Characterizing Coarse-Scale Representativeness of in situ Soil Moisture Measurements from the International Soil Moisture Network

In situ soil moisture measurements play a key role for a variety of large-scale applications. A deep understanding of their quality, especially in terms of spatial representativeness, is crucial for reliably using them as reference data. This study assesses random errors in the coarse-scale representation of in situ soil moisture measurements from more than 1400 globally distributed stations, drawn from the International Soil Moisture Network (ISMN), using the triple collocation method. The method was applied on the original measurements as well as on soil moisture anomalies. Error estimates were summarized for different networks, depths, and measurement principles and furthermore related to the respective climate class, soil type, average soil moisture condition, and soil moisture variability to find possible relationships between measurement errors and local properties. The average network error varies from about 0.02 to 0.06 m$^3$m$^{-3}$ with generally increasing error variability with increasing average error. Trends of (i) decreasing errors with increasing measurement depth and of (ii) increasing errors with increasing average soil moisture conditions and soil moisture variability were found for most networks and sensor types. The errors when looking into anomalies are in general lower than for absolute values. No statistically reliable trends for climate- and soil texture classes were found. These results highlighted the necessity of developing a comprehensive quality control process for in situ measurements to reliably exploit existing data sets and to select representative sites and sensors most appropriate for the requirements of a particular larger-scale application.

Abbreviations: CCI, Climate Change Initiative; CDF, cumulative distribution function; FDR, frequency domain reflectometry; IQR, interquartile range; ISMN, International Soil Moisture Network; NRT, near real time; TDR, time domain reflectometry; WACMOS, water cycle multi-mission observation strategy.

A deep understanding of soil moisture is crucial to describe hydrologic processes, earth-atmosphere interactions, and climate variability. The large variety of soil moisture applications require data from local (e.g., for agricultural support) to global scales (e.g., for climate change studies; Dorigo et al., 2012a).

A large number of local to regional scale meteorological and experimental networks measuring soil moisture in situ is available worldwide. Nevertheless, soil moisture is highly variable in space and time and a globally representative in situ network would require an extremely large number of stations. The high costs for operation and maintenance, together with the limited accessibility of certain regions, make the setup of such an in situ network financially infeasible. To fill this gap, remotely sensed data from optical and microwave instruments has been used to retrieve soil moisture on a global scale (Chauhan et al., 2003; Njoku et al., 2003; Wagner et al., 1999). Several missions such as the Soil Moisture and Ocean Salinity mission (SMOS; Kerr et al., 2010) or the Soil Moisture Active Passive (SMAP; Entekhabi et al., 2010), especially dedicated to estimate soil moisture with footprint sizes of several kilometers, have been launched recently or will be launched in the near future. Satellite sensors provide data with a maximum temporal resolution of 1 to 3 d, which is not sufficient for a large variety of applications. Land surface models have been implemented to fill this temporal gap providing data in about the same spatial resolution of meters to kilometers but several times a day.

Nevertheless, ground based measurements are still crucial not only for studying the spatial and temporal dynamics of soil moisture on a local scale (Brocca et al., 2010a; Brocca et al., 2007; Brocca et al., 2012; Mittelbach et al., 2011; Robinson et al., 2008; Vereecken et
al., 2010), but also for the calibration and validation of those data sets (Albergel et al., 2012; Parrens et al., 2011). In this context, in situ data sets are often seen as ground “truth,” which in fact is an inappropriate term since (i) in situ sensors and their deployment underlie inherent errors (Mittelbach et al., 2012; Plauborg et al., 2005; Walker et al., 2004) and (ii) differences in the spatial scale, the represented depth, and the underlying physical measurement principles of the sensors introduce systematic differences between the represented extent of the observation as well as the actual represented physical quantity (Brocca et al., 2007; Famiglietti et al., 1999; Famiglietti et al., 2008; Miralles et al., 2010). In situ sensors typically represent only a few centimeters of the soil and are usually placed in depths between the surface and 1 to 2 m to represent the plant root zone, whereas the penetration depth of satellite signals is about the size of the wavelength, i.e., about 0 to 5 cm (Schmugge 1983), covering an area from several square-meters to square-kilometers. Vegetation coverage, topography, soil type, spatial weather variability, and many other factors introduce subfootprint scale soil moisture variations that may cause differences between shallow spatial average surface soil moisture estimates from satellites or models and in situ measurements, depending on where the in situ sensor is located within the footprint. In fact, satellite and in situ sensors will never look into the same soil sample and hence do not measure the same water volume.

