Potential and limitations of multidecadal satellite soil moisture observations for climate model evaluation studies

A. Loew\textsuperscript{1}, T. Stacke\textsuperscript{1}, W. Dorigo\textsuperscript{2}, R. de Jeu\textsuperscript{3}, and S. Hagemann\textsuperscript{1}

\textsuperscript{1}Max-Planck-Institute for Meteorology, KlimaCampus, Hamburg, Germany
\textsuperscript{2}Department of Geodesy and Geo-Information, Vienna University of Technology, Vienna, Austria
\textsuperscript{3}Department of Earth Sciences, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, the Netherlands

Received: 15 February 2013 – Accepted: 9 March 2013 – Published: 18 March 2013
Correspondence to: A. Loew (alexander.loew@zmaw.de)
Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Soil moisture is an essential climate variable of major importance for land-atmosphere interactions and global hydrology. An appropriate representation of soil moisture dynamics in global climate models is therefore important. Recently, a first multidecadal, observational based soil moisture data set has become available that provides information on soil moisture dynamics from satellite observations (ECVSM). The present study investigates the potential and limitations of this new dataset for several applications for climate model evaluation. We compare soil moisture data from satellite observations, reanalysis data and simulation results from a state-of-the-art climate model and analyze relationships between soil moisture and precipitation anomalies in the different datasets. In a detailed regional study, we show that ECVSM is capable to capture well interannual and intraannual soil moisture and precipitation dynamics in the Sahelian region. Current deficits of the new dataset are critically discussed and summarized at the end of the paper to provide guidance for an appropriate usage of the ECVSM dataset for climate studies.

1 Introduction

Soil moisture is an Essential Climate Variable (ECV) which has impact on regional to global terrestrial water, energy and carbon fluxes. Soil moisture controls the partitioning of the available energy into latent and sensible heat flux and conditions the amount of surface runoff. By controlling evapotranspiration, it is linking the energy, water and carbon fluxes (Koster et al., 2004; Seneviratne and Stöckli, 2008) and has a direct feedback on precipitation (Taylor et al., 2012a) as well as temperature (Miralles et al., 2012; Mueller and Seneviratne, 2012) at the regional to global scale.

An appropriate knowledge of soil moisture conditions is important for the initialization and quality of seasonal to yearly climate predictions. Fischer et al. (2007) indicated that the record breaking European heat wave in 2003 was enhanced by the large soil
moisture anomalies that were caused by a large precipitation deficit together with early vegetation green-up in the months preceding the extreme summer event. Loew et al. (2009) showed that these soil moisture anomalies were observable by remote sensing and Miralles et al. (2012) started to use satellite soil moisture anomalies to explain temperature extremes. All these studies indicate that due to its longterm memory, soil moisture can be an important factor for seasonal climate forecasts (Fischer et al., 2007).

Soil moisture shows a high variability from daily to interannual time scales and is therefore difficult to measure. Up to now, soil moisture observations are based on either in situ measurements (Robock et al., 2000; Dorigo et al., 2011) or satellite observations (Wagner et al., 1999; Njoku et al., 2003; Loew et al., 2006; Owe et al., 2008; Kerr et al., 2010). Several quasi-global soil moisture datasets have been developed during the last decade based on either active or passive microwave satellite observations from AMSR-E (Owe et al., 2008; Njoku et al., 2003), ERS AMI, MetOp ASCAT (Bartalis et al., 2007), SMMR, SSM/I, WindSat (Li et al., 2010; Parinussa et al., 2012) and SMOS (Kerr et al., 2010; Mecklenburg et al., 2012). Even though, except for SMOS, none of the used instruments has been specifically designed for soil moisture retrievals, the different products have been proven to show good correspondence with in situ observations over a large variety of environmental conditions (Draper et al., 2009; Gruhier et al., 2010; Brocca et al., 2011; Albergel et al., 2012; Dorigo et al., 2012b) and demonstrated their usefulness for applications like hydrological modeling (Beck et al., 2009; Brocca et al., 2012; Draper et al., 2012; Tramblay et al., 2012), numerical weather prediction (Drusch, 2007; Dharssi et al., 2011) or studies of land-atmosphere interactions (Seneviratne et al., 2010; Taylor et al., 2011, 2012a) as well as climate model evaluation and initialization (Schumann et al., 2009; Bisselink et al., 2011).

Recently, the first multidecadal satellite-based global soil moisture record (ECVSM) has become available (Liu et al., 2012). The new ECVSM data product has been generated by homogenizing different existing soil moisture products within the framework of
ESA’s Water Cycle Multi-Mission Observation Strategy (WACMOS) project (Su et al., 2010).

The purpose of the present study is to evaluate potential applications of this novel soil moisture data record for climate modelling applications based on the Max-Planck-Institute for Meteorology Earth System Model (MPI-ESM). The overarching objectives of the analysis in this study are to:

– Evaluate the potential of using ECVSM satellite soil moisture observations for climate model evaluation at regional to global scales.

– Analyze how ECVSM captures intra- and interannual soil moisture variability.

– Analyze how ECVSM can be used to study and evaluate soil moisture and precipitation relationship and dynamics in observations and models.

This paper is the first study that provides a comprehensive analysis of potentials and limitations of the novel multidecadal soil moisture dataset for climate studies and climate modeling applications. It compares soil moisture data from a state-of-the-art climate model, reanalysis as well as satellite observations.

2 Data and models

2.1 Soil moisture data

2.1.1 Multidecadal satellite soil moisture observations (ECVSM)

The ECVSM product is the first ever multidecadal satellite based soil moisture product and is available for the time period 1978–2010 on a daily basis and at a spatial resolution of 0.25°. It has been generated by merging active and passive microwave based soil moisture products (Wagner et al., 1999; Naeimi et al., 2009; Owe et al., 2008) from the following satellite instruments: SMMR (November 1978–August 1987), SSM/I
(July 1987–2007), TMI (1998–2008), AMSR-E (July 2002–December 2010), ERS-1/2 (July 1991–May 2006), and ASCAT (2007–2010). The data harmonization procedure is described in Liu et al. (2012) and is based on a rescaling of the remote sensing soil moisture data using the soil moisture statistics from the Noah-land surface model (Liu et al., 2011, 2012). A cumulative distribution function (CDF) matching technique is employed for that purpose. The CDF matching is applied on each grid cell individually and rescales the satellite observed soil moisture to the Noah land surface model statistics. As a consequence, the soil moisture statistics (percentiles) of the blended ECVSM soil moisture product resembles the soil moisture dynamics of the Noah land surface model. While this rescaling affects the percentile distribution of the ECVSM product, the temporal structures (e.g. autocorrelation, trends) are not changed by this rescaling approach (Liu et al., 2012).

