A Possible Solution for the Problem of Estimating the Error Structure of Global Soil Moisture Datasets

Scipal, K.; Holmes, T.R.H.; de Jeu, R.A.M.; Naeimi, V.; Wagner, W.W.

published in
Geophysical Research Letters
2008

DOI (link to publisher)
10.1029/2008GL035599

document version
Publisher's PDF, also known as Version of record

A possible solution for the problem of estimating the error structure of global soil moisture data sets

K. Scipal,[1] T. Holmes,[2] R. de Jeu,[2] V. Naeimi,[3] and W. Wagner[3]

Received 6 August 2008; revised 8 October 2008; accepted 22 October 2008; published 17 December 2008.

[1] In the last few years, research made significant progress towards operational soil moisture remote sensing which lead to the availability of several global data sets. For an optimal use of these data, an accurate estimation of the error structure is an important condition. To solve for the validation problem we introduce the triple collocation error estimation technique. The triple collocation technique is a powerful tool to estimate the root mean square error while simultaneously solving for systematic differences in the climatologies of a set of three independent data sources. We evaluate the method by applying it to a passive microwave (TRMM radiometer) derived, an active microwave (ERS-2 scatterometer) derived and a modeled (ERA-Interim reanalysis) soil moisture data sets. The results suggest that the method provides realistic error estimates.
Citation: Scipal, K., T. Holmes, R. de Jeu, V. Naeimi, and W. Wagner (2008), A possible solution for the problem of estimating the error structure of global soil moisture data sets, *Geophys. Res. Lett.*, *35*, L24403, doi:10.1029/2008GL035599.

1. Introduction

[2] Soil moisture is a crucial parameter for a large number of applications. Consequently, remote sensing of soil moisture has been an important research topic since the 1970s. But only in the last few years significant progress towards operational soil moisture services has been made which lead to a greater diversity of methods and, consequently, to more successful algorithms [*Wagner et al.*, 2007]. With these improved algorithms it has been possible to derive soil moisture from existing operational passive microwave satellite systems such as the Advanced Microwave Scanning Radiometer (AMSR-E) [*Njoku et al.*, 2003; *Owe et al.*, 2008], the TRMM Microwave Imager (TMI) [*Owe et al.*, 2008] and active microwave systems such as the scatterometers on-board of ERS-1/2 [*Wagner et al.*, 2003] and METOP [*Bartalis et al.*, 2007].

[3] For an optimal use of these data, an accurate estimation of the error structure is an important condition. Traditional error estimation methods, i.e. the validation with ground based observations, are however cumbersome and often unreliable. While the gravimetric technique or one of the numerous indirect approaches allows measuring soil moisture accurately in the field, the problem is that these measurements are only representative for very small areas and the error estimation problem becomes distorted by scaling errors, which can be larger than the actual retrieval error. In addition, available ground observations are restricted to a few locations worldwide and often cover only limited observation periods. Other validation approaches [e.g., *Crow*, 2007] can be applied globally but are limited to validate the relative variation in the signal as they are based on the analysis of correlations.

[4] In this study, we propose to use the triple collocation error estimation technique. This technique has been used previously in oceanography to evaluate wind and wave height observations [*Stoffelen*, 1998; *Caires and Sterl*, 2003; *Janssen et al.*, 2007]. The method allows the estimation of the root mean square error e^2 while simultaneously solving for systematic differences in each collocated data set. In previous studies the problem that satellite and model derived soil moisture data are characterized by large systematic differences in the mean and the variance has been highlighted while the inter-annual variations comply [*Entin et al.*, 1999; *Dirmeyer et al.*, 2004]. The proposed method directly accounts for such differences and therefore appears specifically useful for soil moisture applications. However, in the absence of an absolute ground truth, one of the collocated data sets has to be defined as reference, i.e. e^2 is expressed in the climatology of the reference data set. In this light, the approach is complementary to classical approaches and together they provide a complete set of validation metrics (capturing both the relative anomaly detection and absolute e^2 characteristics of a product). To test the method we use a passive microwave (TMI) derived, an active microwave (ERS-2) derived and a modeled (ERA-Interim reanalysis) soil moisture data set.