Nevertheless, Vachaud et al. (1985) introduced the temporal stability concept for soil moisture, stating that even though soil moisture is highly variable in space and time, spatial fields of soil moisture exist, which persist in time. Single stations within those fields can be used to represent the areal mean soil moisture behavior over larger areas. Based on this temporal stability concept, many other studies investigated the spatiotemporal variability of soil moisture over a large range of scales to assess the feasibility of using pointscale in situ measurements as a representation also for larger scale average soil moisture (Brocca et al., 2010a; Brocca et al., 2007; Brocca et al., 2012; Cosh et al., 2006; Famiglietti et al., 1999; Famiglietti et al., 2008; Martínez-Fernández and Ceballos, 2003; Martínez-Fernández and Ceballos, 2005). The main findings of these studies were that (i) a limited number of sites distributed over an area of interest can be used to reliably represent its average soil moisture behavior, (ii) the number of sites required to obtain a certain quality within a given confidence depends on the scale difference and the soil moisture conditions since spatiotemporal variability increases with increasing scale and reaches a maximum under intermediate wetness conditions, and (iii) almost all stations followed the temporal behavior of the areal mean in terms of the correlation whereas only few stations were able to represent the areal mean absolute soil moisture level. Differences in the latter mainly result from variations in vegetation, topography, soil texture, and climate, and are often from a systematic nature. Hence, appropriate scaling techniques can be used to remove those differences (Crow et al., 2012; Famiglietti et al., 2008; Kumar et al., 2012; Reichle and Koster, 2004). Single point-scale in situ measurements might then be properly used as a reference for coarse scale mean soil moisture (Albergel et al., 2012; Jackson et al., 2010). Further studies investigated also the temporal stability of soil moisture in depth, indicating a correlation of soil moisture along the soil profile (Martínez-Fernández and Ceballos, 2003; Pachepsky et al., 2005; Starks et al., 2006). In fact, Wagner et al. (1999) developed a method to estimate coarse scale profile soil moisture using surface soil moisture estimates from satellites. Various studies verified the usefulness of this method by comparing the estimated profile soil moisture to in situ measurements in depth (Albergel et al., 2009; Albergel et al., 2008; Brocca et al., 2010b). Their results proved an existing correlation for soil moisture along the profile.

From the previous it becomes clear that the use of in situ measurements for evaluating satellite or modeled data requires a detailed knowledge of their quality. One should be aware that the term quality for in situ data describes several properties, whose importance varies with the application. Four different properties should be distinguished: (i) the capability of an in situ sensor to measure absolute soil moisture levels, (ii) the capability of capturing the temporal dynamics (drying and wetting events), (iii) the spatial representativeness of a single station for a certain area, and (iv) the inherent sensor reliability (e.g., the probability for sensor-dropouts or outliers, sensor drifts, or random noise). Different approaches are available for assessing one or more of these properties. Common metrics are the correlation coefficient, which assesses the temporal relationship between two data sets, and the bias and RMSD for quantifying relative measurement differences. Those metrics can be applied on the soil moisture measurements directly or on soil moisture anomalies, which are usually defined as the difference between actual measurements and a long-term climatology (Albergel et al., 2012; Dorigo et al., 2010). Looking into the direct measurements addresses the sensor’s capability of measuring absolute soil moisture levels, whereas anomalies can be used to assess the capability for capturing drying and wetting events. Using a high spatial sensor density allows the assessment of the spatial representativeness of in situ sensors when comparing the respective measurements with the areal average. All the mentioned metrics require reliable reference data sets, which are usually manually collected gravimetric samples or high quality sensor measurements taken under laboratory conditions with prepared soil samples or under field conditions (Mittelbach et al., 2012; Cataldo et al., 2009; Plauborg et al., 2005). A new approach of assessing
the occurrence of measurement outliers without the need of any reference data was introduced by Dorigo et al. (2012b), who used spectrum-based analysis to find and flag spikes, jumps, saturated responses, missing precipitation responses, and sensor drop-outs.

One possible method for quantifying measurement errors without relying on the quality of reference data sets might be the so-called triple collocation method. This is a method for estimating the random errors of three collocated data sets which can be assumed to represent the same physical parameter while simultaneously solving for systematic differences. The method assumes independent (uncorrelated) error structures, which means that the errors must not have the same origin. This is given when using e.g., any combination of in situ measurements, active or passive satellite observations, and land surface model estimates, provided that the model is not driven by one of the others. Several studies highlighted the high potential of the triple collocation in becoming a standard procedure in a comprehensive satellite validation process (Dorigo et al., 2010; Miralles et al., 2010; Scipal et al., 2008; Stoffelen, 1998). Nevertheless, studies also showed that the result is highly sensitive to its input configuration, including different scales and represented physical quantities of the sources, the use of absolute values or anomalies, the time span under observation, and the available number of measurement triplets (Loew and Schlenz, 2011; Zwieback et al., 2012b; Zwieback et al., 2012a). As mentioned, large scale differences between the input data sets introduce errors caused by the spatiotemporal variability of soil moisture, leading to a mismatch in the spatial representativeness. These errors are reflected in the triple collocation result and can take a high fraction of the overall error, leading to an overestimation of the actual inherent sensor error (Miralles et al., 2010). The only way to remove these scaling errors is the assessment of the spatial representativeness, which requires a high spatial sensor density. Unfortunately, only few available in situ networks provide such a density.

The main objective of this study is to use the triple collocation to characterize the random errors of globally available in situ measurements in their purpose of representing footprint-scale (~0.25°) soil moisture. It should be emphasized that, in the context of this study, the term random error describes not only the inherent random sensor noise but also the nonsystematic part in the scaling process, caused by the spatiotemporal variability of soil moisture (referred as scaling error), which is most likely dominating the overall error estimate. A secondary objective of this study is to investigate possible relationships between error levels and site-specific properties which are expected to have a large impact on in situ measurements. These properties are (i) observation depths, (ii) sensor types, (iii) climate regions, and (iv) soil types. The data sets used for this study are drawn from the ISMN.