Dorigo et al. (2012b) provide a comprehensive validation of ECVSM using 932 in situ observation sites from 29 different observing networks (Dorigo et al., 2011, 2013). Despite the large difficulties in validating coarse resolution satellite soil moisture products with in situ point like observations (Crow et al., 2012), Dorigo et al. (2012b) conclude that the ECVSM product has an average unbiased RMSE of 0.05 [m$^3$ m$^{-3}$]. Besides it was shown that trends in ECVSM largely agree with those in various reanalysis products, precipitation, and vegetation vigor (Albergel et al., 2012; Dorigo et al., 2012a).

The ECVSM data set provides a multitude of quality flags and only reliable soil moisture estimates are preserved in the data product. Snow covered areas or frozen ground is typically masked as well as dense or heterogeneously vegetated areas with high optical depth that are not expected to provide reliable soil moisture estimates (Loew, 2008; Parinussa et al., 2011). A pre-processing of the ECVSM data product is required to match it in space and time with the other datasets used in the present study. Thus, the ECVSM soil moisture data is regridded to the MPI-ESM T63 model grid ($\approx$ 1.85°).

A conservative remapping technique is used for that purpose which is implemented in the Climate Data Operators (cdo, https://code.zmaw.de/projects/cdo). Some data gaps were removed by this procedure due to spatial interpolation.
2.1.2 MPI-ESM land surface model (JSBACH)

MPI-ESM couples processes in the atmosphere, ocean and land surface through the exchange of momentum, water and important trace gases such as carbon dioxide. It has been widely used for comparative model calculations in the context of the Coupled Model Intercomparison Project 5 (CMIP5) (Taylor et al., 2012b).

The land surface component of MPI-ESM, JSBACH (Raddatz et al., 2007; Brovkin et al., 2009; Reick et al., 2013), is implicitly coupled to the atmospheric component of MPI-ESM, ECHAM6 (Stevens et al., 2013) and simulates all relevant land surface water, energy and carbon fluxes. The present study uses version 2.03 of JSBACH which is comparable to the model version which was used for CMIP5 experiments, but which includes a layered soil hydrology scheme instead of a standard bucket scheme (Hagemann and Stacke, 2013). In this version, the soil moisture dynamics in the unsaturated zone are simulated using five discretized soil layers with a thickness of $d_z = [0.065, 0.254, 0.913, 2.902, 5.7]$ [m]. A validation of global scale energy and water flux components of MPI-ESM CMIP5 simulations is given in Hagemann et al. (2013) and Brovkin et al. (2013).

JSBACH model simulations can be either performed by forcing the land surface scheme with any kind of meteorological forcing data (e.g. station measurements, re-analysis data) or by coupling JSBACH directly to a Global Circulation Model (GCM), like ECHAM6.

For the present study, model simulations were conducted for a 31-yr period (1979–2009) using an offline forcing dataset. This allows for a more realistic comparison of satellite soil moisture observations with the model simulations as it minimizes the effect of precipitation errors. MPI-ESM model simulations are performed on a Gaussian model grid with rather coarse spatial resolution (T63 $\approx$ 1.85 degree).

The model forcing data used in the present study is based on a methodology created in the EU WATCH project (http://www.eu-watch.org). The Watch Forcing Data (WFD) is documented in Weedon et al. (2011) and includes a wide range of meteorological
forcing variables like surface temperature, precipitation, radiation fluxes and wind speed for the period 1958–2001. The data set is available at 0.5 degree spatial resolution and sub-daily time scales.

The WFD is based on ERA40 reanalysis data (Uppala et al., 2005) and extensive post-processing was conducted in which the data were interpolated from the original Gaussian 1.125 degree grid to a rectangular 0.5 degree resolution grid. An elevation correction was applied for most variables. Furthermore, rainfall and snowfall were subject to extensive corrections to correct for biases in the reanalysis data. An undercatch correction was applied for precipitation to ensure that the monthly statistics are similar to in situ observations of the Global Precipitation Climatology Centre (Schneider et al., 2008) while the daily variability of ERA40 is retained. The WFD were validated against observations from seven FLUXNET sites. Weedon et al. (2010) and Weedon et al. (2011) demonstrated that most variables in the WFD are similar to the FLUXNET data. Daily averaged precipitation rates, which are difficult to compare between site measurements and half degree grid cell averages, agree well for at least three of the sites. The forcing data used in the present study (WFDEI) was generated applying the WFD methodology to the ERA-interim reanalysis (Dee et al., 2011) data (Weedon et al., 2013). The WFDEI dataset covers the period 1979–2009 so far.

2.1.3 ERA-interim soil moisture

ERA-interim is the latest global reanalysis of the European Centre for Medium Range Weather Forecasts (ECMWF). It covers the period from 1979 until present and is based on a variational data assimilation system that assimilates a multitude of in situ and satellite observations in a consistent framework (Dee et al., 2011).

Soil moisture is a prognostic variable in ERA-interim and is provided for four soil layers with a thickness of 0.07, 0.21, 0.72 and 1.89 [m], respectively. The ERA-interim soil moisture data was extracted from the ERA-interim data archive for the period 1979–2009 and regridded to the same spatial grid as MPI-ESM using conservative remapping. As ERA-interim data is available every 6-h, daily mean soil moisture fields were
calculated by averaging data from each day and monthly means were calculated subsequently from the daily means.

2.2 Precipitation data

The precipitation data used in the present study comprise different data sources. Precipitation information from a satellite based product as well as reanalysis data and a bias corrected reanalysis product are used for comparison of precipitation with soil moisture dynamics. The satellite based product and bias corrected reanalysis dataset use a common in situ dataset to correct for monthly biases in the precipitation record. Figure 1 shows the annual means and coefficient of variation of the precipitation rate of the different data products used. All show similar mean spatial precipitation fields, while the global mean values and coefficient of variation differ among the different datasets.