2. Data

[5] For this study we used soil moisture data from the ERS-2 scatterometer (Θ_S), the TMI radiometer (Θ_T) and the ERA-Interim re-analysis project (Θ_E) for the years 1998, 1999 and 2000. For later processing, the data are binned to daily files and collocated to a 0.25° regular grid using a nearest neighbor re-sampling. Such, for each grid point approximately 300 collocated samples are available. To avoid numerical problems we do not consider points with less than 100 collocated samples during the error estimation.

2.1. ERA-Interim

[6] The ERA-Interim reanalysis data set contains consistent atmosphere and surface analyses for the period from 1989 until real time based on the ECMWF NWP model. The reanalysis makes use of the ECMWF Integrated Forecast System at T255 spectral resolution (~80 km horizontal

[1]European Centre for Medium-Range Weather Forecasts, Reading, UK.
[2]Department of Geo-Environmental Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.
[3]Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Vienna, Austria.

0094-8276/08/2008GL035599$05.00

resolution) with 91 vertical levels. In the IFS, land surface processes are described by the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) [*Viterbo and Beljaars*, 1995]. In TESSEL soil processes are calculated in four layers. The lower boundary of each layer is at 0.07, 0.28, 1.0 and 2.68 m depth. To keep the land surface model simple, TESSEL uses a globally uniform soil type with fixed soil hydraulic parameters. Saturation is prescribed with a value of 0.472 $m^3\ m^{-3}$, field capacity with 0.323 $m^3 m^{-3}$ and the wilting point with 0.171 $m^3\ m^{-3}$.

2.2. ERS Scatterometer

[7] The scatterometer on-board ERS-2 is an active microwave instrument operating in C-band (5.6 GHz) at VV polarization. The three SCAT antennas generate radar beams, at incidence angles ranging from 18° to 59°. The three antenna beams continuously illuminate a 500 km wide swath, each measuring the radar backscatter for overlapping 50 km wide cells. The backscatter measurements are converted to soil moisture estimates by applying the TUWien model [*Wagner et al.*, 1999]. To this end, the TUWien 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 Θ_S 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.

2.3. TMI Radiometer

[8] The TRMM mission is an earth observation satellite in an equatorial path between 40°N and 40°S. The TRMM Microwave Imager (TMI) is a passive microwave scanning radiometer, operating at five different wavelengths within the microwave spectrum (10.7, 19.4, 21.3, 37.0, and 85.5 GHz). The sensor measures the microwave brightness temperature at horizontal and vertical polarization. The spatial resolution of the different channels varies, and is 10 km for 37 GHz and 38 km for 10.7 GHz. The equatorial orbit results in varying overpass times for any given location, and a swath-width of approximately 800 km results in a return time in the order of five days. The brightness temperatures measured by TMI are converted to surface soil moisture applying the Land Parameter Retrieval Model (LPRM) [*Owe et al.*, 2008]. The LPRM is based on the solution of a microwave radiative transfer model and solves simultaneously for the surface soil moisture Θ_T, vegetation optical depth and land surface temperature without a-priori information of land surface characteristics. In this case, the LPRM is applied to the X-band (10.7 GHz) observations. VU University Amsterdam together with NASA Goddard Space Flight Centre provides the resulting global soil moisture data set(v03b).

3. Triple Collocation Error Model

[9] Our derivation of the error model closely follows the notation of *Janssen et al.* [2007]. In contrast to *Janssen et al.* [2007] we have to adopt the data calibration step to comply with the requirements of the soil moisture data. To account for the systematic differences of the soil moisture data sets we assume a linear relationship between our three estimates Θ_E, Θ_S and Θ_T and the hypothetical truth Θ (equation (1)), where r_E, r_S and r_T denote the residual error of the estimates Θ_E, Θ_S and Θ_T. The aim of our error model is to derive an estimate of the root mean square error e^2 which expresses the variance of the residual errors r of each data set.

$$\begin{aligned} \Theta_E &= \alpha_E + \beta_E \Theta + r_E \\ \Theta_S &= \alpha_S + \beta_S \Theta + r_S \\ \Theta_T &= \alpha_T + \beta_T \Theta + r_T \end{aligned} \tag{1}$$

[10] From equation (1) we eliminate the calibration constants by introducing the new variables $\Theta^*_E = \Theta_E/\beta_E - \alpha_E/\beta_E$ and $r^*_E = r_E/\beta_E$ etc., to obtain equation (2). It is important to note that this transformation also affects the residuals and consequently the error estimates e^2. To highlight this fact, we flag the transformed variables by a * symbol. Accordingly, the actual retrieval of e^2 has to follow a stepwise approach. In a first step we retrieve e^{*2}. In a second step we solve for the linear calibration and transform e^{*2} to e^2.