Data Sets

Random errors of the ISMN are assessed using the triple collocation method, which requires two additional data sources with independent error structures. A blended active and passive remotely sensed data set and the ERA-Interim reanalysis data set were used for this purpose as they have the biggest temporal overlap with the available in situ measurements. The global Köppen–Geiger map was used to relate the stations to climate classes. Soil texture information was drawn from the Harmonized World Soil Database (HWSD).

The International Soil Moisture Network

The ISMN (http://www.ipf.tuwien.ac.at/insitu) is a centralized data hosting facility. It collects soil moisture ground measurements and, if available, ancillary measurements such as precipitation, soil temperature, air temperature, snow depth, and snow water equivalent from operational and experimental networks worldwide (Dorigo et al., 2011a; Dorigo et al., 2011b). As there are no standard methods yet for collecting soil moisture data in situ, the data sets are usually highly different in terms of sensor installation depths and placement, temporal sampling, used units, and data formats. The ISMN harmonizes incoming data sets in terms of units, sampling interval, data format, and metadata and makes them available to users cost-free from a single web portal. Initiated by ESA in 2010 and operated by the Vienna University of Technology (TU Wien), it has evolved as one of the most important in situ soil moisture data platforms for satellite and land surface model validation (e.g., Albergel et al., 2012; Liu et al., 2012). Providing networks operate one or more geographically distributed stations which typically place a variety of sensors to cover different depths but also to increase the measurement reliability by making redundant measurements with equal sensors in the same depths close to each other. Currently (October 2012), the ISMN holds the measurements of over 6100 soil moisture sensors, provided by 35 different networks, which operate together more than 1400 stations. Data providers are listed in the Acknowledgments (i.e. Brocca et al., 2011; Marczewski et al., 2010; Robock et al., 2000; Su et al., 2011). Figure 1 illustrates the global station distribution. Most of them are located in Northern America and Eurasia and spread over a variety of climate regions, land cover types, and soil textures. The temporal coverage of each network is shown in Fig. 2. Data sets cover a time period from 1952 (historical data sets) to now, while six networks and together with more than 200 stations are operating in near real time (NRT). A large variety of sensor types are used, placed at different depths, and representing different depth intervals. Table 1 gives an overview of the used sensors and depth placements for each network. Available sensors make use of different measurement principles such as time domain reflectometry (TDR), frequency domain reflectometry (FDR), capacitance probes, impedance probes, neutron probes, cosmic ray probes, and gravimetric measurements, all of which result in different sampling intervals and data accuracy (Mittelbach et al., 2011; Mittelbach et al., 2012; Plauborg et al., 2005; Walker et al., 2004). For this study, the sensor types were summarized into five groups after similar properties:
Fig. 1. Station distribution of the International Soil Moisture Network [ISNM (October 2012)]. Pins represent single stations, colors represent different networks.
Fig. 2. Temporal coverage of the International Soil Moisture Network (ISMN) networks, Water Cycle Multi-Mission Observation Strategy (WAC-MOS) and ERA-Interim.

Table 1. Overview of the number of stations, the used sensors, the total observed depth range, and the number of sensor placements in different depths for each network, respectively.

<table>
<thead>
<tr>
<th>Network</th>
<th>Stations</th>
<th>Sensors</th>
<th>Depth covered</th>
<th>Different depth placements</th>
</tr>
</thead>
<tbody>
<tr>
<td>AACES</td>
<td>49</td>
<td>ThetaProbe ML2X</td>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td>AMMA</td>
<td>7</td>
<td>CS616</td>
<td>0.05–1.20</td>
<td>13</td>
</tr>
<tr>
<td>ARM</td>
<td>25</td>
<td>SMP1</td>
<td>0.03–1.75</td>
<td>10</td>
</tr>
<tr>
<td>CALABRIA</td>
<td>5</td>
<td>ThetaProbe ML2X</td>
<td>0.30–0.90</td>
<td>3</td>
</tr>
<tr>
<td>CAMPANIA</td>
<td>2</td>
<td>ThetaProbe ML2X</td>
<td>0.30–0.30</td>
<td>1</td>
</tr>
<tr>
<td>CHINA</td>
<td>40</td>
<td>Coring device/auger</td>
<td>0.00–1.00</td>
<td>11</td>
</tr>
<tr>
<td>COSMOS</td>
<td>67</td>
<td>Cosmic-ray Probe</td>
<td>Dependent on water content</td>
<td>–</td>
</tr>
<tr>
<td>FLUXNET-AMERIFLUX</td>
<td>2</td>
<td>Moisture Point PRB-K</td>
<td>0.00–0.50</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ThetaProbe ML2X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMI</td>
<td>1</td>
<td>ThetaProbe ML2X</td>
<td>0.02–0.10</td>
<td>2</td>
</tr>
<tr>
<td>HOBE</td>
<td>30</td>
<td>Decagon 5TE</td>
<td>0.00–0.55</td>
<td>3</td>
</tr>
<tr>
<td>HSC_SELMACHEON</td>
<td>1</td>
<td>Hydraprobe Analog (CR800)</td>
<td>0.00–0.10</td>
<td>1</td>
</tr>
</tbody>
</table>

Table continued..
gravimetric, capacitance, TDR/FDR, impedance, cosmic ray, and neutron probes.