2.2.1 GPCP

The Global Precipitation Climatology Project (GPCP, v2.2) data product is based on satellite observations (Adler et al., 2003, 2011). The monthly product used in this study has a spatial resolution of 2.5 x 2.5 degrees and provides data since 1979. It is based on a blended gauge-satellite product which combines precipitation retrievals from polar-orbiting passive microwave imagers (SSM/I) as well as geostationary observations (IR data). The satellite retrievals are further bias corrected using rain gauge data from the Global Precipitation Climatology Centre’s (GPCC) Monitoring Product (Schneider et al., 2008).

2.2.2 ERA-interim precipitation

The ERA-interim precipitation is produced by the ERA-interim forecasting model based on temperature and humidity information as derived by the assimilation of atmospheric and terrestrial observations. Total precipitation estimates are only available for forecasting timesteps at 00:00 and 12:00 UTC. The 12-h segment following each forecast
step is used to obtain daily estimates of the rainfall rate by integrating all forecast steps within the timeperiods 00:00–12:00 UTC and 12:00–24:00 UTC. This sampling approach is similar to that of Dee et al. (2011). In general, ERA-interim overestimates precipitation especially in tropical regions compared to GPCP (Dee et al., 2011). As reanalysis precipitation is model generated, it also suffers from biases in annual or seasonal means over different regions (Lorenz and Kunstmann, 2012). The constraints imposed on precipitation by the data assimilation of other variables leads to reasonable variabilities, especially on a day-to-day basis, which is, e.g. utilized in the WFD. The usage of different observing systems in the data assimilation of re-analyses may lead to different biases or spurious trends (Hagemann et al., 2005) and should therefore be interpreted very carefully. The original ERA-interim spatial resolution (∼0.7 degree) was resampled to the T63 grid of MPI-ESM (∼1.85 degree) for further analysis.

### 2.2.3 WFDEI precipitation

The WFDEI data is based on ERA-interim reanalysis data which was corrected using in situ gauge information. It thus mediates between ERA-interim and GPCP datasets which are solely based on observations or reanalysis data.

WFDEI combines ERA-interim reanalysis rainfall data with the version 4 of the Global Precipitation Climatology Centre (GPCC) rain gauge product (Schneider et al., 2008). After a gauge undercatch correction and an adjustment of wet days based on the CRU TS2.1 observations (New et al., 1999, 2000; Mitchell and Jones, 2005), the GPCCv4 product is used to correct for monthly biases in the reanalysis precipitation data while the temporal dynamics is preserved from the reanalysis fields. Thus, the WFDEI accounts for systematic known weaknesses in the reanalysis datasets like an overestimation of wet days in the tropics (Uppala et al., 2005).
3 Methods

The data in the present study corresponds to geospatial data which can be represented by a matrix $X_{m \times n} = [x_1 \ldots x_n]$ whereas $m$ corresponds to the number of timesteps and $n$ to the number of grid cells. Each column vector $x_k, k \in 1 \ldots n$, corresponds to a timeseries of soil moisture or precipitation. Monthly means are calculated for $x_k$ from all valid samples.

3.1 Definitions

3.1.1 Calculation of anomalies

Anomalies of precipitation and surface soil moisture are calculated by removing the mean seasonality from the monthly mean timeseries $x_{l,j}$ whereas $l \in [1,2,\ldots,12]$ is an index for the month and $j$ is an index for the year. The anomaly timeseries $x'_{l,j}$ is given by

$$x'_{l,j} = x_{l,j} - \bar{x}_{l,j} = x_{l,j} - \frac{1}{n} \sum_{k} x_{l,k}$$

(1)

whereas $n$ is the number of years used to calculate the anomaly timeseries. Note that a linear detrending of the timeseries will be applied in some cases prior to calculating the monthly anomaly timeseries, like will be discussed in the manuscript.

3.1.2 Correlation and partial correlation analysis

The Pearson product-moment correlation coefficient is used as a measure for linear correlation between the different soil moisture and precipitation datasets. The correlation coefficient between two variables of size $m$ is calculated as
Precipitation is the major forcing for soil moisture variability. Soil moisture dynamics is however also affected by, e.g. soil hydraulic properties, vegetation or evapotranspiration. Partial correlation analysis is therefore used in addition to the general linear correlation analysis to analyze the general skill of ECVSM to capture the soil moisture dynamics. Partial correlation corresponds to the correlation of two datasets \((x, y)\) where the effect of an additional controlling variable \((z)\) has been removed. Formally the partial correlation coefficient between variables \(x\) and \(y\) removing the effect of the controlling variable \(z\) is given by

\[
\rho_{xy|z} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1 - \rho_{xz}^2} \sqrt{1 - \rho_{zy}^2}}
\]

Partial correlation analysis is used in the present study to investigate the relationship between two soil moisture datasets under the condition that the variability due to the rainfall dynamics has been removed before comparing the soil moisture datasets. This gives additional insight into the capabilities of ECVSM to represent soil moisture dynamics independent of precipitation dynamics.

### 3.1.3 Percentile correlation

The percentile distribution of soil moisture gives insight into the spatial and temporal patterns of soil moisture dynamics as represented in either the simulated or observed soil moisture fields. The advantage of comparing two soil moisture datasets by their percentile distribution is that it is independent from the absolute soil moisture values, but addresses only the similarity of relative soil moisture dynamics.
The $p$-th percentiles $\rho(p)$ of a dataset with $p \in [0.05, 0.1, \ldots, 0.95]$ are calculated from the timeseries $x$ for each grid cell which results in a spatial map, stored in a vector of size $n$ for each $\rho(p)$. The similarity between the percentile maps of the $p$-th percentile of two variables is then calculated as the linear correlation $\rho(\rho_x(p), \rho_y(p))$.

3.2 An example application of ECVSM for regional climate studies

The capability to capture significant climate anomalies is a crucial property of an ECV data record. It has been shown that satellite based surface soil moisture records are capable to capture well regional drought and flooding events like, e.g. the year 2011 droughts in Australia and in the Horn of Africa and the Southern US (de Jeu et al., 2012).