$$\begin{aligned} \Theta^*_E &= \Theta + r^*_E \\ \Theta^*_S &= \Theta + r^*_S \\ \Theta^*_T &= \Theta + r^*_T \end{aligned} \tag{2}$$

[11] The unknown truth can now be removed by a simple elimination procedure and we obtain equation (3).

$$\begin{aligned} \Theta^*_E - \Theta^*_S &= r^*_E - r^*_S \\ \Theta^*_E - \Theta^*_T &= r^*_E - r^*_T \\ \Theta^*_S - \Theta^*_T &= r^*_S - r^*_T \end{aligned} \tag{3}$$

[12] From equation (4) the variance of the residual errors e^{*2} can be retrieved by pairwise multiplying the lines of equation (3) and taking the average over a sufficiently large sample population (indicated by angle brackets). Under the assumption that the residual errors r_E, r_S and r_T are uncorrelated, the residual covariances $\langle r^*_E r^*_S \rangle = \langle r^*_E r^*_T \rangle = \langle r^*_S r^*_T \rangle$ become 0 and we get a direct estimate of $e^{*2}_E = \langle r^{*2}_E \rangle$, $e^{*2}_S = \langle r^{*2}_S \rangle$ and $e^{*2}_T = \langle r^{*2}_T \rangle$. The error variances are hence fully determined by three independent, calibrated soil moisture estimates (equation (4)).

$$\begin{aligned} e^{*2}_E &= \langle (\Theta^*_E - \Theta^*_S)(\Theta^*_E - \Theta^*_T) \rangle \\ e^{*2}_S &= \langle (\Theta^*_E - \Theta^*_S)(\Theta^*_S - \Theta^*_T) \rangle \\ e^{*2}_T &= \langle (\Theta^*_E - \Theta^*_T)(\Theta^*_S - \Theta^*_T) \rangle \end{aligned} \tag{4}$$

[13] Finally, to estimate e^2_E, e^2_S and e^2_T, we have to solve for the calibration expressed in equation (1). Since we do not know the truth, we have to arbitrarily chose one of the data sets as a reference. Hence only two of the three calibration constant pairs can be determined. We arbitrarily chose Θ_E as a reference, i.e., we set $\alpha_E = 0$ and $\beta_E = 1$. The calibration constants α_S, β_S and α_T, β_T can then be calculated by a simple linear least-squares approximation that considering errors in both variables [e.g., *Press and*

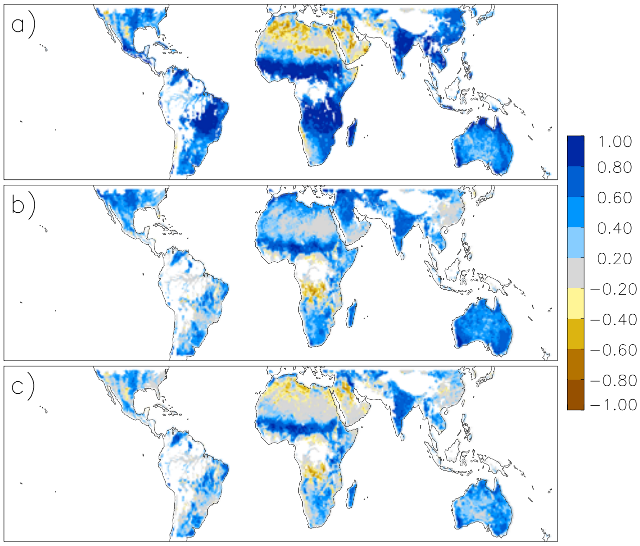

Figure 1. Correlation between (a) Θ_S and Θ_E; (b) Θ_T and Θ_E; and (c) Θ_S and Θ_T for the years 1998, 1999 and 2000.

Teukolsky, 1992]. Considering the symmetric nature of equation (1), this choice does not influence the estimation of $e^2{}_E$, $e^2{}_S$ and $e^2{}_T$. As the calibration affects the estimation of e_E^2, e_S^2 and e_T^2 we have to follow an iterative scheme. In this scheme, we start with estimating the initial guess of the calibration parameters (by assuming $e_E^2 = e_S^2 = e_T^2$) and subsequently solving the calibration and error equations until convergence is achieved.

