ESA CCI Soil Moisture GAPFILLED: an independent global gap-free satellite climate data record with uncertainty estimates

Wolfgang Preimesberger, Pietro Stradiotti, and Wouter Dorigo

Department of Geodesy and Geoinformation, TU Wien, Wiedner Hauptstraße 8, 1040 Vienna, Austria

Correspondence: Wolfgang Preimesberger (wolfgang.preimesberger@geo.tuwien.ac.at)

Abstract. The ESA CCI Soil Moisture multi-satellite climate data record is a widely used dataset for large-scale hydrological and climatological applications and studies. However, data gaps in the record can affect derived statistics such as long-term trends and – if not taken into account – can potentially lead to inaccurate conclusions. Here, we present a novel gap-free dataset, covering the period from January 1991 to December 2023. Our dataset distinguishes itself from other gap-filled products, as it is purely based on the available soil moisture (SM) measurements (independent of ancillary variables to make predictions), and further due to the inclusion of uncertainty estimates for all interpolated data points.

Our gap-filling framework is based on a well-established univariate discrete cosine transform with the penalized least-squares (DCT-PLS) algorithm. This ensures that the dataset remains fully independent of other soil moisture and biogeophysical datasets and eliminates the risk of introducing non-soil-moisture features from other variables. We apply DCT-PLS on a spatial moving window basis to predict missing data points based on temporal and regional neighbourhood information. The challenge of providing gap-free estimates during extended periods of frozen soils is addressed by applying a linear interpolation for these periods, which approximates the retention of frozen water in the soil. To quantify the inherent uncertainties in our predictions, we developed an uncertainty estimation model that considers the input observations quality and the performance of the gap-filling algorithm under different surface conditions. We evaluate our algorithm through performance metrics with independent in situ reference measurements and by its ability to restore GLDAS Noah reanalysis data in artificially introduced satellite-like gaps. We find that the gap-filled data perform comparably to the original observations in terms of correlation and unbiased root mean squared difference (ubRMSD) with in situ data (global median $R = 0.72$, ubRMSD $= 0.05\,\mathrm{m^3\,m^{-3}}$). However, in some complex environments with sparse observation coverage, performance is lower.

The new ESA CCI SM v09.1 GAPFILLED dataset is publicly available at https://doi.org/10.48436/hcm6n-t4m35 (Preimesberger et al., 2024) and will see yearly updates due to its inclusion in the operational ESA CCI SM production.

1 Introduction

Satellite soil moisture data are relevant for a wide range of applications, including water resource management, agriculture, disaster risk assessment and response, weather prediction, and climate monitoring (Dorigo et al., 2021a; Srivastava, 2017). Soil moisture (SM) is classified as an essential climate variable (ECV) (GCOS, 2022), and dedicated satellite missions such as ESA's SMOS (Kerr et al., 2010) or NASA's SMAP (Entekhabi et al., 2010) have been launched to measure it. However, the data coverage of individual satellites is limited by their lifespan and revisit frequency, requiring multiple days for a complete set of global measurements. Long-term analyses of climate variability and change require at least 30 years of data (GCOS, 2022), which surpasses the lifespan of any single Earth observation satellite. Measurements from multiple platforms are therefore harmonized and merged to create consistent multi-decadal records, such as the ESA CCI soil moisture (ESA CCI SM) products (Dorigo et al., 2017; Gruber et al., 2019; Preimesberger et al., 2021). Despite the high number of currently operational satellites integrated in ESA CCI SM, it is still impossible to provide global gap-free daily soil moisture because of several physical limitations to measure soil moisture, as discussed below. Leaving these data gaps unfilled can perturb derived statistics such as anomalies or long-term trends and, in the worst case, lead to incorrect conclusions (Bessenbacher et al., 2023; Liu et al., 2020c). Yet, observational datasets like ESA CCI SM are still valued in addition to gap-free reanalysis products because they can capture features that models might miss (van der Schalie et al., 2022). This creates a challenge as gap-free satellite data are often required, especially in various machine learning frameworks, such as those used for spatial downscaling (Kovačević et al., 2020; Nadeem et al., 2023), or for applications that require water balance closure (Dorigo et al., 2021a). In an attempt to fill these gaps, users unfamiliar with the underlying geophysical processes may resort to oversimplified approaches, such as filling missing values ("NaNs") with zeros, which risks introducing further bias into the analysis. Understanding the underlying causes of data gaps in ESA CCI SM is therefore a prerequisite for filling them.

As mentioned above, there are several reasons for data gaps in ESA CCI SM. Apart from missing satellite overpasses in earlier periods, data gaps originate from the underlying retrieval models to convert raw satellite measurements (radiances, backscatter) into soil moisture. These models become unreliable under certain surface conditions, leading to flagged/masked data points and, therefore, gaps in the final record. The most common cause are frozen soils (van der Vliet et al., 2020). Retrieval models are designed to estimate the amount of liquid water in the soil (Owe et al., 2008; Wagner et al., 1999; Entekhabi et al., 2010). The dielectric properties of water change drastically between liquid and frozen aggregate state (Naeimi et al., 2012). Consequently, measured soil moisture normally drops when part of the satellite footprint is frozen (Amankwah et al., 2021; Dorigo et al., 2021b). The second-most-common cause for unreliable retrievals is dense vegetation, which can mask contributions from the soil. This issue is more pronounced at shorter wavelengths (K- and X-band), while L- and C-band measurements are less affected (van der Schalie et al., 2021; Jackson et al., 1982). Other reasons for failed retrievals include signal perturbation due to radio frequency interference (RFI), mainly for C- and L-band and overly densely populated areas (Uranga et al., 2022; Oliva et al., 2012), as well as barren, dry soils, where the received signal is noise-like and sometimes dominated by responses from sub-surface sediments (Wagner et al., 2024; Petropoulos et al., 2015).

Data gaps can be broadly classified into two – not strictly separable – categories (based on Bessenbacher et al., 2022; Rubin, 1976; Dorigo et al., 2017): (i) (quasi-)random and (ii) systematic retrieval gaps. Random gaps appear noise-like over time and affect only short periods. They are not directly related to processes at the surface. Members of this category include cases where there is no satellite overpass for a given day, when the retrieval model does not converge for various reasons or when it converges at soil moisture values outside the physically possible boundaries, and RFI. Systematic gaps are related to the surface state and, therefore, usually cover longer periods. In the most extreme cases, systematic gaps can be permanent, meaning no data are available for a location (e.g. rainforests in ESA CCI SM). Systematic gaps are more likely to obscure spatial and/or temporal features in the data than random gaps. Gap-filling methods aim to restore these features as accurately as possible without introducing artificial (non-soil-moisture) patterns and without overfitting the data.

There are generally two approaches to fill gaps in satellite soil moisture records: (i) univariate, stand-alone interpolation methods and (ii) multivariate, covariate-enhanced interpolation methods. The former uses only the available (neighbourhood) information to predict missing values with (geo)statistical methods, while the latter relies on supporting data to improve predictions. For soil moisture, most recent gap-filling studies focus on the applicability of different multivariate machine learning (ML) methods in small to medium-sized study regions. Random forest (RF) has been the dominant algorithm (Nadeem et al., 2023; Liu et al., 2020c; Nadeem et al., 2023; Wang et al., 2023; Mao et al., 2019; Bessenbacher et al., 2022) and was found to outperform other covariate supported approaches such as neural networks (NNs), XGBoost, support vector machines (SVMs), or multivariate linear regression (Liu et al., 2023; Sun and Xu, 2021; Tong et al., 2021; Almendra-Martín et al., 2021). To predict missing soil moisture, these studies use physically related variables such as air or land surface temperature, precipitation, soil type, topography, land cover, and vegetation properties. While multivariate approaches can leverage additional datasets to improve predictions, univari-

ate methods such as ordinary kriging (OK) or multiple linear regression (MLR) do not. However, they remain viable alternatives. In homogeneous areas, they have often been found to perform similarly to multivariate methods (Tong et al., 2021; Sun and Xu, 2021; Yang et al., 2018).

In this study, we adopt a univariate approach to fill gaps in ESA CCI SM. The choice of using a univariate method is motivated by the fact that it allows for being totally independent of any ancillary data and therefore mitigating the risk of introducing spurious features. The greatest possible independence of ancillary or model data is also one of the top priorities expressed by the climate user community (Dorigo et al., 2017). Gap filling of ESA CCI SM is based on discrete cosine transform with penalized least squares (DCT-PLS) (Garcia, 2010) because (i) even without the use of any ancillary data, it often performs similarly to multivariate methods such as XGBoost, LSTM, ANN, or CNN (Shangguan et al., 2023; Yang et al., 2018). (ii) It is considered a well-proven method that has been successfully applied to gap fill soil moisture (Wang et al., 2012; Yang et al., 2018) and other geophysical variables such as land surface temperature (LST) (Liu et al., 2020a; Pham et al., 2019), ocean surface current (Kongkulsiri et al., 2018), chlorophyll-a concentration (Wang et al., 2022), and lake surface water temperature (Fan et al., 2022) data.

The multitude of gap-filling studies highlights the importance of this topic and the community's need for a gap-free ESA CCI SM data record. With the ESA CCI SM v09.1 GAPFILLED product, we aim to meet this community requirement by providing an independent, gap-filled, global, long-term, and satellite-based record. Consistently with other ESA CCI SM products, we also provide uncertainty estimates to quantify potential errors associated with the interpolation process – a feature which to our knowledge has not yet been included in any gap-filled satellite soil moisture record. This will allow users to account for the expected accuracy of our predictions and, for example, weigh them accordingly in their analyses.

2 Data

2.1 ESA CCI SM

ESA CCI SM is a multi-satellite climate data record of daily global surface soil moisture. The latest release, v09.1, covers the period from 1 November 1978 to 31 December 2023 (Dorigo et al., 2017; Gruber et al., 2019; Preimesberger et al., 2021; Dorigo et al., 2024a). The COMBINED product merges soil moisture derived from radiometer measurements in the L-, C-, X-, and K-bands by the Land Parameter Retrieval Model (LPRM) (Owe et al., 2008; van der Schalie et al., 2015, 2016; van der Schalie et al., 2021) and from C-band scatterometer observations using the TU Wien change detection method (H SAF, 2022; Wagner et al., 1999, 2013; Hahn et al., 2021). A timeline of all used satellite sensors

and frequency bands is shown in Fig. 1a. Soil moisture values are provided in volumetric units ($m^3 m^{-3}$). ESA CCI SM data include quality flags, which are applied to mask unreliable data points and inform users about the underlying cause. Currently, there are quality flags for (i) frozen soils, (ii) dense vegetation, (iii) unsuccessful retrievals in all sensors, (iv) low signal-to-noise ratio in all sensors, and (v) barren ground (van der Vliet et al., 2020; Parinussa et al., 2011). While flags (i)–(iv) coincide with missing data points in ESA CCI SM, flag (v) only tags the observations but does not remove them. ESA CCI SM data come with associated estimates of measurement uncertainties (Gruber et al., 2017, 2019). These uncertainties are propagated from sensor-level estimates through triple collocation analysis on a day-of-year basis (Stradiotti et al., 2025). ESA CCI SM represents moisture in the upper soil layer (~ 0–5 cm) with a regular spatial sampling of $0.25°$ (~ 25 km). The algorithm is regularly updated with new data from available sensors to incorporate the latest scientific advances and extend the product's temporal coverage. This study aims to fill data gaps in the period from 1 January 1991 onward, chosen to make the data suitable for long-term (multi-decadal) studies (GCOS, 2022). From 1978 to 1991, no overlapping sensors were available (Fig. 1a), resulting in very low data coverage (Fig. 1b) and high measurement uncertainties (Dorigo et al., 2017). As this would lead to even higher uncertainties after gap filling, this period is excluded.

2.2 Model and reanalysis data

We use daily mean top-layer (0–7 cm) soil temperature from ERA5 at a $0.25°$ resolution (Hersbach et al., 2020). In our study, these data are used to classify which gaps occurred during periods of frozen soils and which did not and to subsequently choose one of two different interpolation approaches. ERA5 is a global reanalysis product that integrates in situ and satellite observations into an open-loop run of the HTESSEL land surface model, producing hourly, gap-free estimates of various surface variables since 1940.

We further use ERA5-Land top-layer (0–7 cm) soil moisture from 1991 to 2023 for a comparison of zonal anomalies and long-term trends after gap filling (Muñoz Sabater et al., 2021). ERA5-Land has been produced by replaying the land component of ERA5 and provides surface variables with an increased spatial resolution ($0.1°$).

Additionally, as reference data to assess the impact of gap distribution and size on interpolation quality, we use daily averages of gap-free soil moisture simulations for the 0–10 cm layer (given in $kg m^{-2}$) from the GLDAS Noah v2.1 model (Rodell et al., 2004). The original data are available at a 3 h temporal sampling and $0.25°$ spatial resolution for the period after 2000. Notably, GLDAS Noah surface soil moisture serves as the scaling reference for the ESA CCI SM COMBINED product (Dorigo et al., 2017). Hence, there is no bias between the two.
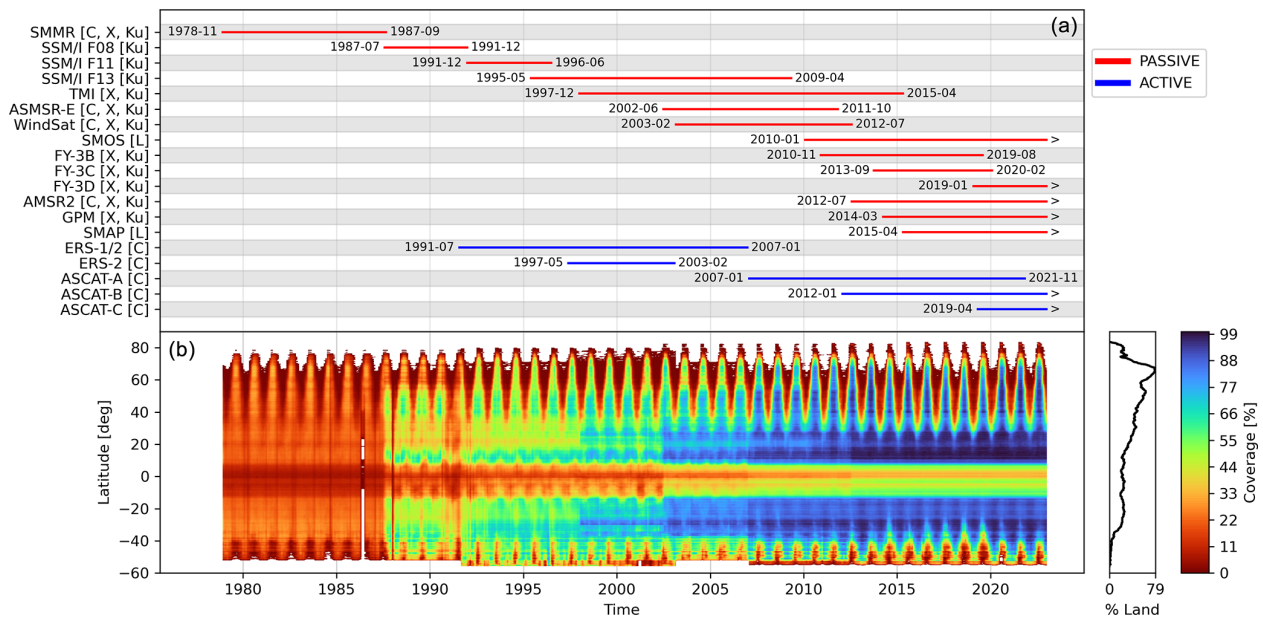
Figure 1. Availability of satellites and sensor frequency bands over time as of ESA CCI SM v09.1 **(a)** and data coverage over time of the merged COMBINED product averaged by latitude **(b)**.