The devastating drought in the Sahel belt in Africa in the last half of the 20th century has been one of the largest and longest climate anomalies observed so far during the satellite era. The negative precipitation anomalies started in the 1960ies with a minimum around 1980 (Fig. 2). Since this minimum, the rainfall recovered and it has been shown in various studies that the vegetation in the Sahel recovered subsequently (Olsson et al., 2005; Hickler et al., 2005; Fensholt et al., 2012).

Land surface-atmosphere feedbacks are likely to have enhanced the Sahelian drought (Charney et al., 1977; Zeng, 1999) and it has been shown that soil moisture patterns affect the convective precipitation in this region Taylor et al. (2011).

The ECVSM data set is the first ever available multidecadal observation-based soil moisture data product that allows to investigate the relationship between soil moisture, precipitation and vegetation dynamics in the Sahel for over three decades and to compare the observations against model simulations.

The present study investigates how the Sahelian drought event is captured by ECVSM and the other soil moisture and precipitation datasets. It will be evaluated if the ECVSM dataset provides suitable information to support regional climate studies on interannual to multidecadal timescales in the Sahel.
In this study, the Sahelian belt is subdivided in five subregions (West = W, Central = C, East = E) to capture different precipitation regimes in the region (Fig. 3). The longitudinal division follows Huber and Fensholt (2011) and Lebel and Ali (2009).

4 Results

4.1 Spatiotemporal data coverage

The ECVSM data set has data gaps which are due to varying temporal and spatial coverage of the observations as well as different quality of the input data. Especially in the first decade of the ECVSM dataset, a large number of data gaps occurred due to the poor spatial coverage of the satellite instruments (Fig. 4). In the first pentad of the 1980ies, nearly 50 % of the year was without any data coverage which was mainly due to the small swath width of the used satellite instruments (Nimbus-SMMR) and a reduction of imaging capabilities due to power constraints of the satellite. This explains why a large portion of the globe contains a large fraction of data gaps (Fig. 5). In Northern latitudes, the fraction of missing data exceeds 80 % of all days of the period 1978–2009. Figure 5 shows further the fraction of missing data for the ESA ECVSM dataset in different preprocessing steps. The raw ECVSM data contains on average a data gap fraction of 73 % (±17 %); brackets indicate standard deviations. The remapping of the dataset to T63 resolution decreases the fraction of data gaps to 60 % (±16 %). For many applications, timeseries without gaps are required which might be achieved, e.g. through temporal smoothing. Figure 5 shows also the effect of a 5-day running mean temporal smoothing filter on the data coverage which is significantly improved (missing data: 30 ± 23 %).
4.2 Comparison of surface soil moisture dynamics

4.2.1 Global mean fields

The global mean soil moisture field for the years 1979–2009 and its coefficient of variation (CV) are shown in Fig. 6. The time-latitude diagrams of zonal mean soil moisture are shown in Fig. 7 for the whole period. ECVSM and JSBACH simulations have mean soil moisture fields and variability that are closer to each other than to ERA-interim which is in general wetter than JSBACH simulations and also shows much smaller temporal variability. The coefficient of variation is almost half as large (CV$_{\text{ERA}} \approx 0.12$) as for JSBACH (CV$_{\text{JSB}} \approx 0.24$) and ECVSM (CV$_{\text{ECVSM}} \approx 0.2$).

Figure 7 shows distinct differences in the spatiotemporal zonal mean soil moisture fields and its anomalies. The biases between the different datasets are clearly observable. JSBACH shows a more pronounced seasonal cycle in the subtropical areas while ECVSM and ERA-interim show a smaller variability in these areas.

While the simulated soil moisture fields are homogeneous over time, the ECVSM dataset clearly shows inconsistencies in the timeseries. A drying of the zonal mean soil moisture is for instance observed in 2002. Further discontinuities can be observed in 1987 and 2006. These discontinuities are also observable in the anomaly plots which show in general much stronger amplitudes for ECVSM than for ERA-interim or JSBACH.

As discontinuities in ECV time series can easily introduce artificial trends in timeseries that could be misinterpreted as a climate signal, a careful analysis of such kind of timeseries is important (Loew and Govaerts, 2010). In particular, a much more thorough analysis of the causes and effects of the observed discontinuities on data analysis is needed, which is beyond the scope of the present study. It is however emphasized that a careful treatment of trends derived from ECVSM is necessary.
4.2.2 Correlation between soil moisture fields

Figure 8 shows the correlation between the absolute monthly mean surface soil moisture of different datasets and the correlation between their anomalies. High significant correlations are observed between all datasets which indicates in general a good skill of all datasets to represent the interannual and seasonal surface soil moisture variability in a consistent way. Larger differences are observed in the Northern Latitudes caused by the poor data coverage in ECVSM due to snow and frozen soil conditions. Negative correlations are here especially observed between ECVSM and both models. While both models show also significant positive correlations in the tropical areas with dense vegetation like the Amazon or Congo basin, the ECVSM does not show any significant correlations in these areas with either of the model data sets. This was expected, as the remote sensing signal is perturbed by dense vegetation and the microwave signal is lacking any soil moisture information in these areas (Jeu et al., 2008; Dorigo et al., 2010). The global mean correlations between soil moisture anomalies of ECVSM and JSBACH (ERA-interim) are $\rho = 0.41 \pm 0.2$ ($\rho = 0.36 \pm 0.19$) while the anomaly correlation between the model simulations is higher ($\rho = 0.64 \pm 0.2$). Thus, in areas where the microwave signal is in general sensitive to soil moisture dynamics, the ECVSM dataset shows reasonable agreement with the two simulated soil moisture datasets for both, absolute values as well as for the soil moisture anomalies. A more detailed comparison of the soil moisture statistics of the different datasets will be made in the following by analyzing the soil moisture percentile distribution.