4. Results

[14] The proposed model will only result in meaningful error estimates if the three data sets represent the same physical quantity. To test this prerequisite, we calculated the correlation between each collocated data set (Figure 1). A low or even negative correlation between one of the data pairs is a clear sign that either of the estimates does not provide meaningful soil moisture information. For example, we generally observe a low correlation in desert areas. This is expected, as the dynamic range of soil moisture is small and hence the correlation becomes distorted by noise. Nevertheless, for the ERS data set we even observe negative correlations, which are caused by unaccounted volume scattering effects of dry sand. Similarly, for the TMI derived soil moisture we observe negative correlation effects in regions with high vegetation cover, which is likely caused by the higher frequency of the TMI sensor, which is more sensitive to the vegetation structure than to the underlying surface soil moisture. To avoid the use of these spurious observations we mask those regions where the correlation drops below 0.2. Correlation coefficients above 0.2 indicate a significant correlation at the 0.05 confidence level according to a t-test.

[15] The results of the error estimation suggest that all three data sets are characterized by a low error (Figure 2). The mean global error is 0.020 m^3 m^{-3} for the ERA-Interim ($e*_E$), 0.028 m^3m^{-3} for the ERS-2 ($e*_S$) and 0.046 m^3m^{-3} for the TMI ($e*_T$) soil moisture data set. It is important to note that these errors refer to the climatology of the reference data set (in our case ERA-Interim). To normalize the errors we can scale them by the dynamic range of the ERA-Interim soil moisture which is defined by the wilting level (0.17 m^3 m^{-3}) and the level of saturation (0.47 m^3 m^{-3}). This results in average relative errors of 6.9%, 9.4% and 15.6% for ERA-Interim, ERS-2 and TMI respectively. It is worth noting that the relative magnitude of the error estimates do not depend on the choice of the reference, i.e. the ERA Interim data set does not profit relative to the other data sets from the selection as the reference. The analysis also reveals clear spatial patterns. The error in the ERA-Interim soil moisture, $e*_E$, is spatial consistent. Slightly larger values of $e*_E$ are found in the Monsoon regions where the reanalysis has problems to correctly reflect precipitation patterns. Although the values of $e*_S$ and $e*_T$ can be locally lower than $e*_E$, they are spatially less consistent. Whereas the higher $e*_T$ values can clearly be linked to regions of high vegetation cover the cause of the high $e*_S$ values are less evident but are possibly caused by azimuthal viewing and/or vegetation effects.

5. Conclusions

[16] The triple collocation technique is a promising method to estimate the error of global soil moisture data sets. The retrieved errors appear reasonable and the observed patterns can be explained by known performance issues of each data set. The results should however be interpreted carefully. Two assumptions are central for the validity of the derived error model: (i) Uncorrelated residual errors; and (ii) A linear relation between the data sets. As the measurement technique and retrieval concept of the data sets used in this study are fundamentally different, the assumption of uncorrelated errors appears justified. The second assumption is however not necessarily true, although all three data sets represent the same physical quantity. Considering that the three systems observe differ-

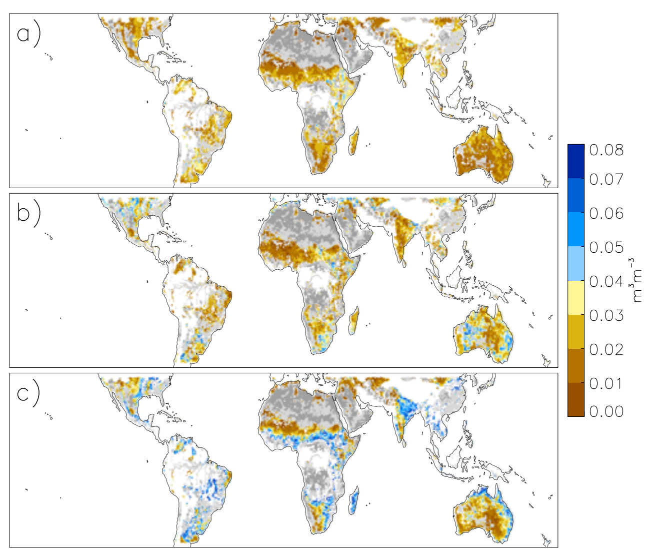

Figure 2. Spatial errors (a) $e*_E$ (ERA-Interim) (b) $e*_S$ (ERS-2); and (c) $e*_T$ (TMI) of derived soil moisture estimates. Grey colour indicates regions where the triple collocation error model cannot be applied as one of the three data sets shows a significantly different soil moisture behavior when compared to the other two data sets.