2.3 Vegetation optical depth

We utilize vegetation optical depth (VOD) from the VODCA v2 CXKu dataset (Zotta et al., 2024; Moesinger et al., 2020) to predict the uncertainty in gap-filled values under different vegetation conditions. VOD is a satellite-derived, unitless measure of vegetation density closely related to vegetation water content and biomass. VODCA v2 CXKu harmonizes and merges VOD retrievals from nine passive sensors operating in the 6.8–19.4 GHz frequency range, creating a consistent long-term record from 1987 to 2023.

2.4 In situ measurements

The International Soil Moisture Network (ISMN) is a data hosting facility that collaborates with partners worldwide to collect in situ measurements of soil moisture and other geophysical variables into a harmonized, quality-controlled database (Dorigo et al., 2021b, 2013, 2011). ISMN in situ measurements serves as the primary source of ground reference data for global satellite soil moisture validation activities. An overview is provided by Dorigo et al. (2021b). In our study, we use ISMN data to evaluate the quality of our gap-filled soil moisture product. We use a subset of the full ISMN database for the 0–10 cm depth range, downloaded in March 2024. Specifically, we use a subset of the available data labelled as fiducial reference measurements (FRMs), which are point measurements found to be representative of soil moisture at the radiometer scale and therefore recommended for satellite validation (Himmelbauer et al., 2023; Goryl et al., 2023). This subset consists of 1314 high-quality time series (Fig. A3 in the Appendix). Soil moisture mea-

surements are provided with associated quality flags for each time stamp and station metadata, such as land cover information extracted from ESA's CCI land cover v1.6.1 dataset. ISMN measurements are unevenly distributed both temporally (with different stations covering different periods) and geographically. This uneven distribution means that our validation results are spatially and temporally biased towards regions and periods with dense station coverage, primarily North America and Europe after the year 2000 (Dorigo et al., 2021b).

3 Preprocessing

The following steps are applied to the original ESA CCI SM v09.1 COMBINED data before the interpolation process takes place.

3.1 Additional retrievals under dense vegetation conditions

Soil moisture values in areas covered by tropical rainforests are associated with high uncertainties (Ulaby et al., 2014). Consequently, these areas are masked out in ESA CCI SM, leading to permanent gaps (Dorigo et al., 2017). Such large gaps are particularly difficult to fill for any algorithm, regardless of whether covariates are used or not (Tong et al., 2021). However, soil moisture is still retrieved, e.g. from deforested patches, and therefore available for some frequency bands in which electromagnetic waves can – at least to some degree – penetrate through vegetation. Specifically, C-band data from ASCAT-A, B, and C, as well as L-band data from SMOS

and SMAP – despite the overall lower data quality compared to other global regions – can serve as useful support points for the interpolation algorithm when available. Hence, these lower-quality observations, despite being masked out in the original ESA CCI v9.1 COMBINED data, are included now in the gap filling to serve as anchor points (compare Fig. 2e). This is preferred over large-scale interpolation without any support data (Guo et al., 2022).

Similar to the rest of the ESA CCI SM data, the additional retrievals available for densely vegetated areas are error-characterized via triple collocation analysis, harmonized through scaling to a common reference, and merged when more than one measurement is available for a grid point on a given day (Liu et al., 2012; Dorigo et al., 2017; Gruber et al., 2017, 2019).

3.2 Gap analysis

The observation density of ESA CCI SM depends on the number of available satellites and surface conditions that affect retrieval success rates. Figure 1b shows the temporal coverage of the product. At the start of the data record, data from only few sensors were available, leading to relatively low data coverage. After 2002, coverage increases significantly, culminating in up-to-daily (i.e. gap-free) measurements for certain latitude bands in recent years. However, some regions still show large data gaps (Fig. 2e), especially around the Equator and above 50° N, due to the dense vegetation cover and seasonally frozen soils, respectively.

Figure 2a–d evaluate data gaps in ESA CCI SM in terms of their 3-dimensional (Euclidean) distance from the nearest available valid data point. Gaps are analysed separately depending on whether the corresponding reanalysis soil temperature is above or below 0 °C. While in most cases a valid observation is found within a range of a few pixels or days, in some exceptional cases, the combined gap size in days and pixels can reach up to 350.

The so derived distances are required for the initial guess for gap filling (nearest neighbour), which is then further adjusted by the DCT-PLS. They are also essential for deriving uncertainty estimates.

4 Gap-filling methods

To fill data gaps between observations in ESA CCI SM and assign uncertainty estimates to the predicted values, we follow the processing chain in Fig. 3. We start from daily ESA CCI SM data with additional observations in the tropics (Sect. 3.1). We then compute the gap statistics (Sect. 3.2) required for all subsequent steps: (i) for each gap, we compute the distance to the nearest observation (across 3 dimensions), and (ii) we differentiate between gaps due to frozen soils and other reasons based on (gap-free) ERA5 temperature information. We then fill data gaps in ESA CCI SM by applying the DCT-PLS algorithm with local parameteriza-

tion (Sect. 4.1) but overwrite the so-derived values with a linear interpolation over time in periods where soil moisture is frozen in a certain place (discussed in Sect. 4.1.1). Prediction uncertainty models are derived based on the performance of our method in restoring GLDAS Noah soil moisture data after introducing gaps from ESA CCI SM (Sect. 5). Finally, the gap-filled soil moisture and uncertainty estimates from gap statistics and pre-computed models are combined in the final GAPFILLED product.

4.1 Core algorithm: DCT-PLS

We only provide a summary of the DCT-PLS algorithm here. For the full description with all intermediate steps, see Garcia (2010). Initially designed as a data smoother, DCT-PLS can also provide estimates for missing data points by predicting a set of smoothed values $\hat{y}$ for the original/input data y. The algorithm optimizes for (i) minimal residual sum of squares (RSS) between the input and smoothed observations and (ii) optimal reduced roughness $P(\hat{y})$ between (smoothed) elements (Eq. 1). DCT-PLS therefore ideally removes noise in the data while retaining relevant information, i.e. finding smoothed versions of input observations y so that $F(\hat{y})$ is minimized.

$$F(\hat{y}) = \text{RSS} + s\,P(\hat{y}) = \|y - \hat{y}\| + s\,P(\hat{y}) \quad (1)$$

The penalty term $P(\hat{y})$ is based on (temporal and spatial) neighbourhood information as differences (D) between elements of $\hat{y}$, thus optimizing for smoothed transitions between them (Eq. 2).

$$P(\hat{y}) = \|D\hat{y}\|^2 \quad (2)$$

Solving the linear system (Eq. 3, where I_n indicates the identity matrix) to minimize $F(\hat{y})$ is a computationally extensive task when y is large.

$$(I_n + s\,D^T D)\hat{y} = y \quad (3)$$

Garcia (2010) provides a step-by-step description of the required (matrix) operations to find $\hat{y}$ and discusses modifications to the base algorithm for improved performance. In particular, computing D and performing eigenvalue (λ) decomposition thereof is greatly simplified by the use of equally spaced input data such as ESA CCI SM. In fact, a predefined formulation for λ_i of D can be used (Eq. 4) where i are the values along the dimension and n the value count.

$$\lambda_{i_j} = -2 + 2\cos((i_j - 1)\pi/n_j) \quad (4)$$

In the 3-dimensional case ($N = 3$; time, latitude, longitude), this leads to tensor Λ (Eq. 5).

$$\Lambda^N_{i_1,\ldots i_N} = \sum_{j=1}^{N} \lambda_{i_j} \quad (5)$$

https://doi.org/10.5194/essd-17-4305-2025

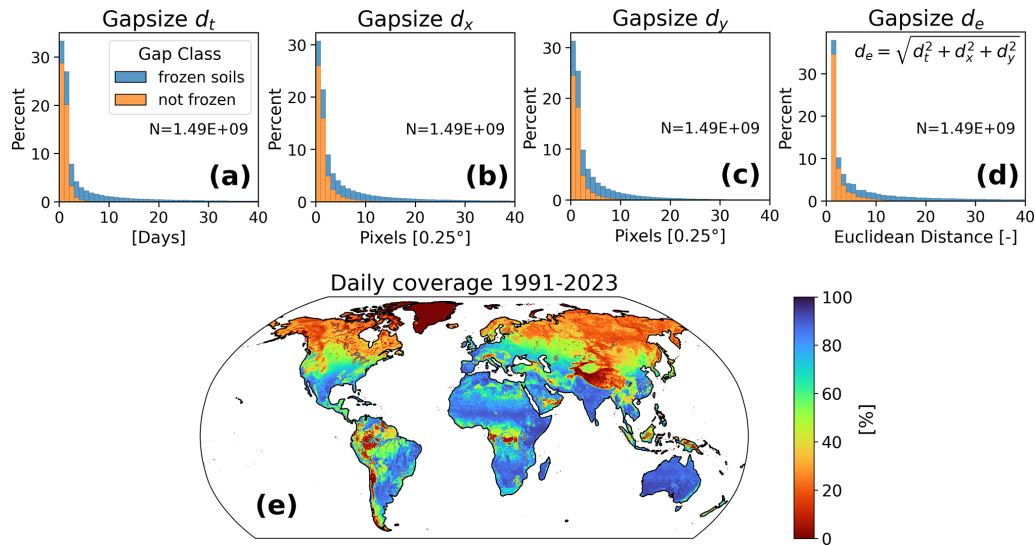Earth Syst. Sci. Data, 17, 4305–4329, 2025

Figure 2. Distance to the nearest neighbour for each missing data point (excluding Greenland) in ESA CCI SM along each dimension: time (**a**), longitude (**b**), latitude (**c**), and Euclidean distance across dimensions (**d**). Very large gaps (40–350 days/pixels) exist but are too rare to be visible in the histograms. The temporal coverage for each grid point with additional anchor observations over tropical rainforests is shown in (**e**). Gap classification (frozen vs. non-frozen) based on ERA5 soil temperature.
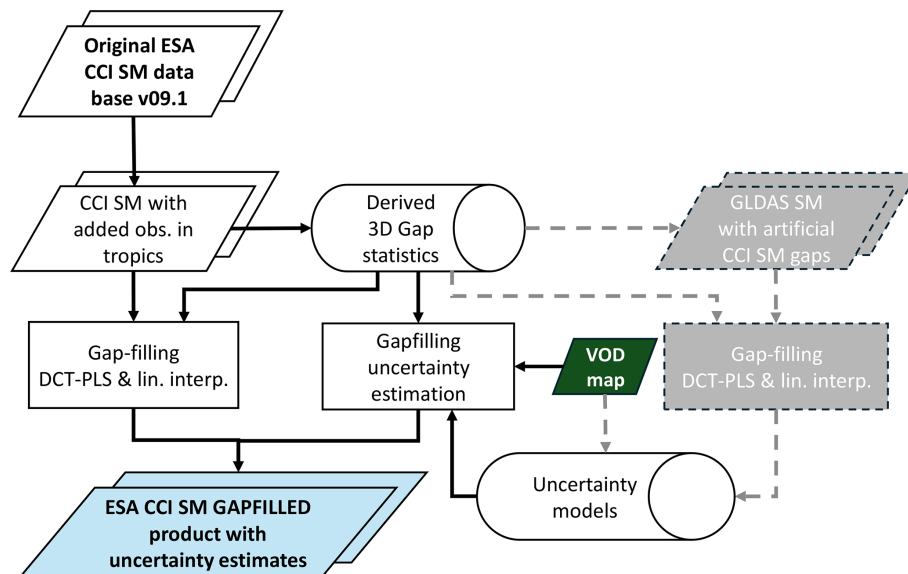

Figure 3. Flow chart of the main process steps to derive the ESA CCI SM GAPFILLED product (blue box). Dashed lines indicate optional processing steps that – once the model parameters are known – can be left out.