4.2.3 Percentiles of soil moisture dynamics

The percentiles of the soil moisture were calculated from the time series of each dataset for each grid cell. Figure 9 shows the 5, 50 and 95% percentiles maps for ECVSM, ERA-interim and JSBACH. The 5 and 95% values correspond to the lower (dry) and upper (wet) limits of the soil moisture dynamics. The similarity between the spatial patterns of each percentile was compared by calculating the correlation
coefficient $\rho$ between the percentile maps of two datasets. Results of this correlation analysis are summarized in Fig. 10 for the different percentiles. Highest correlations ($\rho \approx 0.75$) were found between ECVSM and JSBACH. The correlations between ERA-interim and JSBACH are lower ($\rho \approx 0.6$) and the relationship between ERA-interim and ECVSM shows the weakest spatial correlations of the percentiles ($\rho \approx 0.4$).

As JSBACH soil moisture and ECVSM soil moisture percentiles are completely independent, the results indicate that the model soil moisture dynamics seems to be in good agreement with the observed soil moisture dynamics in ECVSM. ERA-interim deviates stronger from ECVSM observations, but is in closer agreement with JSBACH simulations. A potential reason for the different skills in reproducing the observed ECVSM soil moisture patterns by the models might be related to the precipitation forcing used. While the JSBACH experiments are based on offline simulations using the WFDEI-forcing data, the ERA-interim precipitation results from coupled land-atmosphere simulations. ERA-interim precipitation biases have been reported in the literature. In particular, ERA-interim shows wet biases for the greater part of the Northern Hemisphere and in parts of South America (Dee et al., 2011). Note that the ERA-Interim surface water balance is not necessarily closed as soil moisture nudging is conducted by the data assimilation system that may lead to sources and sinks in the surface water balance, and, thus to less consistency of soil moisture time series (Hagemann et al., 2005).

4.2.4 Relationship of soil moisture and precipitation dynamics

The relationship between precipitation and soil moisture dynamics was analyzed by comparing GPCP, ERA-interim and WFDEI monthly precipitation data with the surface soil moisture data, assuming that the precipitation dynamics is the major driver for the surface soil moisture dynamics on monthly timescales. The effect of evaporation is therefore not explicitly considered in the analysis, but the effect of evapotranspiration is however reflected in the soil moisture dynamics of ECVSM as well as the model based soil moisture fields.
Figure 11 shows the correlation coefficients between monthly precipitation and surface soil moisture anomalies. A corresponding figure which is based on the actual soil moisture and precipitation data instead of the anomaly values is given in the Appendix (Fig. A1). ERA-interim and JSBACH show both high anomaly correlations with all precipitation datasets. Highest anomaly correlations are observed for ERA-interim with ERA-interim precipitation data and for JSBACH with WFDEI precipitation data, like would have been expected as these are the respective precipitation forcing datasets used for the generation of the soil moisture datasets. The anomaly correlation of ERA-interim with the ERA-interim precipitation data is highest ($\rho > 0.8$) in tropical areas. In Northern latitudes, the ERA-interim soil moisture dynamics seems to be decoupled from the actual precipitation dynamics, as significant negative correlations with all precipitation datasets are found in boreal latitudes (Fig. A1). The correlation patterns of the soil moisture datasets with either GPCP or WFDEI precipitation are very similar for all soil moisture datasets. This is explained by the bias correction of GPCP and WFDEI, which is applied on monthly time scales using the same GPCC observations (see Sect. 2.2). The monthly mean (anomalies) are therefore correlated with each other.

The ECVSM soil moisture shows coherent anomaly correlation patterns with all precipitation datasets. In general, the correlations are lower than between the simulated soil moisture fields and precipitation data. However, significant positive correlations ($\rho \approx 0.3$) between ECVSM and precipitation anomalies are found for the entire globe, which indicates a general skill of the ECVSM dataset in representing intra- and inter-annual soil moisture variability.

### 4.2.5 Partial correlation results

It has been demonstrated that the different soil moisture datasets show a significant correlation with different precipitation data as well as between each other. A correlation between simulated soil moisture and ECVSM might be however spurious as both might be dependent on a common precipitation anomalies.
For model evaluation purposes it is however of particular interest whether the land surface model is capable of simulating the observed anomalies of a geophysical variable. This requires to remove the common forcing effects, precipitation in this case, in two soil moisture datasets.

The simulated soil moisture fields were correlated against ECVSM using partial correlations where the effect of the precipitation forcing was removed (control variable). For ERA-interim, the ERA-interim precipitation was removed while WFDEI precipitation was removed for JSBACH simulations. As the true precipitation is unknown, the GPCP dataset is assumed to be the precipitation dataset which best captures the temporal and spatial precipitation dynamics. It is therefore used as a control variable on the ECVSM soil moisture fields for the partial correlation analysis.

Figure 12 shows partial correlations results which are based on soil moisture anomaly timeseries. The corresponding figure for the partial correlation of the original data is given in the Annex (Fig. A2). The partial correlation coefficients of ERA-interim and JSBACH are very similar. Significant partial correlations between the simulated and observed soil moisture data are especially observed in semiarid regions. In general, JSBACH shows higher partial correlation coefficients for the soil moisture data than ERA-interim.

The partial correlation coefficients are smaller for the anomalies than for the original data and JSBACH has higher correlation coefficients with ECVSM than ERA-interim in most areas. Highest partial correlation coefficients are observed in regions which are not affected by snow or by dense vegetation. These areas correspond to regions where the satellite observations are most sensitive to soil moisture variability and where the evapotranspiration is largely affected by soil moisture limitations. Largest differences between the partial correlation coefficients are observed in the Sahelian belt in Africa, where JSBACH has partial correlation coefficients between 0.2 and 0.6 whereas ERA-interim does not show any correlation in this region.

The fact that both datasets show correlations between simulated and observed surface soil moisture after removal of the effect of precipitation is an indication that the
ECVSM as well as the model simulations capture a similar intra- and interannual soil moisture variability, independent from rainfall. The partial correlation analysis might be therefore considered as a diagnostic for the similarity of ECVSM and model simulated interannual soil moisture dynamics.

4.3 Sahelian drought and interannual soil moisture variability

The precipitation in the Sahelian Region recovered since its minimum in the mid of the 1980ies. An increase in vegetation and precipitation was observed since then (Olsson et al., 2005; Ali and Lebel, 2009; Huber et al., 2011; Fensholt and Proud, 2012). Figure 13 shows linear trends of monthly mean precipitation and soil moisture for the period 1979–2009. A clear increase in precipitation in the Sahelian rainfall season (JJAS) is observed in the GPCP dataset which is consistent with the literature. The WFDEI precipitation data shows also a positive precipitation trend, but the areas with a significant trend are smaller. Contrary, ERA-interim shows a significant negative trend in the precipitation timeseries which contradicts the in situ ground observations from GPCC which are included in both the WFDEI data as well as GPCP precipitation estimates.