Only data sets that achieve the statistical requirements of the triple collocation were used (See Methodology). All available sensor depths were used to investigate the error dependency on the measurement depth. Soil texture and climate region analysis were based on surface measurements only, i.e., on measurements of which the start of the measurement interval lies between 0 and 10 cm.
Remotely Sensed Soil Moisture
The satellite data set used in this study was the data set created within the Water Cycle Multi-Mission Observation Strategy project (WACMOS; http://www.esa-soilmoisture-cci.org), released in June 2012 within the framework of the Climate Change Initiative (CCI). It is the first available long-term remotely sensed soil moisture product covering a 32 yr period from 1978 to 2010, providing data in a 0.25° spatial resolution and was generated by merging active and passive soil moisture estimates from various satellite missions (Liu et al., 2012; Liu et al., 2011). Merging different instruments from various satellites based on their temporal availability causes an increase of data quality with time. The temporal resolution is approximately 1–3 d.

ERA-Interim
ERA-Interim is a global atmospheric reanalysis data set combined with an ocean and land surface model produced by the European Centre for Medium-range Weather Forecasts (ECMWF; Dee et al., 2011). It covers a time period from 1979 to June 2012 and provides data in a spatial resolution of ~0.7° at the equator. Soil moisture estimates are provided for four different layers (0–7, 7–28, 28–100, and 100–255 cm) four times each day (0:00, 6:00, 12:00, and 18:00) for two different land surface schemes (TESSEL and HTESSEL; Balsamo et al., 2010). The HTESSEL scheme was used in this study because it provides a more realistic representation of the soil than the TESSEL scheme by distinguishing between six different soil types around the globe instead of just one. ERA-Interim also provides soil temperature estimates for the respective layers and an estimate of snow height. These were used to mask soil moisture measurements for which the temperature is below 0°C and for which the snow cover estimate is greater than zero, as both satellites and most of the in situ sensors are making use of electromagnetic properties of the soil, which significantly change when the soil is frozen (Ulaby et al., 1982).

Koeppen–Geiger Climate Classification
The Koeppen–Geiger classification divides the globe into climate regions based on their temperature and precipitation regime. The updated world map from Peel et al. (2007) was used for this study (http://www.hydrol-earth-syst-sci.net/11/1633/2007/hess-11-1633-2007-supplement.zip, accessed 25 July 2011). It contains a static map on a 0.1 degree grid based on long-term in situ observations between 1951 and 2000.

Harmonized World Soil Database
The Harmonized World Soil Database (HWSD; http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/) is a merged and harmonized product from various soil information sources providing information about soil properties such as material fractions, bulk density, or texture classes on a 1-km grid (Nachtergaele et al., 2009). For this study, only the USDA soil texture classification is used, which for two layers (topsoil: 0–30 cm, subsoil: 30–100 cm) classifies the soil in the dominant fraction of clay, silt, and sand.

Methodology

Theory of the Triple Collocation
The triple collocation estimates the random error of three collocated data sets with independent error structures. In a first step, a relationship to the “truth” must be defined, e.g., a linear one:

\[ \Theta_1 = \alpha_1 + \beta_1 \Theta + \varepsilon_1 \]
\[ \Theta_2 = \alpha_2 + \beta_2 \Theta + \varepsilon_2 \]
\[ \Theta_3 = \alpha_3 + \beta_3 \Theta + \varepsilon_3 \]

\[ \Theta_1, \Theta_2, \text{and } \Theta_3 \text{ represent the measurement, } \alpha_1 \text{ and } \beta_1 \text{ the linear coefficients, } \varepsilon_1 \text{ the random error and } \Theta \text{ the “true” soil moisture } (i = 1, 2, 3). \]

By selecting one data set as the reference, i.e., setting \( \alpha_R = 0 \) and \( \beta_R = 1 \), and rescaling the others into its data space, Eq. [2] is obtained (the index \( R \) indicates the number of the chosen reference):

\[ \Theta_1^* = \Theta + \varepsilon_1^* \]
\[ \Theta_2^* = \Theta + \varepsilon_2^* \]
\[ \Theta_3^* = \Theta + \varepsilon_3^* \]

\( \Theta_1^*, \Theta_2^*, \text{and } \Theta_3^* \text{ represent the rescaled measurements and the rescaled random errors, respectively. By subtracting the equations in Eq. [2] and averaging over the cross-multiplied differences we finally obtain Eq. [3]:} \]

\[ \frac{\langle (\Theta_1^* - \Theta_2^*) (\Theta_1^* - \Theta_3^*) \rangle}{\langle \varepsilon_1^* \rangle} = \varepsilon_1^* \]
\[ \frac{\langle (\Theta_1^* - \Theta_2^*) (\Theta_2^* - \Theta_3^*) \rangle}{\langle \varepsilon_2^* \rangle} = \varepsilon_2^* \]
\[ \frac{\langle (\Theta_1^* - \Theta_3^*) (\Theta_2^* - \Theta_3^*) \rangle}{\langle \varepsilon_3^* \rangle} = \varepsilon_3^* \]

\[ \varepsilon_i^* = \left\{ \varepsilon_i^* \right\} \]

\( \varepsilon_i^* \) represent the averaged random errors of the data sets expressed in the data space of the chosen reference. Theoretically, the choice of the reference should have no other impact on the error estimates of the other sets than the definition of their data space and since the linear coefficients are solved to obtain Eq. [2], one could transform them back into their own. For a reasonable comparison, they are usually kept in the same data space.