Together with a realization of the smoothing parameter s, we can define the tensor Γ^N (Eq. 6; $\div$ and $\circ$ symbolize the element-by-element division and multiplication).

$$\Gamma^N = 1^N \div (1^N + s\Lambda^N \circ \Lambda^N) \qquad (6)$$

This allows us to efficiently solve for $\hat{y}$ using the discrete cosine transform matrix (DCTN) of y and the inverse form (IDCTN) for data with the dimension $N = 3$, respectively (Strang, 1999). As data gaps are present in ESA CCI SM, as-

signing weights (W) to observations is required (Eq. 7). Data gaps are assigned a weight of zero and therefore interpolated as part of the (robust) smoothing process, while observations have an initial weight of 1. k here refers to one realization of $\hat{y}$ because this step is repeated multiple times when optimizing s, as explained later.

$$\hat{y}_{k+1} = \text{IDCTN}\left(\Gamma^N \circ \text{DCTN}(W \circ (y - \hat{y}_k) + \hat{y}_k)\right) \qquad (7)$$

Finally, having an estimate for $\hat{y}$, the generalized cross-validation (GCV) score is computed (Eq. 8), where n_{miss} is the number of missing values of n overall samples.

$$
\begin{aligned}
\text{GCV}(s) &= \frac{\text{wRSS}/(n - n_{\text{miss}})}{(1 - Tr(H)/n)^2} \\
&= \frac{\|W^{1/2}(\hat{y} - y)\|^2/(n - n_{\text{miss}})}{(1 - Tr(H)/n)^2}
\end{aligned}
\tag{8}
$$

$Tr(H)$ is computed from λ (Eq. 4) as described in Eq. (9).

$$
Tr(H) = \sum_{i=1}^{n} [1 + s(2 - 2\cos((i-1)\pi/n))^2]^{-1}
\tag{9}
$$

A bounded minimization – as implemented in SciPy's "optimize" module (Virtanen et al., 2020) – is now applied to find the smoothing parameter s for minimal GCV, and $\hat{y}$ is computed again using this s. s is the only free parameter in the model that needs to be tuned using GCV.

As outliers can be present in ESA CCI SM data, we use the "robust" implementation of DCT-PLS (Garcia, 2010), which includes further iterations of the above described process to detect outliers in the data and gradually reduce weights assigned to these observations until the optimization converges at $\hat{y}_{k+1} - \hat{y}_k \rightarrow 0$, meaning that no more relevant changes between y and $\hat{y}$ – before and after calibrating s, respectively – are found.

DCT-PLS can take n-dimensional tensors of any size. We apply it to (3-dimensional) spatial subsets of the dataset via a $15 \times 15°$ moving window. We prefer a moving window over a global s value to locally allow for different degrees of smoothing and therefore reducing potential overfitting or over-smoothing when using a global s value (Wang et al., 2012). Therefore, local parameterization of s is also expected to improve the interpolation results between areas with different levels of autocorrelation. In most cases, the chosen window is large enough to provide sufficient data for a robust estimation. However, for some edge cases, such as remote islands, the initial window size is gradually extended by $5°$ (in eight directions, in practice, the maximum was $35°$) until the GCV minimization process has sufficient data to converge for s.

Predictions are made not only for missing data points, but also at locations where observations are available. While these (smoothed) observations are not used in the final gap-filled product (because the original observations are kept when available), they can be used to further harmonize predictions and observations before they are combined (see Sect. 4.1.2).

4.1.1 Frozen soils

During periods of frozen soil moisture, it is a matter of definition whether a dataset should represent only the liquid or the total (frozen and non-frozen) soil moisture content. In the first case, soil moisture would start to decline as soon as water begins to freeze in a scene until it reached $\sim 0\,\text{m}^3\,\text{m}^{-3}$. Factors such as freezing point depression due to solutes can result in the presence of liquid soil moisture even when the soil temperature is below $0\,°\text{C}$ (Amankwah et al., 2021). In the second case, which we try to cover in our data record, soil moisture remains constant over the period when it is frozen; i.e. there is no soil water loss due to evaporation of percolation. However, when predicted observations are based on insufficiently-masked, low, and often noisy soil moisture observations from transitional zones/periods around a frozen soil gap, these are often too low (Wang et al., 2012; Liu et al., 2023). In addition, since DCT-PLS does not use any information on soil temperature, it may predict temporal fluctuations in soil moisture even when the soil is frozen. One could argue that this is the result of regional (sub-pixel) freeze/thaw processes, but there is no reason to assume that univariate algorithms, which do not account for soil temperature on a sub-pixel scale, can accurately predict this.

We therefore carry the last (observed or interpolated) soil moisture level before freezing starts forward over time until thawing occurs and soil moisture changes can be measured again directly by the satellite. In practice, there can be differences between the last and first available observations around a period of frozen soil, e.g. due to measurement noise. We therefore use the mean of 30 d before and after an affected period to perform a simple linear interpolation over time between these points. This is in line with our definition of frozen soil moisture content, deals with the expected noise in observations around the affected period, and avoids temporal fluctuations in predictions when soil moisture should remain stable. These values are then used to replace those from the DCT-PLS algorithm.

4.1.2 Combining predictions and observations

For the final GAPFILLED product, we use the predictions (from either DCT-PLS or linear interpolation) to fill data gaps in the original record. The available original observations are not replaced. However, since DCT-PLS uses temporal and spatial neighbourhood information to create predictions, the mean and variance of the predictions for a grid point are adjusted towards those of its neighbours, giving the data a (spatially) smooth appearance. This characteristic, typical of any spatial interpolation method (Llamas et al., 2020), can lead to temporal inconsistencies between predictions and observations. To address this, we apply a linear transformation (derived from linear regression on shared data points) to scale the mean and variance of the predictions to match those of the observations (Steven et al., 2003; Paulik et al., 2025). Finally, the scaled predictions are used to fill gaps in the observation time series.

4.2 Validation

The gap-filled ESA CCI SM product is compared to ISMN in situ reference measurements using the QA4SM online validation platform (https://qa4sm.eu, last access: 20 August 2025). Time series for grid points collocated with FRMs (see Appendix Fig. A3) are extracted from the original ESA CCI SM COMBINED and the GAPFILLED product. For the latter, we extracted four different subsets: (i) all available data points from the GAPFILLED product (original observations plus filled values), (ii) only the filled values, (iii) only the filled values from "small" gaps (classification based on the Euclidean distance (ED) to the nearest observation) with ED < 2 (i.e. adjacent to a valid data point in any dimension), and (iv) only the filled values within "medium/large" gaps (ED ≥ 2). All four datasets were uploaded to the service. We select a temporal matching window of ±1 h around 00:00 UTC to align in situ measurements with the satellite data (temporal reference). QA4SM offers filtering options to include only in situ time stamps flagged as "good" – excluding erroneous measurements from, for example, frozen soils (Dorigo et al., 2013) – and "(very) representative" time series in terms of FRM qualification. Biases between satellite and in situ time series are removed by matching their mean and standard deviation using the in situ data as the (scaling) reference (Gruber et al., 2020). QA4SM validation results include ESA CCI land cover information at the in situ sites, which we use to stratify our results.

5 Uncertainty estimation

An uncertainty estimate is provided for each interpolated value. This uncertainty depends on the uncertainty of the observations used ($\sigma_{\text{observations}}$) and the availability of support data and other inherent factors affecting the quality of DCT-PLS predictions (summarized as $\sigma_{\text{gapfilling}}$). Combining both as in Eq. (10) yields the uncertainty for our predictions.

$$\sigma_{\text{prediction}} = \sqrt{\sigma_{\text{observations}}^2 + \sigma_{\text{gapfilling}}^2} \qquad (10)$$

$\sigma_{\text{observations}}$ is characterized in ESA CCI SM using triple collocation analysis and uncertainty propagation (Gruber et al., 2016). The random error level of observations mainly depends on the merged sensor frequency bands in conjunction with surface characteristics, such as vegetation cover (Parinussa et al., 2011). Recently, a temporal component was added to account for (sub)seasonal error dynamics (Stradiotti et al., 2025). We use $\sigma_{\text{observations}}$ as the baseline for uncertainties in the gap-filled values. For a conservative estimate, we choose the 95th percentile of measurement uncertainties for any grid point series (or the nearest-neighbour series in the case of a permanent gap).

To estimate $\sigma_{\text{gapfilling}}$, we impose the gaps in the satellite data on gap-free GLDAS Noah soil moisture and subsequently restore these gaps using the presented methodology

Table 1. Model parameters found from fitting $f(x)$ for different VOD levels (VOD classification in Fig. A1 in the Appendix).

$f(x) = b \cdot (1 - e^{a \cdot x})$	a	b
high VOD	−0.089	0.128
medium VOD	−0.218	0.08
low VOD	−0.234	0.035

(Fig. 3). This approach preserves the original satellite gap systematics as opposed to randomly splitting the input data into training and validation sets, as is commonly done (e.g. Wang et al., 2012; Liu et al., 2020b; Zhang et al., 2021). The goal of this analysis is to derive a set of model functions for the expected discrepancy between predictions and observations based on the gap size. Uncertainty models are preferred over empirical values because they remain applicable even in scenarios where GLDAS data are unavailable, ensuring broader utility of the methodology (e.g. for the period before 2000, when GLDAS Noah v2.1 data are not available). For improved accuracy, these functions should account for both realistic gap conditions and varying surface complexities. To address the former, we use the previously computed gap statistics (Euclidean distances). To address the latter, we use a static VOD classification map for low, medium, and high VOD based on the 2000–2023 average, derived from the VODCA v2 CXKu archive (compare Fig. A1 in the Appendix). We therefore assume that prediction uncertainty will generally increase for larger gaps and with denser vegetation cover, consistent with the observations (Parinussa et al., 2011).

From analysing the differences between original and restored values with regard to the gap systematics and VOD level (compare Fig. 4a), a log-like function of form $f(x) = b \cdot (1 - e^{a \cdot x})$ is fitted for different vegetation conditions. Each a and b is found from fitting 1000 realizations of this function (least squares with boundary conditions for $a < 0$ and $b > 0$) to randomly drawn samples without replacement. Using the medians of the obtained parameter sets for a and b (Table 1), we defined the required models for the uncertainty of the gap-filling process ($\sigma_{\text{gapfilling}}$) itself as shown in Fig. 4b.

6 Results

6.1 DCT-PLS parameterization

Figure 5 shows the result of the local parameterization of the DCT-PLS model. A lower s means that more (day-to-day temporal and/or regional spatial) variability is found in the prediction (Fig. 5a). The highest s values are found in arid climates, in northeast Africa and Arabia. This coincides with the expected low variability in soil moisture in these regions. Low s values (i.e. high variability) are found in subtropical regions in central Africa and South America, South
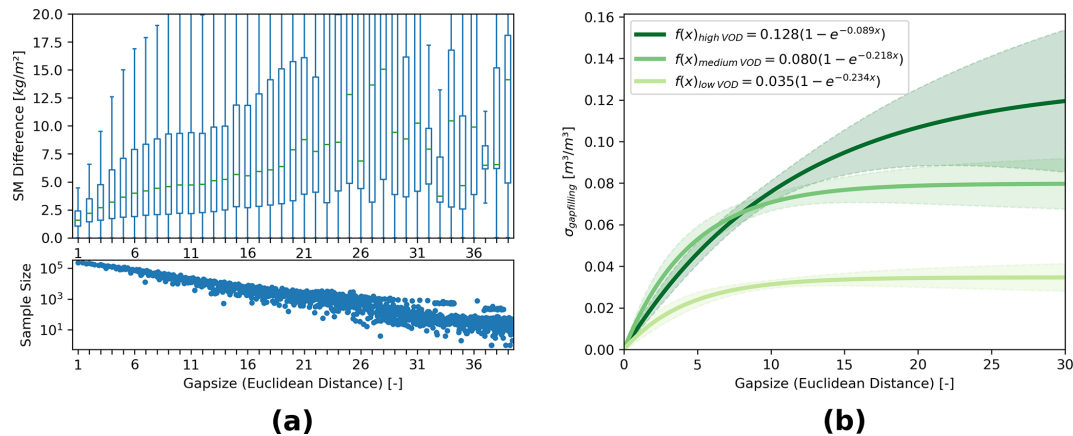
Figure 4. Absolute differences between original and restored GLDAS Noah soil moisture after gap filling **(a)** and function fits **(b)** (median and standard deviation from 1000 randomly drawn samples) to predict $\sigma_{\text{gapfilling}}$ for high, medium, and low VOD conditions.

Asia, and East Australia. The GCV scores in Fig. 5b were minimized to find the optimal s in each cell. For most regions, the GCV score is close to zero, which indicates a good fit. Higher scores are found in subtropical regions, which indicates a slightly larger discrepancy between the predictions and available observations in these areas. This could be due to underestimating the seasonal and/or overfitting of the day-to-day soil moisture variability in the measurements.

6.2 Spatial data characteristics

The GAPFILLED product provides full daily coverage for both soil moisture and uncertainty estimates. For a selected day, Fig. 6 shows both fields before and after gap filling, respectively. The gap filling has produced a smooth, spatially consistent image without any outlier values or edges that could arise from sudden differences in regional parameterization of the DCT-PLS model. As expected, soil moisture in tropical latitudes is high as a result of interpolating anchor retrievals from L- and C-band measurements within and surrounding these areas. In high northern latitudes, the predictions come mainly from linear time interpolation between autumn values of the preceding year and spring values observed later in the year.

The uncertainties are within the range of 0 to 0.1 m³ m⁻³, which is expected from the defined models for $\sigma_{\text{gapfilling}}$ and the value range of the original (triple collocation) error estimates $\sigma_{\text{observations}}$ (Gruber et al., 2016, 2019). The highest values are found in (sub)tropical and boreal ecozones, the Tibetan Plateau, and southeast Asia. This is due to the low data coverage and quality of the used retrievals in these regions, which are noisy for dense vegetation and often missing in mountainous regions.

While uncertainties increase as expected in most regions covered by dense vegetation, an inverse effect appears in some areas of the northwestern Amazon and the central African rainforest (Fig. A2 in the Appendix), where the un-

certainties of the GAPFILLED product are lower than in parts of the surrounding subtropical zones. Our analysis of this phenomenon shows that – although uncertainties in the GAPFILLED product increase according to $\sigma_{\text{gapfilling}}$ – the initial uncertainty estimates based on triple collocation analysis $\sigma_{\text{observations}}$ in the affected regions are too low (Fig. A2a). This likely results from the poor agreement between satellite and model data in these areas, whereas similar soil moisture climatologies are found in the (independent) active and passive satellite products (Fig. A2d).

