture data sets," *Geophysical Research Letters*, vol. 35, 2008, pp. 0-3.

[4] D.G. Miralles, W.T. Crow, and M.H. Cosh, "Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations," *Journal of Hydrometeorology, submitted*, 2010.

[5] K. Scipal, M. Drusch, and W. Wagner, "Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system," *Advances in Water Resources*, vol. 31, 2008, pp. 1101-1112.

[6] R.A. Jeu, W. Wagner, T.R. Holmes, A.J. Dolman, N.C. Giesen, and J. Friesen, "Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. Surveys in Geophysics, 29, 399-420.," *Surveys in Geophysics*, vol. 29, 2008.

[7] M. Owe, R.A. de Jeu, and J.P. Walker, "A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, 2001, pp. 1643-1654.