The success of Eq. [3] requires three basic assumptions to be fulfilled: (i) the three data sets must represent the same physical quantity, i.e., must be significantly correlated, (ii) the errors must be uncorrelated, i.e., the covariances must be zero, and (iii) an appropriate number of triplets must be available to obtain reliable estimates in the averaging step. Based on Zwieback et al. (2012b), we decided that at least about 100 triplets must be available. Nevertheless, the limited number of triplets might artificially inflate the error estimates with a magnitude higher than the actual inherent...
sensor error or the scaling error. Furthermore, the results in this study are filtered using the Student’s t test for checking the significance of the correlation between all three data sets with a probability threshold of 0.05.

Implementation of the Triple Collocation

Varying input settings have a high impact on the reliability of the result as well as on its actual meaning (Dorigo et al., 2010; Zwieback et al., 2012b). Applied to the original soil moisture values, the result provides information about the ability of measuring absolute soil moisture whereas the use of soil moisture anomalies gives information about the capability of catching drying and wetting events, e.g., through precipitation (Dorigo et al., 2010). In this study, the triple collocation was applied on both original values and anomalies. Furthermore, the choice of the rescaling method, which is required to obtain Eq. [2], may significantly change the result if the data sets are different in their statistical properties. The choice of an inappropriate rescaling technique will artificially inflate the error estimates in addition to the inherent errors caused by a mismatch in the spatial representativeness. As mentioned in the Theory of the Triple Collocation, the choice of the reference data set effects in which data space the errors are expressed. The in situ data was always used as the reference in this study.

Rescaling of Original Estimates

Kumar et al. (2012) found that slight nonlinearities exist between soil moisture data sources, leading to improved scaling results for cumulative distribution function (CDF)-matching techniques compared to linear approaches. The CDF-method matches the cumulated distribution function of the data sets and hence corrects for nonlinearities by correcting theoretically all higher statistical moments (Reichle and Koster, 2004). Depending on the implementation, mainly mean, standard deviation, skewness, and kurtosis are affected. We choose to apply it when using absolute soil moisture measurements as a stepwise linear scaling between a set of percentiles of the data (Liu et al., 2011).

Anomalies

Anomalies usually are the difference between actual measurements and the long-term climatology. However, if data sets do not have a sufficient temporal coverage, climatologies cannot be computed reliably. Another way of calculating anomalies is to use a moving average window to create a baseline for the subtraction (e.g., Albergel et al., 2012), as shown in Eq. [4].

$$\Theta_{A}(t) = \Theta(t) - \overline{\Theta(W)}$$  

$$\Theta_{A}$$ is the soil moisture anomaly, $$\Theta$$ the observed absolute soil moisture, $$t$$ the time of acquisition, and $$W$$ the length of the moving window. An appropriate window length allows for removing systematic differences between different data sources which would be interpreted as random errors, while preserving the response of the individual data sets to short-term drying and wetting events (e.g., the seasonal vegetation growing cycles strongly affect satellite observations, but not in situ measurements). We chose a window length of 5 wk ($$t \pm 17$$ d) according to Albergel et al. (2012). The anomaly is only computed when at least five measurements are available within the respective window, even though most in situ stations provide an hourly measurement rate.

The anomalies of remotely sensed and ERA-Interim soil moisture were scaled to those of the in situ sensors using a normalization approach that matches the mean and the standard deviation as shown in Eq. [5] (Dorigo et al., 2010):

$$\Theta'_{A,S}(t) = \Theta_{A,R}(t) + \frac{\text{Var}(\Theta_{A,R})}{\text{Var}(\Theta_{A,S})} [\Theta_{A,S}(t) - \overline{\Theta_{A,S}}]$$  

[5]

$$\Theta_{A,S}(t)$$ is the anomaly at the time $$t$$, $$\Theta_{A,R}$$ the reference anomaly data set, $$\text{Var}()$$ refers to the variance, and the overline to the mean value. $$\Theta'_{A,S}(t)$$ is the rescaled measurement at the time $$t$$.

Spatial and temporal collocation

The spatial collocation was performed by using the in situ stations as a reference and searching the respective satellite and ERA-Interim ground point closest to the station.

In situ sensors cover a varying depth-range, whereas the ERA-Interim data set represents four fixed layers (see ERA-Interim). Since the in situ measurements sometimes overlap with more than one ERA-Interim layer, a depth-collocation was performed by assigning the sensor depth to the ERA-Interim layer that covers the start of its measurement interval (e.g., the 0–7 cm layer of ERA-Interim is used when the sensor is placed in 5–10 cm). The same depth collocation was done for the two layers of the Harmonized World Soil Database.