6.2.1 Time series characteristics

Figure 7 shows temporal subsets for five grid points in the Northern Hemisphere, where also in situ measurements are available. The locations were selected to represent a wide range of environments and gap types as far as the spatial distribution of in situ sensors allows (for more information on the chosen sites, see Table A1 in the Appendix). For visualization purposes, a linear least-squares regression scaling was applied to remove biases between point- and satellite-scale time series (Steven et al., 2003; Paulik et al., 2025). Anomalies are computed relative to the 1991–2023 average conditions. All R scores are based only on data points from the visualized period and are statistically significant ($p < 0.05$). Figure 7a illustrates the performance of the gap-filling algorithm in predicting soil moisture for gaps because of frozen conditions and gaps emerging for other reasons in an area with a moderate number of missing data points (40 %). The predicted values align reasonably well with the available in situ data, with an anomaly correlation of 0.59 between the in situ measurements and the filled values only ($R_{\text{FILLED VALUES}}$, without the original observations). This is lower than the 0.70 correlation between in situ measurements and original satellite observation anomalies. However, this is hardly surprising, as systematic gaps appear during more challenging conditions. This plot also shows how combining
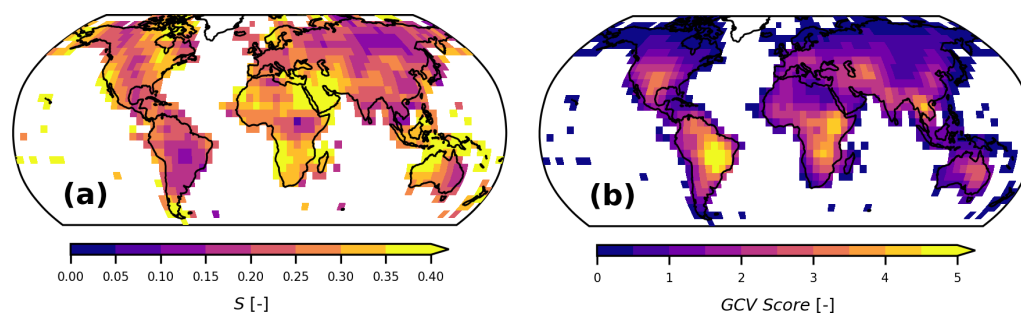
Figure 5. Local parameterization of the DCT-PLS function to find the optimal value for *s* (**a**) that minimizes the GCV score (**b**).
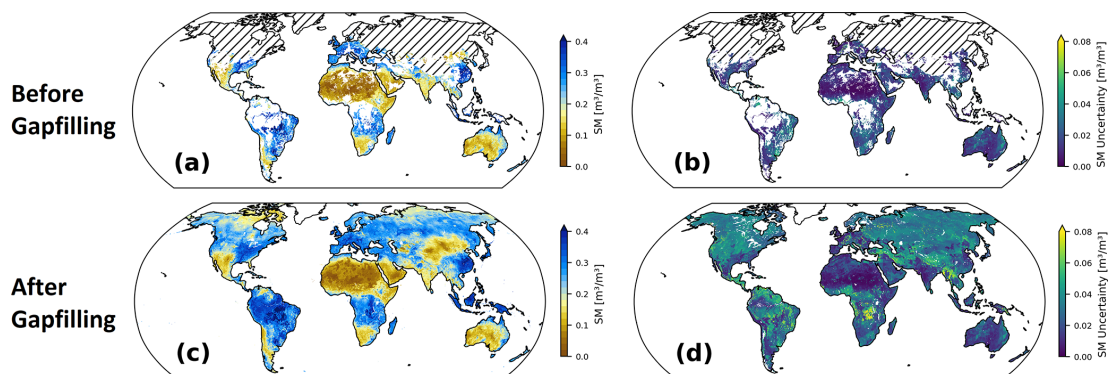

Figure 6. Soil moisture and uncertainty before (**a, b**) and after (**c, d**) gap filling on 1 March 2007. Ice sheets (Greenland, Antarctica) are excluded from the GAPFILLED product. Hatched regions indicate ERA5 soil temperature below 0 °C.

data from both the DCT-PLS and linear interpolation in a single time series based on soil temperature can lead to inconsistencies for some edge cases (February 2018). In this case, an individual, unobserved "non-frozen" data point is found in the middle of a prolonged frozen period. Differences in the predictions of the two interpolation algorithms can lead to outliers in this case. Figure 7b is an example of a location affected strongly by seasonally missing data. The periods of linear interpolation (frozen) align well with missing in situ values due to the (independent) quality-flag-based masking based on in situ temperature measurements (Dorigo et al., 2021b). The reanalysis temperature data below 0 °C therefore, for this location, provides a good approximation of the period when soil moisture is frozen and kept at an almost constant level. However, it also means that a direct comparison between gap-filled values and in situ measurements is usually not possible during winter. For the onset and termination of this period, the absolute values between gap-filled and in situ data match well. Figure 7c represents a location with only few missing data points and no relevant periods of frozen soil moisture. In this case, we find a higher correlation with the in situ data for the gap-filled values compared to the original data (0.89 vs. 0.80), which we attribute to the smoothing effect of DCT-PLS and therefore reduced noise in the predictions. Figure 7d is for a location near the Equator, where ESA CCI SM is normally permanently masked due

to the presence of dense vegetation (i.e. 100 % filled values). The example time series shows how the additional C- and L-band retrievals in this region form a consistent although somewhat noisy record, which matches reasonably well with the in situ data (anomaly correlation of 0.54). Figure 7e is located in a complex topography on the Tibetan Plateau, with significant amounts of missing data points due to soil freezing. The performance of ESA CCI SM is generally poor in this region and the gap-filled product shows only little variation and often contradicts the (sparse) in situ measurements. The available observations are very noisy in this location, and hence do not form a consistent time series with the gap-filled data points.

6.2.2 Global anomalies and uncertainties

Global monthly anomalies and uncertainties over time are shown in Fig. 8. The similarity in anomalies before (Fig. 8a) and after (Fig. 8c) gap filling indicates that large-scale deviations from normal conditions are equally well represented in the gap-filled as in the original data. The main differences (Fig. 8e) are found in equatorial regions (mainly due to the additional observations used) and in transition zones/periods between frozen and unfrozen soil moisture.

Uncertainties are high at the start of the record, when data coverage is lowest (compare Fig. 8b, d to Fig. 1b), and for
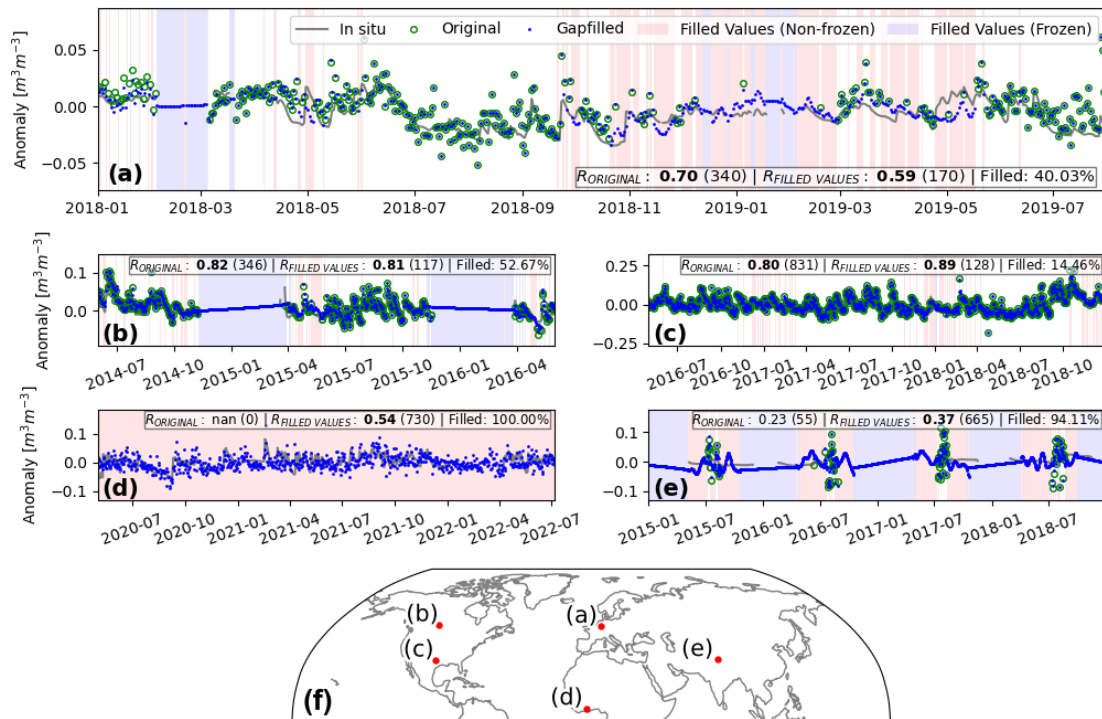
Figure 7. Selection of soil moisture anomaly time series from the original and gap-filled ESA CCI SM data set and available in situ measurements (relative to the 1991–2023 climatological average). The locations of **(a)**–**(e)** are indicated in **(f)**. All statistics are based only on data from the shown sub-periods. R scores shown in bold are statistically significant ($p < 0.05$) and are based on the number of data points indicated in parentheses. For more information about the chosen locations, see Table A1 in the Appendix.

equatorial latitudes due to reduced coverage and observation quality. Uncertainties in the gap-filled values often exceed $0.1\,\mathrm{m^3\,m^{-3}}$ in areas where ESA CCI SM measurements are seasonally masked due to frozen soils, and linear interpolation potentially does not sufficiently account for diurnal freeze–thaw processes, evaporation, and other factors. However, it is important to note that these higher uncertainties reflect the inclusion of additional data points rather than a decline in dataset quality.

6.2.3 Zonal anomalies and trends

Figure 9 shows annual anomalies expressed as Z scores (Vreugdenhil et al., 2022) for different latitude bands of the original observation-only ESA CCI dataset, its gap-filled counterpart, and two reanalysis products. Gap filling can influence annual anomaly estimates, particularly when large portions of observational data are (systematically) masked. This effect is evident in the global estimates (Fig. 9a), where the 1991–2023 trend shifts from negative to neutral after gap filling. The most significant discrepancies between the original and gap-filled ESA CCI SM data occur during the first third of the time series (1992–2002), a period with limited observation coverage compared to the era following the launch of AMSR-E in 2002. The gap-filled data are drier than the original and the reanalysis products. From 2002 to 2016,

we find good agreement between all four datasets until the two CCI (and reanalysis) records start to diverge again, with gap-filled anomalies exceeding those of the original data. The extreme dry anomaly in the original dataset in 2019 is mitigated after gap filling, resulting in a closer alignment with reanalysis. During the last 4 years, CCI and GLDAS agree well, while ERA5-Land reached the lowest value on record in 2023. For the Southern Hemisphere (Fig. 9b), we find only minor differences between the original and gap-filled data, mainly in the beginning and at the end of the time series. We attribute this partly to the good initial data coverage, but it also indicates that the integration of additional observations for the previously permanently masked tropical rainforest regions was successful. The gap-filled time series is now slightly less variable and closer to GLDAS Noah. The overall trend for the Southern Hemisphere has slightly changed, from neutral to negative, in line with the reanalysis products. Looking at Z scores for the tropical zone (Fig. 9c) separately, we find a negative soil moisture trend after gap filling. This trend, as well as the annual anomalies in general, align better with the reanalysis products, compared to the original data. Some differences are found during the first 5 and the last 10 years, when gap filling has a better aligned CCI with GLDAS. Figure 9d comprises the data from a wide range of climatic regions, from arid over temperate to boreal conditions. This subset contains the majority of points
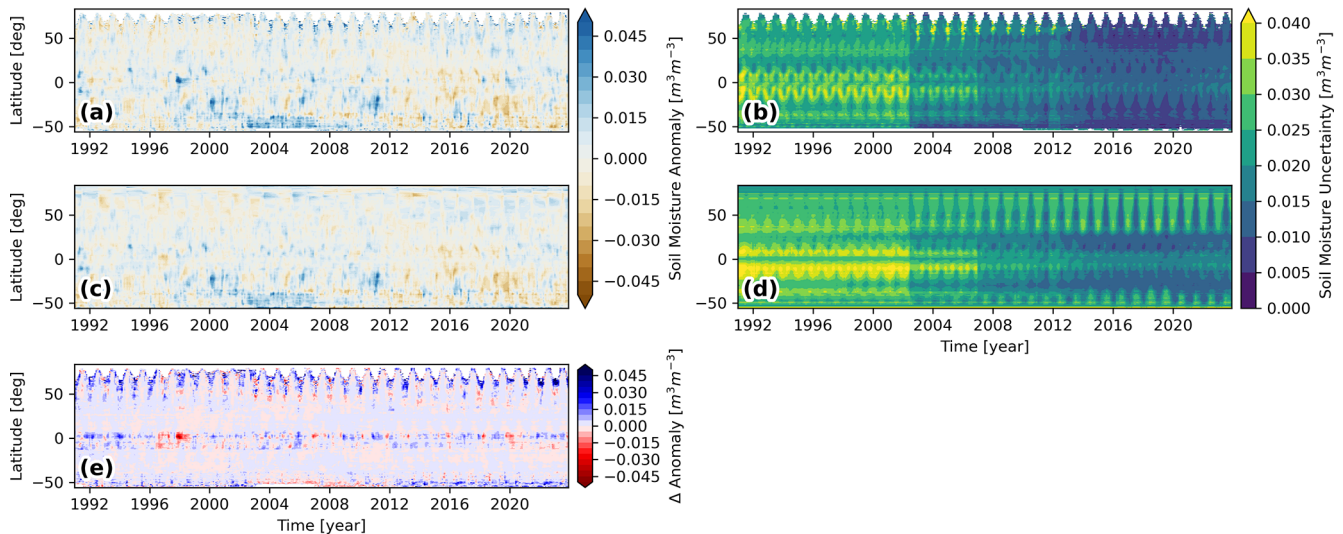
Figure 8. Time-latitude diagrams of monthly soil moisture anomalies and soil moisture uncertainties as in the original COMBINED **(a, b)** and GAPFILLED **(c, d)** product. Differences between **(a)** and **(c)** are shown in **(e)**.

over land for the Northern Hemisphere (excluding Greenland). From 1991 to 2003 we find drier and from 2013 to 2020 wetter conditions after gap filling, which matches the observations from Fig. 9a in the same period, where it had a similar effect on the long-term trend in the data. The largest discrepancies between the original and gap-filled data are found for the boreal zone (Fig. 9e). The effect of the interpolation of seasonally missing values is clearly visible and has led to a switch from a slight negative to a strong positive trend, which is in contrast to ERA5-Land. However, it should be noted that the reliability of reanalyses in this case is also unknown. This is also evident from the discrepancies between GLDAS Noah and ERA5-Land in this zone.