Figure 14 shows the spatially mean timeseries of precipitation in the Sahel as derived from all datasets. The WFDEI and GPCP datasets show a very similar temporal evolution with a positive trend, whereas the longterm trend in the WFDEI precipitation data is less significant than for GPCP. The soil moisture in ERA-interim and ECVSM shows significant negative soil moisture trends in the Sahel, while the JSBACH simulations do not show any significant change over the investigated timeperiod. It has been however shown by Loew (2013) that the significant positive trend in GPCP data is mainly caused by the strong negative precipitation anomalies in the beginning of the 1980ies and not significant in the years thereafter and that the longterm trend is not significant if a few years in the 1980ies are discarded from the analysis.

Dorigo et al. (2012a) found a significant negative trend of June-July-August surface soil moisture in the same region, as derived from the ECVSM dataset for the period

The ECVSM as well as the ERA-interim soil moisture show significant negative trends in wide parts of the Sahel (Fig. 13) while the JSBACH simulations show a more diverse picture. In the Central part a significant negative temporal trend is here also observed, while a dipole of significant trends is observed in the Western part with an increase in the North and a decrease in the South. The Eastern part shows no significant trends in the JSBACH simulations. The positive trend in precipitation and negative trend in surface soil moisture seem to be contradictory, but could be related to an increase in evapotranspiration by the increased abundance of vegetation in the area.

In the following, we will investigate the interannual and decadal variability of surface soil moisture as observed by ECVSM and as simulated by ERA-interim and JSBACH. The analysis will focus exclusively on the anomalies of soil moisture and precipitation to be independent of biases between the different datasets.

The interannual soil moisture and precipitation anomalies for the Sahel (20° W–45° E, 10° N–20° N) are shown in Fig. 15 in time-latitude diagrams. The ECVSM and especially ERA-interim surface soil moisture show a decline in the surface soil moisture content throughout the period and all latitudes, like has been already discussed, while the JSBACH soil moisture shows no clear temporal trend. The ERA-interim timeseries shows some discontinuities. A drier period (1983–1995) is followed abruptly by a wetter period which lasts until approx. 2001. A further drier period follows after 2006.

The precipitation anomalies between GPCP and WFD are consistent with each other as would have been expected, while ERA-interim precipitation shows a pronounced decline throughout the period 1979–2009.

The ERA-interim and ECVSM soil moisture anomalies show linear temporal trend of soil moisture for the entire time period. Linear detrended anomaly timeseries were therefore used for the further analysis to focus the analysis exclusively on interannual soil moisture anomalies and to avoid any suspicious correlations due to common longterm trends. Figure A3 displays the time-latitude diagrams for soil moisture.
and precipitation anomalies, where the linear longterm trend has been removed. The ECVSM and JSBACH surface soil moisture shows very consistent anomaly patterns with precipitation if the linear trends are removed from the data. The ECVSM data reproduces very well the dry and wet anomalies that are observed in the GPCP record.

To analyze the relationship between soil moisture and precipitation anomalies, the correlation between the anomaly timeseries was calculated for different regions (Fig. 3). The anomaly correlation coefficients between precipitation and soil moisture are summarized in Fig. 16. Very similar partial correlation coefficients are found for GPCP and WFDEI as both use the same in situ observations from GPCC on monthly timescales to compensate for biases in their data products. JSBACH soil moisture anomalies show high correlations ($r > 0.6$) for all precipitation datasets, while ERA-interim soil moisture anomalies are best correlated with ERA-interim precipitation data. The ECVSM anomalies show highest correlations with WFDEI/GPCP precipitation anomalies in all regions.

As the soil moisture anomalies are likely to be highly correlated with the precipitation anomalies, a partial correlation analysis was conducted to quantify how the ECVSM observations are related to the simulated surface soil moisture dynamics when the effect of the precipitation boundary condition is removed. It has been shown that the ERA-interim and ECV soil moisture both show a longterm decline of soil moisture in the analysis period. This common linear trend can introduce a significant correlation in the partial correlation analysis. The partial correlation was therefore conducted twice, once based on soil moisture and precipitation anomalies and second on anomalies which have been detrended prior to the analysis (Fig. 17).

In all regions, the ERA-interim and JSBACH soil moisture anomalies show a positive correlation with the ECVSM anomalies. In the Western Sahel (W1, W2), ERA-interim and JSBACH show these positive correlations with ECVSM also when the data was detrended before the analysis. Only JSBACH shows a positive correlation with the ECVSM soil moisture anomalies in the central and eastern Sahel when the linear trend in the timeseries has been removed. This clearly indicates that the partial correlation
of ERA-interim anomalies is largely dependent on the longterm linear trend in the dataset and can not be interpreted as a common interannual soil moisture dynamics with ECVSM. Contrary, negative correlations between ERA-interim and ECVSM soil moisture are observed for the Eastern and Central regions which indicates that the ERA-interim soil moisture dynamics is highly dependent on the ERA-interim precipitation data in these regions.

5 Summary and conclusions

The objectives of the present study were to assess the potential and limitations of the novel ECVSM soil moisture dataset for climate modeling applications. The identified potentials and limitations are briefly summarized in Table 1.

The analysis in the present paper focused on a limited set of potential application of ECVSM for climate model evaluation, general study of soil moisture-precipitation interdependencies as well as on the applicability of ECVSM to capture interannual soil moisture dynamics and anomalies at the regional scale. All analyses were done on monthly time scales.

The ECVSM dataset is unique, as it is the first and only existing observation-based soil moisture data record for multiple decades. The analysis has shown that the present dataset is generally in good agreement with other soil moisture datasets from modeling studies as well as rainfall data.