For the temporal collocation, the data source with the lowest measuring frequency, which was the remote sensing data set, was taken as a reference to search for the closest valid measurement of the other sources with a maximum difference of ±3 h.

Analysis

Sensor Quality on a Network, Depth and Sensor Type Level

The aim is to evaluate (i) whether the measurement quality changes with sensor positioning or for different sensor types, (ii) whether some measurements show different error levels when looking into anomalies instead of absolute values, and (iii) whether there are networks that provide more reliable measurements than others. The error estimates were therefore grouped with respect to the networks, the observation depths (the four ERA-Interim depth intervals were used to summarize the in situ measurement depths as described in Spatial and Temporal Collocation) and the used
measurement techniques. Standard statistics (median, inter-quartile-range, and outliers) were computed for the comparison and the estimated errors of the absolute soil moisture measurements were plotted against the estimated errors of the anomalies.

Climate Class Analysis
The error estimates of the entire ISMN were grouped after the Köppen–Geiger climate classes to evaluate a possible impact of the climate conditions on the measurement quality. We assumed that even though daily temperature fluctuations are known to cause variations in the sensor readings (Dorigo et al., 2012b), an influence would be mainly driven by precipitation rather than by temperature regimes. Hence, we summarized climate classes with similar or equal precipitation patterns to increase the statistical significance. The grouping is shown in Table 2. The distribution of available measurements within the climate classes is shown in Fig. 3. The estimated errors within the classes were furthermore compared to the median soil moisture state of the entire measurement period. We decided to consider only surface measurements for this analysis, i.e., measurements of which the start of the depth interval lies between 0 and 10 cm, since we assumed that the biggest impact, if apparent, will be at the surface. Besides, the majority of sensors in the ISMN are placed close to the surface, so the surface measurements allow for the most meaningful inter-comparison because possible impacts of the depth mismatch between in situ sensors and the satellite signal are reduced.

Soil Texture Analysis
In situ sensors are placed over a large variety of soil types, which on the one hand influences soil moisture storage and redistribution properties (e.g., infiltration and evaporation rates or total water storage volume), and on the other hand shows different responses to the physical measurement principle of the sensors. To evaluate a possible influence of the soil type on the measurement quality, we grouped the error estimates with respect to the USDA soil texture classes according to three dominant soil constituent types: Clay, silt, and sand (Table 3). The distribution of available measurements

Table 2. Köppen–Geiger climate classes summarized by similar or equal precipitation regimes.

<table>
<thead>
<tr>
<th>Summarized classes</th>
<th>Original classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aw</td>
<td>Aw (Tropical–Savannah)</td>
</tr>
<tr>
<td>BWx</td>
<td>BW (Arid–Desert–Hot)</td>
</tr>
<tr>
<td>BWk</td>
<td>BW (Arid–Desert–Cold)</td>
</tr>
<tr>
<td>BSx</td>
<td>BS (Arid–Steppe–Hot)</td>
</tr>
<tr>
<td>BSk</td>
<td>BS (Arid–Steppe–Cold)</td>
</tr>
<tr>
<td>Csx/Dsx</td>
<td>Cs (Temperate–Dry Summer–Hot Summer)</td>
</tr>
<tr>
<td>Csb</td>
<td>Cs (Temperate–Dry Summer–Warm Summer)</td>
</tr>
<tr>
<td>Dsa</td>
<td>Dsa (Cold–Dry Summer–Hot Summer)</td>
</tr>
<tr>
<td>Db</td>
<td>Dsa (Cold–Dry Summer–Cold Summer)</td>
</tr>
<tr>
<td>Dwx</td>
<td>Dwa (Cold–Dry Winter–Hot Summer)</td>
</tr>
<tr>
<td>Dwc</td>
<td>Dwa (Cold–Dry Winter–Cold Summer)</td>
</tr>
<tr>
<td>Cf/Dfx</td>
<td>Cf (Temperate–Without dry season–Hot Summer)</td>
</tr>
<tr>
<td>Cfb</td>
<td>Cf (Temperate–Without dry season–Warm Summer)</td>
</tr>
<tr>
<td>Dfa</td>
<td>Dfa (Cold–Without dry season–Hot Summer)</td>
</tr>
<tr>
<td>Db</td>
<td>Dfa (Cold–Without dry season–Cold Summer)</td>
</tr>
<tr>
<td>Dfc</td>
<td>Dfa (Cold–Without dry season–Cold Summer)</td>
</tr>
<tr>
<td>ETH</td>
<td>ETH (Polar–Tundra–High Elevation)</td>
</tr>
</tbody>
</table>

Table 3. USDA soil texture classes summarized by dominant soil types.