We further assess global trend changes in Fig. 10, showing Theil–Sen slopes computed from annual averages between 1991 and 2023, with a Mann–Kendall test for statistical significance (Dorigo et al., 2012). Figure 10a is for the original ESA CCI SM data (with gaps), and Fig. 10b is the same for the gap-filled product. Figure 10c summarizes the differences in terms of change in trend direction and significance. Only a very small number of initially detected significant trends changed their direction due to gap filling, either from positive to negative (1.5 %) or vice versa (1.4 %). Affected regions are in the northwest United States, Siberia, and Myanmar, where positive trends are found after gap filling, as well as central Africa and some spots in South America, where the trend direction was inverted to negative. More prominently, large parts of the Northern Hemisphere, where previously no significant increase in soil moisture was found, now show a positive trend. This is the case for large parts of Russia, Alaska, the northwest United States, and Canada. The opposite is often found in the Southern Hemisphere, mainly in central and western Africa and South America and also in northeast

China and Mongolia. The previously masked regions covered by tropical rainforest do not show spatially consistent trends after gap filling. For parts of Brazil and Columbia in particular, trends vary spatially, while a consistent negative trend is found for the lower half of South America, which only changed in some spots in the far south due to gap filling.

6.3　Evaluation of $\sigma_{\text{gapfilling}}$

Figure 11 shows Pearson's R (masked for $p < 0.05$) and unbiased root mean squared difference (ubRMSD), between original daily soil moisture from GLDAS Noah and restored GLDAS Noah points for the imposed ESA CCI SM gaps. The gap-filling algorithm manages to restore the data well (global median $R = 0.81$, ubRMSD $= 2.69\,\text{kg\,m}^{-2}$, corresponding to $\sim 10\,\%$ relative error). When data from gaps classified as "frozen" (Sect. 3.2) are excluded, i.e. without values filled by linear interpolation, performance metrics improve (global median $R = 0.89$, ubRMSD $= 2.24\,\text{kg\,m}^{-2}$). By design, regions with a low R and high ubRMSD in this analysis match with regions where our uncertainty models predict a lower gap-filling accuracy (Figs. 6d, 8d).

6.4　GAPFILLED soil moisture validation with in situ data

Figure 12 shows aggregated performance metrics (R and ubRMSD) between FRM in situ sites and the original ESA CCI SM COMBINED product, the new GAPFILLED product (original data with filled values), and the subset consisting only of the filled values. All three datasets generally perform on a similar level. In terms of absolute value correlations, the GAPFILLED product slightly outperforms the original data. Note, however, that the temporal and spatial coverage differs among the three products. Logically,
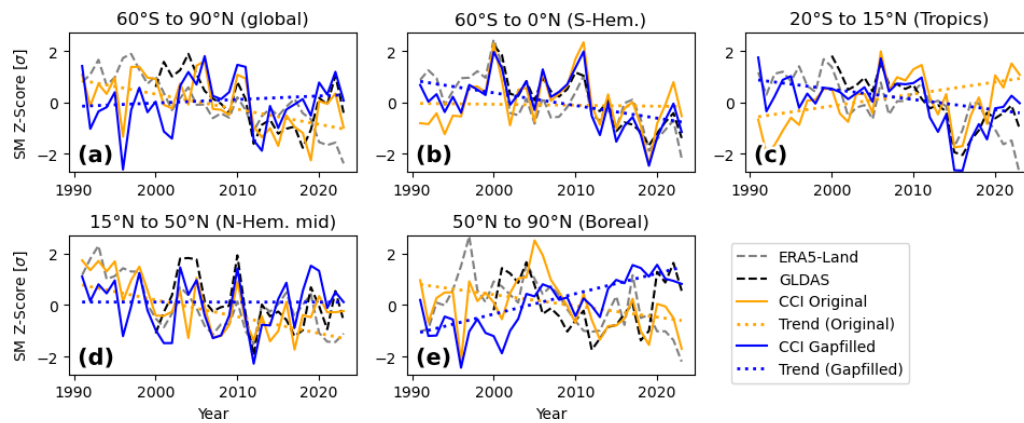
Figure 9. Annual soil moisture anomaly Z scores across different latitude bands: **(a)** global, **(b)** Southern Hemisphere, **(c)** (sub)tropical zone, **(d)** arid/temperate zone, **(e)** boreal zone. The dotted coloured lines indicate the zonal trends in the original and gap-filled data.
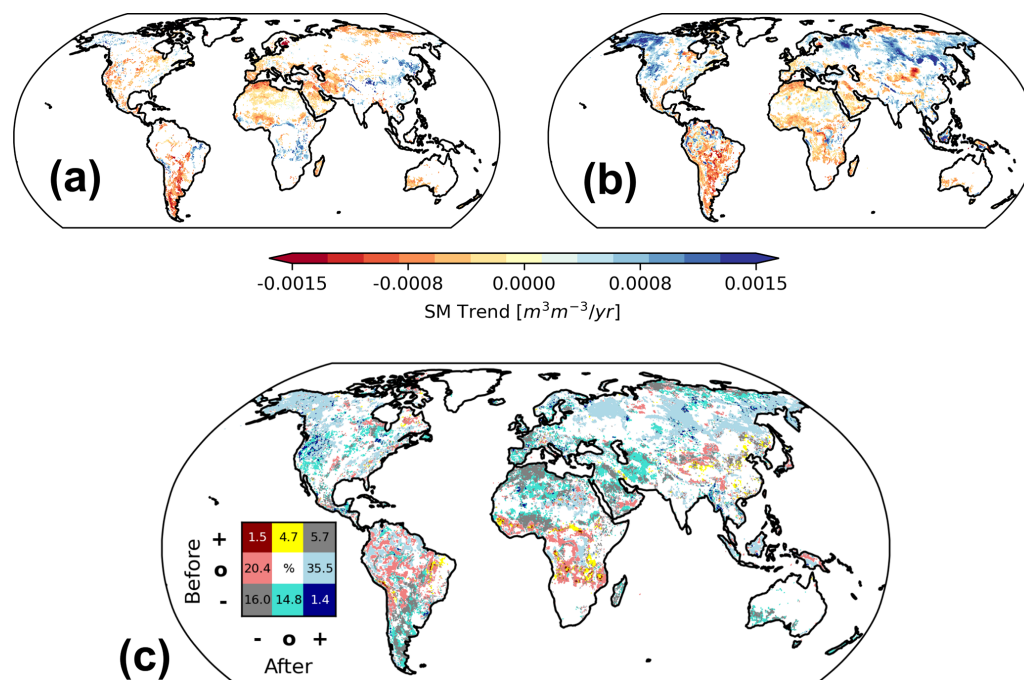

Figure 10. Significant ($p < 0.05$) long-term (1991–2023) soil moisture trends before **(a)** and after **(b)** gap filling. Panel **(c)** shows the change in trend direction and significance: "+" indicates wetting, "−" drying, and "0" non-significant trends before and after gap-filling, respectively. Inset numbers (%) are relative to the total number of points with a significant trend.

there is no overlap in observations between the original and the filled values. Fill values typically comprise challenging cases, where satellite retrievals were not possible in the first place. Consequently, it is unsurprising that the performance of only the filled values slightly lags behind the original dataset in terms of absolute values R (Fig. 12a) and ubRMSD (Fig. 12c), especially for stations near or under dense vegetation ("Tree Cover"). Metrics are slightly more spread for the filled values. A good performance of the GAPFILLED product is also found in terms of anomaly R and ubRMSD, indicating that the gap-filling algorithm manages to capture

not only the seasonality, but also short-term events on par with the observations (based on the same 1991–2023 reference period as in Sect. 6.2.1). The station count for anomaly metrics can be lower than for absolute values as QA4SM excludes time series for which a reliable anomaly cannot be computed (e.g. due to insufficient temporal coverage).

In Fig. 13, we assess the performance of the filled values separately in "small" and "medium/large" gaps with respect to the same in situ reference measurements. Figure 13a shows that the temporal agreement between in situ and filled values decreases for larger gaps. However, it is less distinct
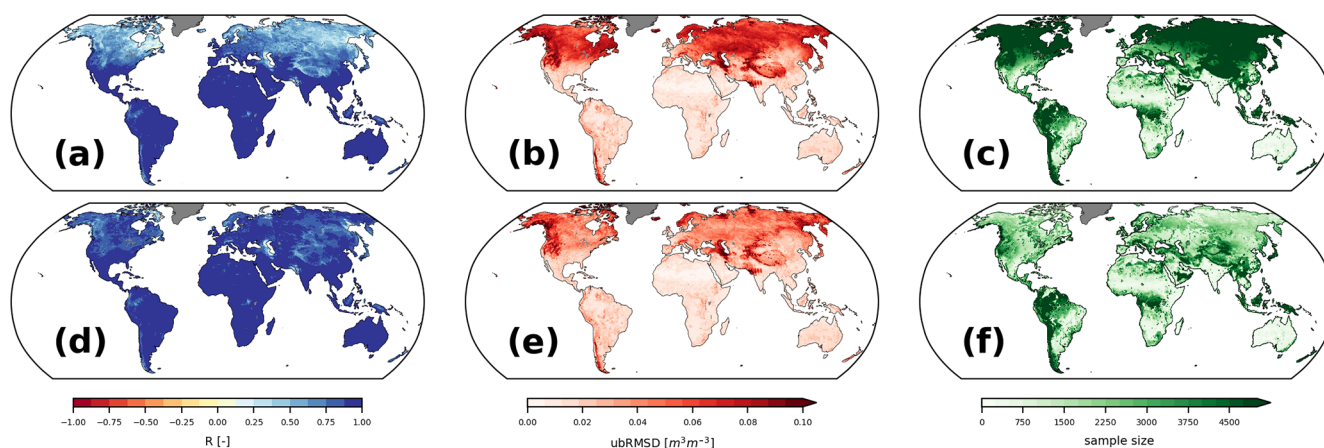
https://doi.org/10.5194/essd-17-4305-2025

Earth Syst. Sci. Data, 17, 4305–4329, 2025

Figure 11. Agreement between original and restored GLDAS Noah surface soil moisture obtained after imposing data gaps from ESA CCI SM and subsequent filling. The top row is based on all restored data points, and the bottom row is with frozen periods excluded. Pearson's R in (**a**) and (**d**) is masked if $p > 0.05$ ($\approx 3\,\%$ of all grid cells). ubRMSD is shown in (**b**) and (**e**), and the respective sample sizes (i.e. number of gaps over time) over the period 2000–2023 are shown in (**c**) and (**f**).
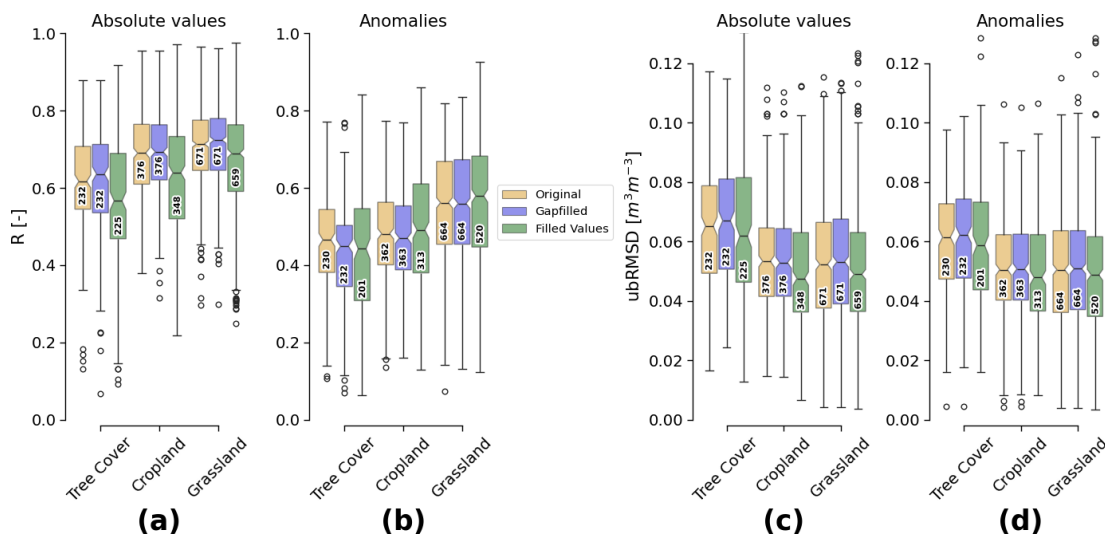

Figure 12. Performance metrics between ISMN soil moisture time series and the "original" ESA CCI SM v09.1 COMBINED product, the GAPFILLED product, and the "filled values" of the GAPFILLED product only. Pearson's R ($p < 0.05$) based on absolute values (**a**) and anomalies (**b**) and ubRMSD (**c, d**), respectively. The numbers in each box indicate the count of contributing time series.

in terms of ubRMSD (Fig. 13b), which barely increases for larger gaps. Note, however, that there is a severe lack of reference data for the larger gaps, mainly due to the unavailability of in situ measurements during winter. Hence, it was not possible to evaluate "medium" and "large" gaps separately.

7 Code and data availability

The ESA CCI SM v09.1 GAPFILLED dataset is available at https://doi.org/10.48436/hcm6n-t4m35 (Preimesberger et al., 2024). It contains the gap-filled original soil moisture observations, gap-free uncertainty estimates, and the pure predictions before reimposing the observations (i.e. for missing

data points as well as in place of the original measurements). We also provide two binary masks in the data to separate (original) observations and filled values in the gap-filled soil moisture field and to differentiate between DCT-PLS predictions and linear interpolation values over frozen periods.