In areas where the microwave signal is in general sensitive to soil moisture dynamics, the ECVSM dataset shows reasonable agreement with the ERA-interim and JSBACH soil moisture datasets for absolute values as well as for the soil moisture anomalies. Using percentiles distributions has been found to be a useful approach to evaluate the general soil moisture dynamics of a land surface scheme used in a climate model. The JSBACH soil moisture fields have shown much higher spatial similarities with the ECVSM observations than the ERA-interim soil moisture field. These differences might be partly attributed to the differences in the precipitation forcing as is indicated by

3562
a partial correlation analysis which revealed that the ECVSM soil moisture anomalies show comparable correlation patterns with ERA-interim and JSBACH soil moisture anomalies after removal of the influence of the precipitation forcing.

While the percentile distribution is in general a very useful method to evaluate the general soil moisture dynamics of a model from observational data, it needs to be emphasized that the soil moisture dynamics in the ECVSM dataset is not purely observational based. The ECVSM final product was generated by harmonizing a multitude of soil moisture products using the Noah land surface model output from the Global Land Data Assimilation System (GLDAS) as a common scaling reference (Liu et al., 2011). This implies that the soil moisture statistics represented in ECVSM depends on the soil moisture dynamics of the Noah land surface model. ECVSM can therefore not provide an independent data source for the statistics of soil moisture dynamics at a particular location.

It is however emphasized that the temporal dynamic in the dataset is not affected by the normalization procedure. Thus, comparing temporal soil moisture and precipitation anomalies provides additional insight in the temporal soil moisture dynamics as represented by ECVSM. It was demonstrated in the present study that the ECVSM dataset shows in general good anomaly correlations with different global precipitation products. Additionally, the ECVSM dataset captures well the intra- and interannual soil moisture variability and has also skill to represent soil moisture dynamics independent from the precipitation forcing. The highest skill in representing soil moisture dynamics was observed in areas that are not affected by dense vegetation or snow and ice. As these areas could be clearly identified from the partial correlation analysis it can be concluded that partial correlation can be used as an indirect validation of the sensitivity of ECVSM to soil moisture variability.

In high latitudes, the data density of ECVSM is limited due to snow cover and frozen ground conditions. Negative correlations between ECVSM and simulated soil moisture were observed in these high latitudes which are likely due to missing snow melt peaks.
in the first part of the ECVSM time series. Further detailed studies on the reasons for these negative correlations are required.

The ECVSM shows discontinuities in its time series which are especially recognized in zonal mean anomaly plots (Fig. 7). These temporal discontinuities are likely to be caused by a change in the observing system which affects the absolute soil moisture values as well as the temporal sampling of the data. Discontinuities are, e.g. observed in 2002 (integration of AMSR-E data), 1987 (change from SMM/R to SSM/I) or 2006 (inclusion of METOP ASCAT). The input data for ECVSM has been also rescaled prior to the generation of the longterm record. Different data rescaling was applied before and after 1987 for the passive microwave observations (see Liu et al., 2012, Sect. 3.1.2 for details). As a consequence of this scaling, some longterm trends in the timeseries might be minimized and any trend analysis performed on ECVSM needs to be interpreted critically. Dorigo et al. (2012a) therefore investigated global trends in surface soil moisture dynamics only after 1988. Loew (2013) has analyzed the importance of different periods on the estimation of longterm trends from satellite ECV records and shows that small changes in the investigation period might have a strong effect on the correlation results obtained. We therefore investigated the robustness of the results of the present study by comparing the results obtained from the whole data record (1979–2009) against results for the period 1987–2009. In general, results (not shown in the paper) from both periods show very similar spatial correlation patterns with slightly higher correlation values for the period 1987–2009, like would have been expected like discussed before.

It has been shown that the ECVSM data set is well suited to study regional climate phenomena on multidecadal timescales. The Sahelian rainfall dynamics and its representation in ECVSM were investigated as an example for the potential applications of ECVSM on the regional scale. It was shown that ECVSM has high correlations of surface soil moisture anomalies with the soil moisture anomalies of JSBACH as well as with observed precipitation anomalies. Partial correlation analysis revealed highest partial correlation coefficients between ECVSM and JSBACH which indicates that
both datasets show a comparable soil moisture residual after removal of the precipitation dynamics. It needs to be emphasized, that the temporal dynamics of ECVSM and JSBACH is completely independent as both are based on different data sources. Significant correlations between the datasets can therefore be considered as a common representation of soil moisture dynamics. ECVSM therefore provides suitable information to support regional climate studies in the Sahel. More detailed studies are however needed to better understand the different drivers of soil moisture dynamics in this region.

Overall, the ECVSM dataset provides a first unique dataset with relevant information for climate studies. A further potential application of ECVSM for climate model evaluation studies is the identification of characteristic time scales in land surface models. By calculating the autocorrelation length from both, models and observations, one can identify characteristic time scales. This information might be used, e.g. to infer relevant model specific soil parameterizations like characteristic timescales which are important for having a realistic soil moisture memory effect in the climate model land surface schemes which is of particular importance for seasonal climate predictions. Further studies will therefore focus on an evaluation of soil moisture memory effects from ECVSM and comparisons to the MPI-ESM land surface scheme.

Acknowledgements. This work was supported through the through the Cluster of Excellence CliSAP (EXC177), University of Hamburg, funded through the German Science Foundation (DFG) and the ESA Climate Change Initiative Climate Modelling User Group (contract: 4000100222/10/l-AM). Tobias Stacke acknowledges funding from the Federal Ministry of Education and Research in Germany (BMBF) through the research programme MiKli (FKZ: 01LP1108A) and Wouter Dorigo and Richard de Jeu are partly sponsored through ESA’s Climate Change Initiative for Soil Moisture. (contract 4000104814/11/I-NB). The WFDEI dataset was generated and provided through the EU FP6 WATCH project (contract number 036946). The GPCP combined precipitation data were provided by the NASA/Goddard Space Flight Center’s Laboratory for Atmospheres, which develops and computes the dataset as a contribution to the GEWEX Global Precipitation Climatology Project.
The service charges for this open access publication have been covered by the Max Planck Society.