<table>
<thead>
<tr>
<th>Summarized classes</th>
<th>Original classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clay</td>
<td>clay (heavy)</td>
</tr>
<tr>
<td></td>
<td>silty clay</td>
</tr>
<tr>
<td></td>
<td>clay</td>
</tr>
<tr>
<td></td>
<td>sandy clay</td>
</tr>
<tr>
<td>Sand</td>
<td>sand</td>
</tr>
<tr>
<td></td>
<td>loamy sand</td>
</tr>
<tr>
<td>Loam</td>
<td>sandy loam</td>
</tr>
<tr>
<td></td>
<td>sandy clay loam</td>
</tr>
<tr>
<td></td>
<td>clay loam</td>
</tr>
<tr>
<td></td>
<td>silty clay loam</td>
</tr>
<tr>
<td></td>
<td>loam</td>
</tr>
<tr>
<td></td>
<td>silt loam</td>
</tr>
</tbody>
</table>

Fig. 3. Distribution of available measurements within the original Köppen–Geiger classes (left) and the summarized classes (right).
within the texture classes is shown in Fig. 4. The estimated errors were again compared to the median soil moisture of the entire period. Only surface measurements were considered for this analysis for the same reasons as for the climate class analysis.

**Results and Discussion**

**Errors on a Network, Depth, and Sensor Type Level**

Figure 5 shows statistics of error estimates of absolute soil moisture measurements for each network and measurement depth. The average error of the networks varies from about 0.02 to 0.06 m$^3$ m$^{-3}$. In addition, the error variability [in terms of the interquartile range (IQR)] changes significantly between the networks from about 0.01 to 0.05 m$^3$ m$^{-3}$ with a pattern of increasing variability with increasing average error. A clear decrease in the measurement errors with increasing measurement depth can be seen for all networks except for the deepest layers of SCAN and SNOTEL. AMMA, ARM, and HOBE show a rather stable behavior. One reason for the observed error decrease with increasing measurement depth could be that the magnitude of daily temperature fluctuations, which are known to have an impact on the sensor readings (Dorigo et al., 2012b) also decreases in deeper layers. Another possible explanation is the smoothing effect on the signal due to the decrease of temporal variability of soil moisture in deeper layers. Localized weather phenomena that affect the spatio-temporal variability and hence the spatial representativeness might also have a lower impact on deeper layers. Figure 6 shows the same statistics for different sensor types and measurement depths, again for absolute soil moisture measurements. All sensor types show a decrease of the errors with increasing depth, except for TDR/FDR and impedance probes placed in the deepest layer. The varying amount of measurements available to calculate the statistics for a certain network, depth or sensor type might lower its reliability (e.g., for AMMA, COSMOS, HYDROL-NET_PERUGIA, ICN, MOL-RAO, UMBRIA, UMSUOL, or cosmic ray and deep layer capacitance sensors) and might give a wrong impression of error variability under certain conditions. Several networks, which are listed in Table 1, are not shown in the results because they either didn’t fulfill the statistical requirements of the triple collocation method (were not significantly correlated with the satellite or the model) or had a too limited or no temporal overlap with the satellite and model data. No error estimates for neutron probes and gravimetric measurements are shown for the same reason.

Above described patterns are similar when using anomalies instead of absolute soil moisture measurements. Figure 7 shows the error estimates for original values against error estimates for anomalies for all available measurements. One can see that, in general, errors of anomalies are lower than for absolute values with an increasing discrepancy for increasing errors. A trend for decreasing errors with increasing depth is again visible except for very deep measurements, which might be caused by the significantly lower number of error estimates in those depths. Soil moisture anomalies typically show a lower dynamic range than absolute soil moisture values, especially when calculating them as the difference from the moving-average baseline like it was done in this paper. Hence, the data space in which the triple collocation expresses the errors show a lower dynamic range, which could support the impression of lower error estimates for anomalies. In addition, the dynamic range of absolute soil moisture measurements varies between the different networks and stations because of the different climate, soil, and terrain properties. Figure 8 shows the relationship between error estimates and the dynamic range of the sensors for all used stations. One can see a clear trend of increasing error with increasing measurement variability.
Fig. 5. Box–Whiskers plot for the triple collocation result summarizing different observation depths of each network. The box represents the upper and lower quartile together with the median; the whisker length is 1.5 times the interquartile range drawn from the respective quartile. Red crosses indicate outliers exceeding the whiskers. Values in brackets show the average number of triplets used for the error estimate/the total number of error estimates used to calculate the statistic in the respective column.

Fig. 6. Box–whiskers plot for the triple collocation result summarizing different sensor types and depths. The box represents the upper and lower quartile together with the median; the whisker length is 1.5 times the interquartile range drawn from the respective quartile. Red crosses indicate outliers exceeding the whiskers. Values in brackets show the average number of triplets used for the error estimate/the total number of error estimates used to calculate the statistic in the respective column.
Possible explanations for this are (i) the different data space in which the triple collocation expresses the error estimates and (ii) that the magnitude of random errors usually depends on the signal variability.