Our Python implementation of the DCT-PLS algorithm is based on the original MATLAB implementation by Garcia (2010) and is provided as part of the pytesmo soil moisture toolbox package in v0.18.0 (https://doi.org/10.5281/zenodo.14975386, Paulik et al., 2025), available at https://github.com/TUW-GEO/pytesmo (last access: 20 August 2025).
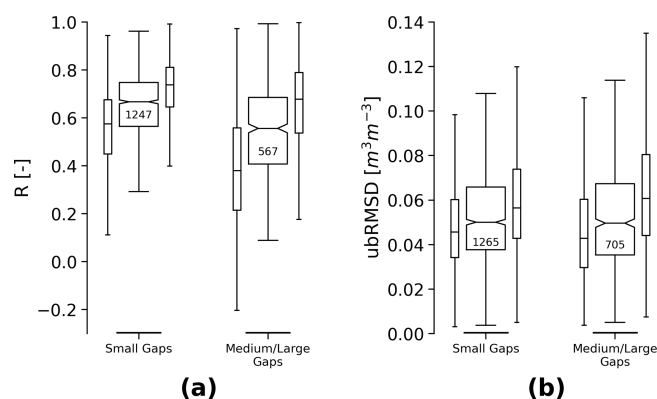
Figure 13. Performance metrics between ISMN soil moisture and only the filled values in the GAPFILLED product, separated by their Euclidean distance (ED) to the nearest observation: small gaps (ED < 2) and medium/large gaps (ED ≥ 2). The shown metrics are **(a)** Pearson's R (only $p < 0.05$) and **(b)** ubRMSD. Narrow boxes indicate the upper and lower boundary values of the 95 % confidence interval for each data point. The numbers in each box refer to the sample size (number of time series).

8 Discussion and conclusions

The gap-filling framework around the DCT-PLS algorithm effectively estimates soil moisture for data gaps in the ESA CCI SM dataset. Choosing a univariate algorithm ensures that the resulting product relies solely on soil moisture observations independent of external satellite or model datasets. Additionally, DCT-PLS is a performant algorithm suitable for operational data production as envisaged for the Copernicus Climate Change Service (C3S) (Dorigo et al., 2024b).

We incorporated additional soil moisture retrievals from L- and C-band in equatorial latitudes with dense vegetation cover as anchor measurements for the interpolation algorithm. Although these measurements are often noisy and therefore justifiably excluded from the original ESA CCI SM data, they serve as viable support points for gap filling, leading to a moderate agreement with available in situ measurements in the area. However, we found that the retrieval uncertainty estimates based on triple collocation analysis for these areas are often too low, which can also affect the (propagated) uncertainties of the GAPFILLED product. We conclude that not only the retrieval of soil moisture under very dense vegetation layers remains challenging, but also that the estimation of associated uncertainties warrants further research. Studies utilizing these retrieval uncertainties, including ours, could apply them to provide better estimates of the quality of derived products based on these data.

Linear interpolation during periods when soil moisture is frozen may appear simplistic but follows the logic that there is no loss of water from the soil resulting in short-term changes. We use the average of observations from 30 d before and after periods when soil moisture is frozen as bounding points for linear interpolation. The threshold was determined based on regional experiments with satellite and in situ data and was selected as the best compromise between stable interpolation and a smooth transition between interpolated and observed data. However, it could be further refined, for example, by considering different onset periods for different climates globally. Our approach has the drawback that it relies on external soil temperature information, which can introduce outliers in the time series when prolonged periods of frozen soils ($< 0\,°C$) are interrupted by brief thawing events. This issue could be mitigated by adopting a slightly stricter threshold value (e.g. $4\,°C$, in line with Gruber et al., 2020) and/or by setting a minimum number of consecutive days above $0\,°C$ before a data point is classified as "unfrozen". Moreover, our method does not account for potential sub-pixel freeze–thaw dynamics. Further research is needed to better understand soil moisture dynamics at coarse scales in such cases, enabling satellite retrievals and improved gap-filling predictions in the future. In the final data files, we provide a separate flag to identify cases where linear interpolation was applied so that users can filter them out or replace them if necessary.

To provide users with a quality estimate for the gap-filled values, we developed a method to quantify the uncertainties in our predictions ($\sigma_{\text{gapfilling}}$). These uncertainties are generally higher for larger gaps and regions with dense vegetation compared to areas with better initial coverage. They are combined with the observation uncertainties and included as a separate field in the dataset. In order to quantify uncertainties associated with the gap-filling process, we used 3-dimensional Euclidean distances to the nearest valid data point as an estimate of gap size, but in practice, it could matter whether this point is close temporally or spatially due to different levels of soil moisture autocorrelation in these dimensions (Piles et al., 2022). Evaluating these uncertainties remains challenging due to limited reference data. While correlation with in situ measurements decreases for larger gaps, this could not be confirmed in terms of ubRMSD. We do not consider the current uncertainty estimates final and encourage further research on this topic. Methods such as Gaussian process regression (Gelfand and Schliep, 2016) have been used to fill measurement gaps in time series data in the past (De Caro et al., 2023). They provide uncertainty estimates based on the variance of multiple predictions but for the same reason were also found to be computationally too demanding (Heaton et al., 2019) for operational application in a global record such as ESA CCI and C3S SM.

We observed that the GAPFILLED product performs similarly to the original observations for the absolute values (including the climatological signals) and short-term fluctuations (anomalies). In some cases, the filled values match better with independent reference data than the original observations, likely due to the smoothing effect of DCT-PLS. Similar outcomes have been observed with other smoothing algorithms, such as the exponential filter model, which is used to

estimate root-zone soil moisture from surface measurements (Pasik et al., 2023; Wagner et al., 1999; Albergel et al., 2008). These conclusions are based on evaluations using in situ data and from successfully restoring artificially introduced gaps in the gap-free GLDAS Noah dataset. One downside of the latter approach is that model and satellite soil moisture can differ, for example, in terms of noise level, representation of extreme events, or autocorrelation. This can be due to limited model representations, for example, due to disregarded irrigation signals (Zaussinger et al., 2019) or unaccounted for latent water influx from rivers (van der Schalie et al., 2022). Thus, conclusions drawn from the restoration of model-based data may not fully apply to satellite observations in some regions, but they still give valuable insights. Regional parameterization of the DCT-PLS algorithm resulted in spatial variations in s. The use of GCV score optimization for s (Garcia, 2010) follows the recommendation of Wang et al. (2012). While the impact on the predictive skill compared to using a single global s value was not extensively evaluated, we found a good global performance in restoring artificially excluded data.

The GAPFILLED dataset is suitable for both long-term and event-based studies. Our comparisons, as well as other studies, have shown that derived statistics (such as changes in annual anomalies) can be affected by gap filling (Bessenbacher et al., 2023). However, in complex environments with little or no input observations, such as mountainous regions, minimal variability is observed in our data. Similarly, using a smoothing algorithm such as DCT-PLS for interpolation may be detrimental in terms of preserving small-scale extreme events. For applications that do not require a pure soil-moisture-based dataset, multivariate methods might therefore still be preferred. These studies could potentially use our data as starting values for further adjustment.
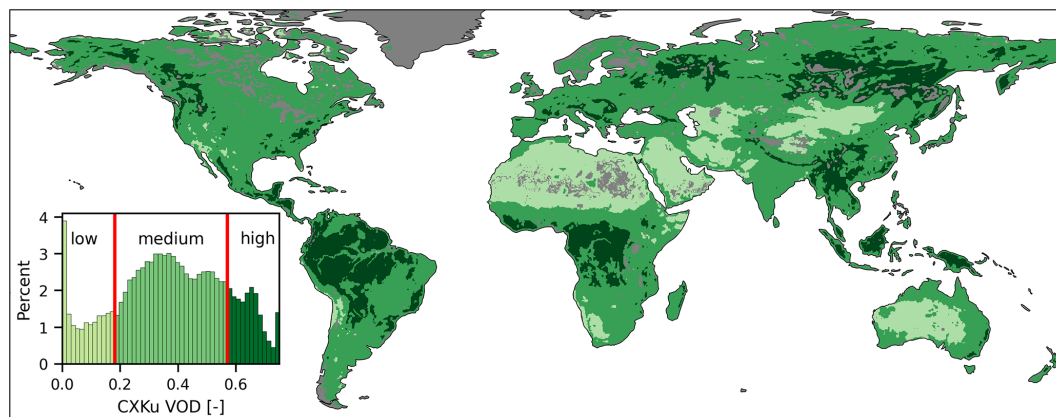
Appendix A

Figure A1. VOD classification based on VODCA v2 CXKu multiband 2000–2023 average conditions. Threshold values are 0.18 for low/medium and 0.57 for medium/high VOD ($\mu \pm 1\sigma$).
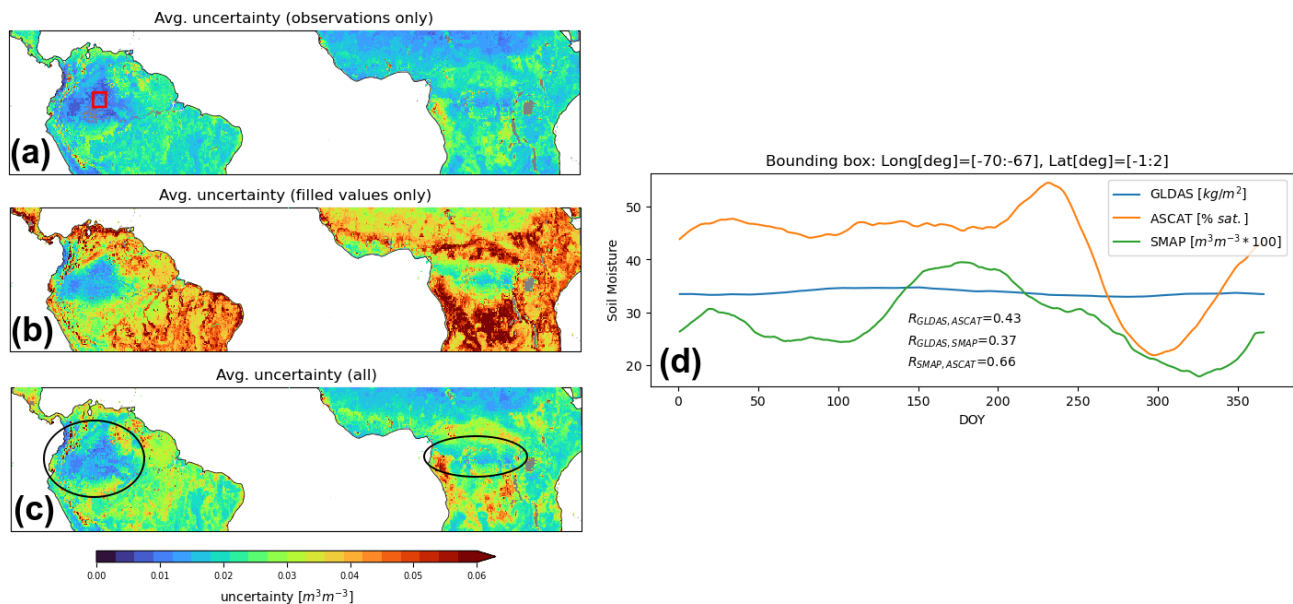
Figure A2. 1991–2023 average uncertainty estimates in parts of the tropical climate zone of **(a)** satellite observations used for the GAP-FILLED product (based on triple collocation analysis), **(b)** the filled (propagated) values in the GAPFILLED product only, and **(c)** the final GAPFILLED product overall (circles indicate areas with uncertainty estimates that are too low). Panel **(d)** shows the mean soil moisture climatologies of SMAP, ASCAT, and GLDAS Noah over the area outlined by the red bounding box in **(a)**.
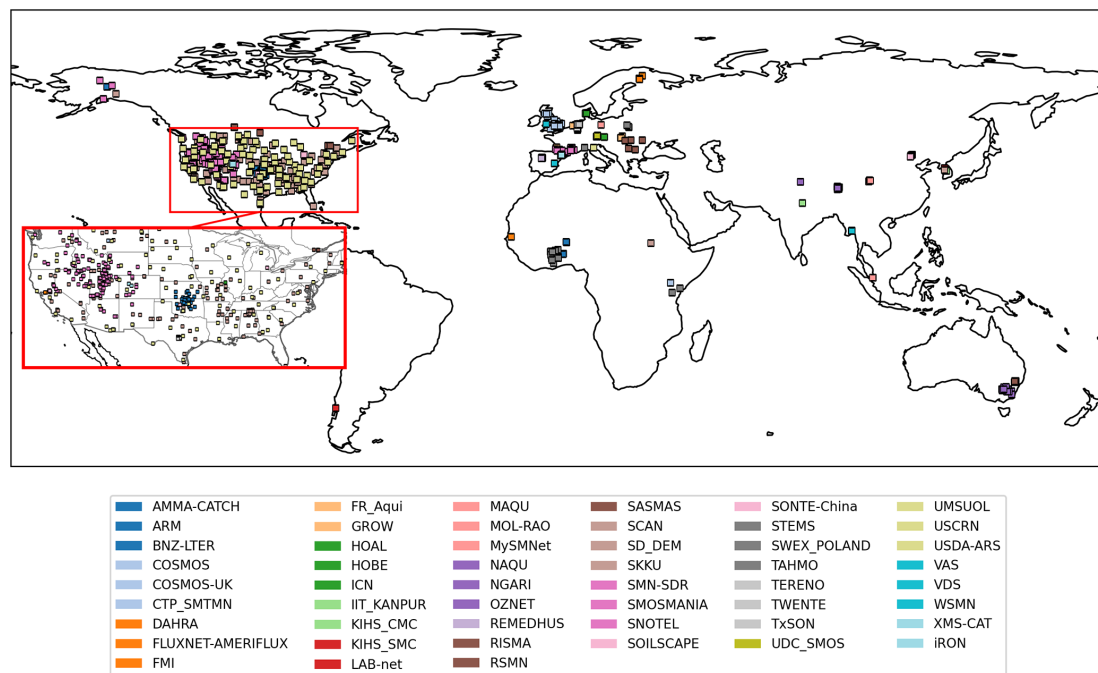

Figure A3. ISMN Fiducial reference measurements (0–0.1 m depth) in ISMN v20240314 on http://qa4sm.eu (last access: 20 August 2025) (network references in Table A2 in the Appendix).

Table A1. ISMN sensor locations and properties from Fig. 7.