References


3570


Schneider, U., Fuchs, T., and Rudolf, B.: Global Precipitation Analysis Products of the GPCC, 2008. 3547, 3548, 3549

3573
Table 1. Summary of potential and limitations of ECVSM for climate modelling applications.

<table>
<thead>
<tr>
<th>Potential</th>
<th>Current limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Climate model evaluation</strong></td>
<td>ECVSM soil moisture probability density function is dependent on Noah GLDAS and therefore does not provide a model independent dataset for soil moisture percentile distribution</td>
</tr>
<tr>
<td>Percentile statistics are useful tool for model evaluation</td>
<td>ECVSM shows discontinuities in the longterm timeseries which are likely to be the result from changing observing systems. Any conclusions from trend analysis therefore needs to be done very carefully.</td>
</tr>
<tr>
<td>ECVSM shows consistent temporal trend patterns with precipitation and ancillary soil moisture data</td>
<td>Time series not homogeneous due to different data rescaling approaches before and after 1988; see Liu et al. (2012), Sect. 3.1.2 for details</td>
</tr>
<tr>
<td><strong>Process and regional studies</strong></td>
<td>Limited data coverage in early years (&lt; 1990)</td>
</tr>
<tr>
<td>Good representation of intra- and interannual soil moisture anomalies at global to regional scale</td>
<td>Risk of regionally missing soil moisture information due to binary like blending technique applied; example: missing heat wave 2003 in Europe</td>
</tr>
<tr>
<td>Suitable soil moisture information for land-atmosphere interactions in the Sahel</td>
<td>High latitude limitation due to snow cover and frozen soil conditions</td>
</tr>
</tbody>
</table>
Fig. 1. Mean precipitation fields and their coefficient of variation (CV) for GPCP (top), ERA-interim (middle), WFDEI forcing data (bottom).
Fig. 2. Sahel rainfall anomalies for the period 1900–2011 and the region from 10° N–20° N and 20° W–10° E (Data from http://jisao.washington.edu/), Ali and Lebel (2009).
Fig. 3. Annual mean precipitation [mm yr$^{-1}$] in the Sahelian region. Bounding boxes for regions used for detailed analysis extend between 10$^\circ$ N and 20$^\circ$ N. Longitudinal extend is given as: W1 (20$^\circ$ W–10$^\circ$ W), W2 (10$^\circ$ W–00$^\circ$ W), C1 (00$^\circ$ W–15$^\circ$ E), C2 (15$^\circ$ E–25$^\circ$ E), E1 (25$^\circ$ E–40$^\circ$ E) (Data: GPCP).
Fig. 4. Number of days per year without any soil moisture observations globally in ECVSM.
Fig. 5. Fractional coverage of missing data for ECMSM in different pre-processing steps: RAW: data at original spatial and temporal resolution, T63: data remapped to T63 grid, SMOOTH: timeseries smoothed with 5-day running average kernel. Numbers in the figure titles indicate global mean and stdv.
Fig. 6. Mean surface soil moisture of different data sets (left) and coefficient of variation (CV) of soil moisture (right) for the period 1979–2009.
Fig. 7. Time-latitude diagrams for monthly mean surface soil moisture and surface soil moisture anomaly data sets [m$^3$ m$^{-3}$].
Fig. 8. Correlation coefficient for the linear correlation between the different surface soil moisture datasets (left) and their anomalies (right): ERA vs. ECVSM (top), ECVSM vs. JSBACH (middle), ERA vs. JSBACH (bottom). Only significant correlations ($p < 0.1$) are shown. Frequency distribution of correlation coefficients between different datasets is shown in the histogram (solid line: soil moisture data; dashed line: soil moisture anomalies).
Fig. 9. Soil moisture percentiles [m$^3$ m$^{-3}$] for ECVSM (left), ERA-interim (middle) and JSBACH (right) for three different percentiles of $p = [0.05, 0.5, 0.95]$ (top-down).
Fig. 10. Pearson product-moment correlation coefficients between the $p$-th percentiles of ERA-interim, JSBACH and ECVSM soil moisture fields.
Fig. 11. Significant correlations ($p < 0.1$) of monthly GPCP (top), ERA-interim (middle) and WFDPEI precipitation anomalies against ERA-interim (left), JSBACH (middle) and ECVSM (right). A similar plot based on the soil moisture and precipitation data instead of the anomalies is given in Fig. A1.
Fig. 12. Correlation coefficients of partial correlation between ERA-interim (left) and JSBACH (right) soil moisture anomalies against ECVSM soil moisture anomalies with removal of the effect of the respective precipitation forcing.
Fig. 13. Pearson product moment coefficient for temporal trend of precipitation (top) and soil moisture (bottom) for the Sahelian rainfall season (June–September) and years 1979–2009: precipitation (from left to right: GPCP, WFDEI, ERA-interim), soil moisture (from left to right: ECVSM, JSBACH, ERA-interim). Grid cells with significant correlations ($p < 0.1$) are indicated by stippled areas.
Fig. 14. Timeseries of mean Sahelian rainfall (top) and soil moisture (bottom) as aggregated over the area from 10° N to 20° N and 20° W to 40° E for the different datasets used in the study.
**Fig. 15.** Time-latitude diagrams for monthly anomalies of surface soil moisture \( [m^3 m^{-3}] \) and precipitation \([mm day^{-1}]\) in the Sahel region (20° W–45° E, 10° N–20° N). Anomalies are calculated by removing the mean seasonal climatology derived from the whole timeseries (1979–2009). A corresponding figure where the linear trend has been removed is given in Fig. A3.
Fig. 16. Correlation coefficient between precipitation and soil moisture anomalies in different regions in the Sahel: GPCP (top), WFD (middle), ERA-interim precipitation (bottom). Data was detrended before analysis.
Fig. 17. Partial correlation coefficients between ERA-interim and JSBACH soil moisture and ECVSM with prior removal of precipitation influence on the correlation coefficient.
**Fig. A1.** Significant correlations ($p < 0.1$) of monthly GPCP (top), ERA-interim (middle) and WFDPEI precipitation against ERA-interim (left), JSBACH (middle) and ECVSM (right). A similar plot based on the soil moisture and precipitation anomalies is given in Fig. 11.
Fig. A2. Correlation coefficients of partial correlation between ERA-interim (left) and JSBACH (right) soil moisture anomalies against ECVSM soil moisture anomalies with removal of the effect of the respective precipitation forcing (same as Fig. 12 but for soil moisture and precipitation data instead of anomalies).
Fig. A3. Same as Fig. 15, but linear longterm trend removed prior to anomaly calculation.