Climate Class Analysis

Figure 9 shows the errors of measurements within different climate classes related to the median soil moisture at the respective site. One can see a clear trend of increasing errors for wetter average conditions in cold arid steppe regions. A similar trend can be seen for temperate and cold regions with dry summer as well as for arid desert regions, but very few data points are available to make this statement statistically reliable. This behavior is consistent with studies investigating inherent sensor errors (e.g., Mittelbach et al., 2011). Temperate and cold regions without a dry season appear to show two interfering phenomena pronounced as an apparent cross in the scatterplot: One trend for increasing errors with increasing average soil moisture and also a second trend for decreasing errors with increasing average soil moisture. Connecting the climate region to the soil type could not explain this behavior (not shown). A possible reason for the convex upward relationship between average soil moisture conditions and error levels is the spatiotemporal variability of soil moisture that reaches a maximum under intermediate wetness conditions (e.g., Brocca et al., 2012). The interfering error peaks for very low and very high average soil moisture levels could not be explained and should be investigated in further studies. Too few or too randomly
Fig. 9. Errors in absolute values against the average soil moisture condition for summarized climate classes. Colors indicate the detailed classes according to the Koeppen–Geiger classification.
spread data points are available to see any pattern for tropical savannah, polar tundra and cold regions with dry winter. No significant differences in error budgets are visible between particular climate regions. The impression of slightly lower errors in tropical savannah and arid desert regions might be caused by the lack of measurements in these regions.

Soil Texture Analysis

Figure 10 shows the errors of measurements within different soil types related to the median soil moisture at the respective site. Again, a slight trend of increasing errors with increasing average soil moisture conditions is visible but not as pronounced as for different climate regions. The same interfering trend of increasing error with decreasing average soil moisture might be apparent in clay and loam and also less pronounced than for the climate classes. Sensors placed in sand appear to have slightly lower errors, which can be again caused by a lack of data points. The general reliability of the soil texture analysis is hampered by three facts: (i) the HWSD only represents dominant soil types in two coarse layers, (ii) the soil texture at the sites might significantly differ from the coarse scale average soil texture provided from the HWSD, and (iii) several soil texture borders are following country borders instead of natural landscape features, questioning the reliability of the data sources. Site-specific soil texture information from the data providers would help to overcome this issue, but it is only available for very few networks.

Conclusions and Outlook

This study investigated the quality of in situ measurements of the ISMN for representing soil moisture at footprint scales on a global basis using random error estimates of the triple collocation method. These errors reflect the actual inherent sensor measurement error, i.e., sensor noise and malfunctions, overlaid with external errors, which are (i) systematic differences between the statistical properties of the different data sources that cannot be removed with the CDF-matching and are hence interpreted as random errors, and (ii) scaling errors caused by the spatiotemporal variability of soil moisture, i.e., the limited spatial representativeness of the in situ sensors (horizontally but also in depth) when comparing it with larger scale satellite and model data. Systematic differences between the in situ measurements and the “true” soil moisture state (e.g., through miscalibration) cannot be resolved and might lead to an additional scaling of the error estimates. Besides, harmed assumptions in the triple collocation (e.g., too few data triplets or nongaussianity of the data sets) might additionally inflate the error estimate. Since many studies show that one single sensor might not be sufficient to represent larger-scale soil moisture, it is very likely that both external error sources dominate the overall error estimate and that the triple collocation result thus mainly reflects the spatial representativeness of the sensors as a function of the spatiotemporal variability of soil moisture. This cannot be proven...
A high variation in average errors levels of particular networks and sensor types as well as in error variability within those was found. A global trend for decreasing errors with increasing measurement depth and for increasing errors with increasing average soil moisture conditions was observed, independent of the soil type and climate region. An interfering trend of decreasing errors with increasing average soil moisture conditions, which could only partly be explained with the spatiotemporal variability of soil moisture, is visible in data sets within temperate and cold regions without clear dry seasons, but less pronounced in clay and loam. Almost all sensors show lower errors when looking into anomalies instead of average soil moisture, which is mainly caused by the lower dynamic range of anomalies. A clear relationship between the dynamic range and error levels was also found for absolute soil moisture measurements. Moreover, 35.8% of the data sets exceed the current satellite mission accuracy requirement of 0.04 m3 m-3 in terms of the triple collocation error estimate for absolute soil moisture values. In situ measurements are often considered as the “true” reference for this requirement, but also for a large variety of other applications and must therefore achieve significantly lower error levels. The limited spatial representativeness of single in situ stations for larger-scale soil moisture levels and the limited knowledge about inherent sensor errors question the meaning of a single number for a direct comparison between in situ sensors and satellite instruments. Almost all stations out of more than 1400 show considerable errors that should be taken into account in most applications, making the development of a standard procedure for a comprehensive quality assessment an essential task, including the development of procedures to reliably select representative existing or future sites for the in situ–satellite inter-comparison.

This study investigated the results of the triple collocation, which is just one approach for estimating random errors and should be seen only complementary to other tools such as correlation, RMSE, and bias analysis or the assessment of spatial representativeness, since all these methods characterize different quality properties. The requirements on these properties highly vary with application and not a single sensor or site is capable to fulfill all of them. Detailed knowledge about the requirements of the particular application is crucial to support the comprehensive quality assessment by allowing the best possible selection of existing sites, but also by supporting the selection of representative locations and the best fitting sensor type for the setup of new sites. Finally, it helps to avoid misinterpretations of results based on in situ data under the assumption of freedom from errors.

Acknowledgments

References


Wagner, G., A. Xaver, M. Vreugdenhil, A. Gruber, A. Hegyiová, A.D. Sanchis-