	Network Name	Station Name	Sensor Type	Latitude [°]	Longitude [°]	Elevation [m]	Depth [cm]	Land Cover Class	Climate Class
(a)	TERENO	Schoeneseiffen	Hydraprobe-II-Sdi-12	50.5149	6.37559	612	5	130	Cfb
(b)	RISMA	SK2	Hydraprobe-II-Sdi-12	51.33504	−106.5639	590	0–5	130	Dfb
(c)	SCAN	Stephenville	Hydraprobe-Digital-Sdi-12-(2.5-Volt)	32.25	−98.2	407.896	5.08	130	Cfa
(d)	TAHMO	BiaSHTSDebiso	TEROS12	6.66462	−3.09725	228.0	10	130	Aw
(e)	NGARI	SQ08	5TM	32.55603	79.84004	4308.0	5	200	Dsb

Table A2. Soil moisture FRM networks available through QA4SM.

Network	Reference
AMMA-CATCH	Pellarin et al. (2009), Mougin et al. (2009), Cappelaere et al. (2009), Rosnay et al. (2009), Lebel et al. (2009), Galle et al. (2015)
ARM	Cook (2016, 2018)
COSMOS	Zreda et al. (2008, 2012)
CTP_SMTMN	Yang et al. (2013)
DAHRA	Tagesson et al. (2014)
FLUXNET-AMERIFLUX	–
FMI	Ikonen et al. (2018, 2016)
FR_Aqui	Al-Yaari et al. (2018), Wigneron et al. (2018)
GROW	Xaver et al. (2020), Zappa et al. (2019, 2020)
HOAL	Blöschl et al. (2016), Vreugdenhil et al. (2013)
HOBE	Bircher et al. (2012), Jensen and Refsgaard (2018)
ICN	Hollinger and Isard (1994)
IMA_CAN1	Biddoccu et al. (2016), Capello et al. (2019), Raffelli et al. (2017)
IPE	Alday et al. (2020)
KIHS_CMC	–
KIHS_SMC	–
LAB-net	Mattar et al. (2016, 2014)
MAQU	Su et al. (2011), Dente et al. (2012)
MOL-RAO	Beyrich and Adam (2007)
NAQU	Su et al. (2011), Dente et al. (2012)
NGARI	Su et al. (2011), Dente et al. (2012)
OZNET	Smith et al. (2012), Young et al. (2008)
PTSMN	Hajdu et al. (2019)
REMEDHUS	González-Zamora et al. (2019)
RISMA	Ojo et al. (2015), L'Heureux (2011), Canisius (2011)
RSMN	–
SASMAS	Rüdiger et al. (2007)
SCAN	Schaefer et al. (2007)
SD_DEM	Ardö (2013)
SMN-SDR	Zhao et al. (2020), Zheng et al. (2022)
SMOSMANIA	Calvet et al. (2016), Albergel et al. (2008), Calvet et al. (2007)
SNOTEL	Leavesley et al. (2008), Leavesley (2010)
SOILSCAPE	Moghaddam et al. (2011, 2016), Shuman et al. (2010)
SWEX_POLAND	Marczewski et al. (2010)
TAHMO	–
TERENO	Zacharias et al. (2011), Bogena et al. (2018, 2012), Bogena (2016)
UDC_SMOS	Schlenz et al. (2012), Loew et al. (2009)
UMBRIA	Brocca et al. (2011, 2008, 2009)
UMSUOL	–
USCRN	Bell et al. (2013)
VAS	–
VDS	–
iRON	Osenga et al. (2021, 2019)

References

Al-Yaari, A., Dayau, S., Chipeaux, C., Aluome, C., Kruszewski, A., Loustau, D., and Wigneron, J.-P.: The AQUI Soil Moisture Network for Satellite Microwave Remote Sensing Validation in South-Western France, Remote Sens., 10, 1839, https://doi.org/10.3390/rs10111839, 2018.

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., and Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, Hydrol. Earth Syst. Sci., 12, 1323–1337, https://doi.org/10.5194/hess-12-1323-2008, 2008.

Alday, J. G., Camarero, J. J., Revilla, J., and Resco de Dios, V.: Similar diurnal, seasonal and annual rhythms in radial root expansion across two coexisting Mediterranean oak species, Tree Physiol., 40, 956–968, https://doi.org/10.1093/treephys/tpaa041, 2020.

Almendra-Martín, L., Martínez-Fernández, J., Piles, M., and Ángel González-Zamora: Comparison of gap-filling techniques applied to the CCI soil moisture database in Southern Europe, Remote Sens. Environ., 258, 112377, https://doi.org/10.1016/j.rse.2021.112377, 2021.

Amankwah, S. K., Ireson, A. M., Maulé, C., Brannen, R., and Mathias, S. A.: A Model for the Soil Freezing Characteristic Curve That Represents the Dominant Role of Salt Exclusion, Water Resour. Res., 57, e2021WR030070, https://doi.org/10.1029/2021WR030070, 2021.

Ardö, J.: A 10-Year Dataset of Basic Meteorology and Soil Properties in Central Sudan, Dataset Papers in Geosciences [data set], 2013, https://doi.org/10.7167/2013/297973/dataset, 2013.

Bell, J., Palecki, M., Baker, B., Collins, W., Lawrimore, J., Leeper, R., Hall, M., Kochendorfer, J., Meyers, T., Wilson, T., and Diamond, H.: U.S. Climate Reference Network Soil Moisture and Temperature Observations, J. Hydrometeorol., 14, 977–988, https://doi.org/10.1175/JHM-D-12-0146.1, 2013.

Bessenbacher, V., Seneviratne, S. I., and Gudmundsson, L.: CLIMFILL v0.9: a framework for intelligently gap filling Earth observations, Geosci. Model Dev., 15, 4569–4596, https://doi.org/10.5194/gmd-15-4569-2022, 2022.

Bessenbacher, V., Schumacher, D. L., Hirschi, M., Seneviratne, S. I., and Gudmundsson, L.: Gap-Filled Multivariate Observations of Global Land–Climate Interactions, J. Geophys. Res.-Atmos., 128, e2023JD039099, https://doi.org/10.1029/2023JD039099, 2023.

Beyrich, F. and Adam, W.: Site and Data Report for the Lindenberg Reference Site in CEOP – Phase 1, Berichte des Deutschen Wetterdienstes, 230, Offenbach am Main, https://www.cen.uni-hamburg.de/en/icdc/data/atmosphere/docs-atmo/dwd230-ceop-report.pdf (last access: 20 August 2025), 2007.

Biddoccu, M., Ferraris, S., Opsi, F., and Cavallo, E.: Long-term monitoring of soil management effects on runoff and soil erosion in sloping vineyards in Alto Monferrato (North West Italy), Soil Till. Res., 155, 176–189, https://doi.org/10.1016/j.still.2015.07.005, 2016.

Bircher, S., Skou, N., Jensen, K. H., Walker, J. P., and Rasmussen, L.: A soil moisture and temperature network for SMOS validation in Western Denmark, Hydrol. Earth Syst. Sci., 16, 1445–1463, https://doi.org/10.5194/hess-16-1445-2012, 2012.

Blöschl, G., Blaschke, A. P., Broer, M., Bucher, C., Carr, G., Chen, X., Eder, A., Exner-Kittridge, M., Farnleitner, A., Flores-Orozco, A., Haas, P., Hogan, P., Kazemi Amiri, A., Oismüller, M., Parajka, J., Silasari, R., Stadler, P., Strauss, P., Vreugdenhil, M., Wagner, W., and Zessner, M.: The Hydrological Open Air Laboratory (HOAL) in Petzenkirchen: a hypothesis-driven observatory, Hydrol. Earth Syst. Sci., 20, 227–255, https://doi.org/10.5194/hess-20-227-2016, 2016.

Bogena, H., Kunkel, R., Pütz, T., Vereecken, H., Kruger, E., Zacharias, S., Dietrich, P., Wollschläger, U., Kunstmann, H., Papen, H., Schmid, H., Munch, J., Priesack, E., Schwank, M., Bens, O., Brauer, A., Borg, E., and Hajnsek, I.: TERENO – Long-term monitoring network for terrestrial environmental research, Hydrol. Wasserbewirts., 56, 138–143, 2012.

Bogena, H., Montzka, C., Huisman, J., Graf, A., Schmidt, M., Stockinger, M., von Hebel, C., Hendricks-Franssen, H., van der Kruk, J., Tappe, W., Lücke, A., Baatz, R., Bol, R., Groh, J., Pütz, T., Jakobi, J., Kunkel, R., Sorg, J., and Vereecken, H.: The TERENO-Rur Hydrological Observatory: A Multiscale Multi-Compartment Research Platform for the Advance-

https://doi.org/10.5194/essd-17-4305-2025

Earth Syst. Sci. Data, 17, 4305–4329, 2025

ment of Hydrological Science, Vadose Zone J., 17, 180055, https://doi.org/10.2136/vzj2018.03.0055, 2018.

Bogena, H. R.: TERENO: German network of terrestrial environmental observatories, Journal of large-scale research facilities, 2, A52, https://doi.org/10.17815/jlsrf-2-98, 2016.

Brocca, L., Melone, F., and Moramarco, T.: On the estimation of antecedent wetness condition in rainfall-runoff modeling, Hydrol. Process., 22, 629–642, https://doi.org/10.1002/hyp.6629, 2008.

Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R.: Antecedent wetness conditions based on ERS scatterometer data, J. Hydrol., 364, 73–87, 2009.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., A, D., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., and Bittelli, M.: Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe, Remote Sens. Environ., 115, 3390–3408, https://doi.org/10.1016/j.rse.2011.08.003, 2011.

Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., and Piguet, B.: In situ soil moisture observations for the CAL/VAL of SMOS: the SMOSMANIA network, in: 2007 IEEE International Geoscience and Remote Sensing Symposium, 1196–1199, https://doi.org/10.1109/IGARSS.2007.4423019, 2007.

Calvet, J.-C., Fritz, N., Berne, C., Piguet, B., Maurel, W., and Meurey, C.: Deriving pedotransfer functions for soil quartz fraction in southern France from reverse modeling, SOIL, 2, 615–629, https://doi.org/10.5194/soil-2-615-2016, 2016.

Canisius, F.: Calibration of Casselman, Ontario Soil Moisture Monitoring Network, Agriculture and Agri-Food Canada, Ottawa, ON, 37 pp., 2011.

Capello, G., Biddoccu, M., Ferraris, S., and Cavallo, E.: Effects of Tractor Passes on Hydrological and Soil Erosion Processes in Tilled and Grassed Vineyards, Water, 11, 2118, https://doi.org/10.3390/w11102118, 2019.

Cappelaere, B., Descroix, L., Lebel, T., Boulain, N., Ramier, D., Laurent, J.-P., Favreau, G., Boubkraoui, S., Boucher, M., Moussa, I., Chaffard, V., Hiernaux, P., Issoufou, H. B.-A., Breton, E., Mamadou, I., Nazoumou, Y., Oi, M., Ottle, C., and Quantin, G.: The AMMA-CATCH experiment in the cultivated Sahelian area of south-west Niger , Investigating water cycle response to a fluctuating climate and changing environment, J. Hydrol., 375, 34–51, https://doi.org/10.1016/j.jhydrol.2009.06.021, 2009.

Cook, D.: Surface Energy Balance System (SEBS) Instrument Handbook, https://doi.org/10.2172/1004944, 2018.

Cook, D. R.: Soil Temperature and Moisture Profile (STAMP) System Handbook, https://doi.org/10.2172/1332724, 2016.

De Caro, D., Ippolito, M., Cannarozzo, M., Provenzano, G., and Ciraolo, G.: Assessing the performance of the Gaussian Process Regression algorithm to fill gaps in the time-series of daily actual evapotranspiration of different crops in temperate and continental zones using ground and remotely sensed data, Agr. Water Manage., 290, 108596, https://doi.org/10.1016/j.agwat.2023.108596, 2023.

Dente, L., Su, Z., and Wen, J.: Validation of SMOS soil moisture products over the Maqu and Twente regions, Sensors, 12, 9965–9986, 2012.

Dorigo, W., de Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., and Fernández-Prieto, D.: Evaluating global trends (1988–2010) in harmonized multi-satellite sur-

face soil moisture, Geophys. Res. Lett., 39, L18405, https://doi.org/10.1029/2012GL052988, 2012.

Dorigo, W., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A., Zamojski, D., Cordes, C., Wagner, W., and Drusch, M.: Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network, Vadose Zone J., 12, vzj2012.0097, https://doi.org/10.2136/vzj2012.0097, 2013.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sens. Environ., 203, 185–215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.

Dorigo, W., Dietrich, S., Aires, F., Brocca, L., Carter, S., Cretaux, J.-F., Dunkerley, D., Enomoto, H., Forsberg, R., Güntner, A., Hegglin, M. I., Hollmann, R., Hurst, D. F., Johannessen, J. A., Kummerow, C., Lee, T., Luojus, K., Looser, U., Miralles, D. G., Pellet, V., Recknagel, T., Vargas, C. R., Schneider, U., Schoeneich, P., Schröder, M., Tapper, N., Vuglinsky, V., Wagner, W., Yu, L., Zappa, L., Zemp, M., and Aich, V.: Closing the Water Cycle from Observations across Scales: Where Do We Stand?, B. Am. Meteorol. Soc., 102, E1897–E1935, https://doi.org/10.1175/BAMS-D-19-0316.1, 2021a.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.-C., Camarero, J. J., Capello, G., Choi, M., Cosh, M. C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K. H., Kanniah, K. D., de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J. P., Osenga, E., Palecki, M. A., Pellarin, T., Petropoulos, G. P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J. P., Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., and Sabia, R.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci., 25, 5749–5804, https://doi.org/10.5194/hess-25-5749-2021, 2021b.

Dorigo, W., Preimesberger, W., Hahn, S., Van der Schalie, R., De Jeu, R., Kidd, R., Rodriguez-Fernandez, N., Hirschi, M., Stradiotti, P., Frederikse, T., Gruber, A., and Duchemin, D.: ESA Soil Moisture Climate Change Initiative (Soil_Moisture_cci): COMBINED product, Version 09.1, CEDA Archive [data set], https://doi.org/10.5285/0e346e1e1e164ac99c60098848537a29, 2024a.

Dorigo, W., Preimesberger, W., Reimer, C., Van der Schalie, R., Pasik, A., De Jeu, R., and Paulik, C.: Soil moisture gridded data from 1978 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.d7782f18, 2024b.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oeve-

len, P., Robock, A., and Jackson, T.: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, Hydrol. Earth Syst. Sci., 15, 1675–1698, https://doi.org/10.5194/hess-15-1675-2011, 2011.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Van Zyl, J.: The Soil Moisture Active Passive (SMAP) Mission, Proc. IEEE, 98, 704–716, https://doi.org/10.1109/JPROC.2010.2043918, 2010.