ent soil layers and hence different dynamics a higher order calibration might be necessary to avoid the introduction of systematic errors [*Drusch et al.*, 2005; *Reichle and Koster*, 2004]. It is worth stressing that the assumption of uncorrelated residual errors is absolutely vital and forms the basis of the approach. A sensible result can not be obtained if this assumption is relaxed, unless the covariance terms can be quantified accurately. The linearity assumption on the other hand could be relaxed by introducing a parametric transform in equation (1) and some type of nonlinear iterative root-finder instead of the linear least-squares approximation. In addition, it is important to note that the proposed calibration does not provide robust results if the signal to noise ratio is small and if the number of observations is low. In future applications we therefore recommend to include a fourth data set to cross check the retrieved errors.

[17] **Acknowledgments.** EUMETSAT's Hydrology Satellite Application Facility H-SAF partly funded this project. ERS scatterometer data were processed within the Austrian Science Fund project GLOBESCAT (project L148-N10). TMI data were processed within the EU 6th Framework program WATCH (project 036946-2).

References

Bartalis, Z., W. Wagner, V. Naeimi, S. Hasenauer, K. Scipal, H. Bonekamp, J. Figa, and C. Anderson (2007), Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT), *Geophys. Res. Lett.*, *34*, L20401, doi:10.1029/2007GL031088.

Caires, S., and A. Sterl (2003), Validation of ocean wind and wave data using triple collocation, *J. Geophys. Res.*, *108*(C3), 3098, doi:10.1029/2002JC001491.

Crow, W. T. (2007), A novel method for quantifying value in spaceborne soil moisture retrievals, *J. Hydrometeorol.*, *8*, 56–57.

Dirmeyer, P. A., Z. Guo, and X. Gao (2004), Comparison, validation, and transferability of eight multiyear global soil wetness products, *J. Hydrometeorol.*, *5*, 1011–1033.

Drusch, M., E. F. Wood, and H. Gao (2005), Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture, *Geophys. Res. Lett.*, *32*, L15403, doi:10.1029/2005GL023623.

Entin, J., A. Robock, K. Y. Vinnikov, S. Qiu, V. Zabelin, S. Liu, A. Namkhai, and T. Adyasuren (1999), Evaluation of global soil wetness project soil moisture simulations, *J. Meteorol. Soc. Jpn.*, *77*, 183–198.

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

Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem (2003), Soil moisture retrieval from AMSR-E, *IEEE Trans. Geosci. Remote Sens.*, *41*, 215–229.

Owe, M., R. de Jeu, and T. Holmes (2008), Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.*, *113*, F01002, doi:10.1029/2007JF000769.

Press, W. H., and S. A. Teukolsky (1992), Fitting straight line data with errors in both coordinates, *Comput. Phys.*, *6*, 274–276.

Reichle, R. H., and R. D. Koster (2004), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, *31*, L19501, doi:10.1029/2004GL020938.

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

Viterbo, P., and A. Beljaars (1995), An improved land surface parameterization scheme in the ECMWF model and its validation, *J. Clim.*, *11*, 2716–2748.

Wagner, W., G. Lemoine, and H. Rott (1999), A method for estimating soil moisture from ERS Scatterometer and soil data, *Remote Sens. Environ.*, *70*, 191–207.

Wagner, W., K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf (2003), Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data, *J. Geophys. Res.*, *108*(D19), 4611, doi:10.1029/2003JD003663.

Wagner, W., P. Pampaloni, G. Blöschl, J.-C. Calvet, B. Bizzarri, J.-P. Wigneron, and Y. Kerr (2007), Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, *Nord. Hydrol.*, *8*, 1–20.

R. de Jeu and T. Holmes, Department of Geo-Environmental Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1085, NL-1081 HV Amsterdam, Netherlands.

V. Naeimi and W. Wagner, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gusshausstrasse 27-29/E122, A-1040 Vienna, Austria.

K. Scipal, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 6AX, UK. (klaus.scipal@ecmwf.int)