Fan, C., Liu, K., Luo, S., Chen, T., Cheng, J., Zhan, P., and Song, C.: Detection of surface water temperature variations of Mongolian lakes benefiting from the spatially and temporally gap-filled MODIS data, Int. J. Appl. Earth Obs., 114, 103073, https://doi.org/10.1016/j.jag.2022.103073, 2022.

Galle, S., Grippa, M., Peugeot, C., Bouzou Moussa, I., Cappelaere, B., Demarty, J., Mougin, E., Lebel, T., and Chaffard, V.: AMMA-CATCH a Hydrological, Meteorological and Ecological Long Term Observatory on West Africa: Some Recent Results, in: AGU Fall Meeting Abstracts, vol. 2015, GC42A–01, 2015.

Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values, Comput. Stat. Data Anal., 54, 1167–1178, https://doi.org/10.1016/j.csda.2009.09.020, 2010.

GCOS: The 2022 GCOS ECVs Requirements, World Meteorological Organisation, 245, https://library.wmo.int/idurl/4/58111 (last access: 20 August 2025), 2022.

Gelfand, A. E. and Schliep, E. M.: Spatial statistics and Gaussian processes: A beautiful marriage, Spatial Stat., 18, 86–104, https://doi.org/10.1016/j.spasta.2016.03.006, 2016.

González-Zamora, Á., Sánchez, N., Pablos, M., and Martínez-Fernández, J.: CCI soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain, Remote Sens. Environ., 225, 469–482, https://doi.org/10.1016/j.rse.2018.02.010, 2019.

Goryl, P., Fox, N., Donlon, C., and Castracane, P.: Fiducial Reference Measurements (FRMs): What Are They?, Remote Sens., 15, 5017, https://doi.org/10.3390/rs15205017, 2023.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, Int. J. Appl. Earth Obs., 45, 200–211, https://doi.org/10.1016/j.jag.2015.09.002, 2016.

Gruber, A., Dorigo, W. A., Crow, W., and Wagner, W.: Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals, IEEE T. Geosci. Remote, 55, 6780–6792, https://doi.org/10.1109/TGRS.2017.2734070, 2017.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W.: Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology, Earth Syst. Sci. Data, 11, 717–739, https://doi.org/10.5194/essd-11-717-2019, 2019.

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.-C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., van der Schalie, R., Wigneron, J.-P., and Wagner, W.: Validation practices for satellite soil moisture retrievals: What are (the) errors?, Remote Sens. En-

viron., 244, 111806, https://doi.org/10.1016/j.rse.2020.111806, 2020.

Guo, X., Fang, X., Cao, Y., Yang, L., Ren, L., Chen, Y., and Zhang, X.: Reconstruction of ESA CCI soil moisture based on DCT-PLS and in situ soil moisture, Hydrol. Res., 53, 1221–1236, https://doi.org/10.2166/nh.2022.058, 2022.

H SAF: Product User Manual, Metop ASCAT Surface Soil Moisture Climate Data Record V7 12.5 Km Sampling (H119) and Extension (H120), V1.2, 2022, H SAF, https://doi.org/10.15770/EUM_SAF_H_0009, 2022.

Hahn, S., Wagner, W., Steele-Dunne, S. C., Vreugdenhil, M., and Melzer, T.: Improving ASCAT Soil Moisture Retrievals With an Enhanced Spatially Variable Vegetation Parameterization, IEEE T. Geosci. Remote, 59, 8241–8256, https://doi.org/10.1109/TGRS.2020.3041340, 2021.

Hajdu, I., Yule, I., Bretherton, M., Singh, R., and Hedley, C.: Field performance assessment and calibration of multi-depth AquaCheck capacitance-based soil moisture probes under permanent pasture for hill country soils, Agr. Water Managem., 217, 332–345, https://doi.org/10.1016/j.agwat.2019.03.002, 2019.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A.: A Case Study Competition Among Methods for Analyzing Large Spatial Data, J. Agr. Biol. Environ. Stat., 24, 398–425, https://doi.org/10.1007/s13253-018-00348-w, 2019.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Himmelbauer, I., Gruber, A., and Crapolicchio, R.: Fiducial Reference Measurements for Soil Moisture (FRM4SM), Technical report, contract No. 4000135204/21//I-BG. DT2-1, European Space Agency, Technische Universitaet Wien Wiedner Hauptstrasse 8 to 10/E120, 1040 Vienna, Austria, https://earth.esa.int/documents/d/earth-online/frm4sm__dt2-1_fpp_sm_v04 (last access: 20 August 2025), 2023.

Hollinger, S. and Isard, S.: A Soil Moisture Climatology of Illinois, J. Climate, 7, 822–833, https://doi.org/10.1175/1520-0442(1994)007<0822:ASMCOI>2.0.CO;2, 1994.

Ikonen, J., Vehviläinen, J., Rautiainen, K., Smolander, T., Lemmetyinen, J., Bircher, S., and Pulliainen, J.: The Sodankylä in situ soil moisture observation network: an example application of ESA CCI soil moisture product evaluation, Geosci. Instrum. Method. Data Syst., 5, 95–108, https://doi.org/10.5194/gi-5-95-2016, 2016.

Ikonen, J., Smolander, T., Rautiainen, K., Cohen, J., Lemmetyinen, J., Salminen, M., and Pulliainen, J.: Spatially distributed evaluation of ESA CCI Soil Moisture products in a northern boreal forest environment, Geosciences, 8, 51, https://doi.org/10.3390/geosciences8020051, 2018.

Jackson, T. J., Schmugge, T. J., and Wang, J. R.: Passive microwave sensing of soil moisture under vegetation canopies, Water Resour. Res., 18, 1137–1142, https://doi.org/10.1029/WR018i004p01137, 1982.

Jensen, K. H. and Refsgaard, J. C.: HOBE: The Danish Hydrological Observatory, Vadose Zone J., 17, 180059, https://doi.org/10.2136/vzj2018.03.0059, 2018.

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., Juglea, S. E., Drinkwater, M. R., Hahne, A., Martín-Neira, M., and Mecklenburg, S.: The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle, Proc. IEEE, 98, 666–687, https://doi.org/10.1109/JPROC.2010.2043032, 2010.

Kongkulsiri, S., Tomkratoke, S., and Sirisup, S.: DCT-PLS Gap Filling and DMD Analysis of the Coastal Ocean Surface Current in the Gulf of Thailand, in: 2018 OCEANS – MTS/IEEE Kobe Techno-Oceans (OTO), 1–6, https://doi.org/10.1109/OCEANSKOBE.2018.8559274, 2018.

Kovačević, J., Cvijetinović, Ž., Stančić, N., Brodić, N., and Mihajlović, D.: New Downscaling Approach Using ESA CCI SM Products for Obtaining High Resolution Surface Soil Moisture, Remote Sens., 12, 1119, https://doi.org/10.3390/rs12071119, 2020.

Leavesley, G. H.: A Modelling Framework for Improved Agricultural Water-Supply Forecasting, 2010.

Leavesley, G. H., David, O., Garen, D. C., Lea, J., Marron, J. K., Pagano, T. C., Perkins, T. R., and Strobel, M. L.: A Modeling Framework for Improved Agricultural Water Supply Forecasting, American Geophysical Union, Fall Meeting 2008, abstract id. C21A-0497, 2008AGUFM.C21A0497L, https://ui.adsabs.harvard.edu/abs/2008AGUFM.C21A0497L (last access: 20 August 2025), 2008.

Lebel, T., Cappelaere, B., Galle, S., Hanan, N., Kergoat, L., Levis, S., Vieux, B., Descroix, L., Gosset, M., Mougin, E., Peugeot, C., and Seguis, L.: AMMA-CATCH studies in the Sahelian region of West-Africa: an overview, J. Hydrol., 375, 3–13, https://doi.org/10.1016/j.jhydrol.2009.03.020, 2009.

L'Heureux, J.: 2011 Installation Report for AAFC-SAGES Soil Moisture Stations in Kenaston, SK, Agriculture and Agri-Food Canada, Regina, SK, 2011.

Liu, H., Lu, N., Jiang, H., Qin, J., and Yao, L.: Filling Gaps of Monthly Terra/MODIS Daytime Land Surface Temperature Using Discrete Cosine Transform Method, Remote Sens., 12, 361, https://doi.org/10.3390/rs12030361, 2020a.

Liu, K., Li, X., Wang, S., and Zhang, H.: A robust gap-filling approach for European Space Agency Climate Change Initiative (ESA CCI) soil moisture integrating satellite observations, model-driven knowledge, and spatiotemporal machine learning, Hydrol. Earth Syst. Sci., 27, 577–598, https://doi.org/10.5194/hess-27-577-2023, 2023.

Liu, Y., Dorigo, W., Parinussa, R., de Jeu, R., Wagner, W., McCabe, M., Evans, J., and van Dijk, A.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sens. Environ., 123, 280–297, https://doi.org/10.1016/j.rse.2012.03.014, 2012.

Liu, Y., Yang, Y., and Jing, W.: Potential Applicability of SMAP in ECV Soil Moisture Gap-Filling: A Case Study in Europe, IEEE Access, 8, 133114–133127, https://doi.org/10.1109/ACCESS.2020.3009977, 2020b.

Liu, Y., Yao, L., Jing, W., Di, L., Yang, J., and Li, Y.: Comparison of two satellite-based soil moisture reconstruction algorithms: A case study in the state of Oklahoma, USA, J. Hydrol., 590, 125406, https://doi.org/10.1016/j.jhydrol.2020.125406, 2020c.

Llamas, R. M., Guevara, M., Rorabaugh, D., Taufer, M., and Vargas, R.: Spatial Gap-Filling of ESA CCI Satellite-Derived Soil Moisture Based on Geostatistical Techniques and Multiple Regression, Remote Sens., 12, 665, https://doi.org/10.3390/rs12040665, 2020.

Loew, A., Dall'Amico, J. T., Schlenz, F., and Mauser, W.: The Upper Danube Soil Moisture Validation Site: Measurements and Activities, in: Earth Observation and Water Cycle Science, edited by: Lacoste, H., ESA Special Publication, 674, p. 56, https://ui.adsabs.harvard.edu/abs/2009ESASP.674E..56L (last access: 20 August 2025), 2009.

Mao, H., Kathuria, D., Duffield, N., and Mohanty, B. P.: Gap Filling of High-Resolution Soil Moisture for SMAP/Sentinel-1: A Two-Layer Machine Learning-Based Framework, Water Resour. Res., 55, 6986–7009, https://doi.org/10.1029/2019WR024902, 2019.

Marczewski, W., Slominski, J., Slominska, E., Usowicz, B., Usowicz, J., Romanov, S., Maryskevych, O., Nastula, J., and Zawadzki, J.: Strategies for validating and directions for employing SMOS data, in the Cal-Val project SWEX (3275) for wetlands, Hydrol. Earth Syst. Sci. Discuss., 7, 7007–7057, https://doi.org/10.5194/hessd-7-7007-2010, 2010.

Mattar, C., Santamaría-Artigas, A., Durán-Alarcón, C., Olivera-Guerra, L., and Fuster, R.: LAB-net: the first Chilean soil moisture network for remote sensing applications, in: Quantitative Remote Sensing Symposium (RAQRS), 22–26, 2014.

Mattar, C., Santamaría-Artigas, A., Durán-Alarcón, C., Olivera-Guerra, L., Fuster, R., and Borvarán, D.: The LAB-Net Soil Moisture Network: Application to Thermal Remote Sensing and Surface Energy Balance, Data, 1, 6, https://doi.org/10.3390/data1010006, 2016.

Moesinger, L., Dorigo, W., de Jeu, R., van der Schalie, R., Scanlon, T., Teubner, I., and Forkel, M.: The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA), Earth Syst. Sci. Data, 12, 177–196, https://doi.org/10.5194/essd-12-177-2020, 2020.

Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., and Teneketzis, D.: A Wireless Soil Moisture Smart Sensor Web Using Physics-Based Optimal Control: Concept and Initial Demonstrations, IEEE J. Sel. Top. Appl. Earth Obs., 3, 522–535, https://doi.org/10.1109/JSTARS.2010.2052918, 2011.

Moghaddam, M., Silva, A., Clewley, D., Akbar, R., Hussaini, S., Whitcomb, J., Devarakonda, R., Shrestha, R., Cook, R., Prakash, G., Santhana Vannan, S., And Boyer, A.: Soil Moisture Profiles and Temperature Data from SoilSCAPE Sites, USA, NASA Oak Ridge National Laboratory Distributed Active Archive Center [data set], https://doi.org/10.3334/ORNLDAAC/1339, 2016.

Mougin, E., Hiernaux, P., Kergoat, L., Manuela, G., Rosnay, P., Timouk, F., Le Dantec, V., Demarez, V., Lavenu, F., Arjounin, M., Lebel, T., Soumaguel, N., Ceschia, E., Mougenot, B., Baup, F., Frappart, F., Frison, P.-L., Gardelle, J., Gruhier, C., and Mazzega, P.: The AMMA-CATCH Gourma observatory site in Mali: Relating climatic variations to changes in vegetation, surface hydrology, fluxes and natural resources, J. Hydrol., 375, 14–33, https://doi.org/10.1016/j.jhydrol.2009.06.045, 2009.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Nadeem, A. A., Zha, Y., Shi, L., Ali, S., Wang, X., Zafar, Z., Afzal, Z., and Tariq, M. A. U. R.: Spatial Downscaling and Gap-Filling of SMAP Soil Moisture to High Resolution Using MODIS Surface Variables and Machine Learning Approaches over ShanDian River Basin, China, Remote Sens., 15, 812, https://doi.org/10.3390/rs15030812, 2023.

Naeimi, V., Paulik, C., Bartsch, A., Wagner, W., Kidd, R., Park, S.-E., Elger, K., and Boike, J.: ASCAT Surface State Flag (SSF): Extracting Information on Surface Freeze/Thaw Conditions From Backscatter Data Using an Empirical Threshold-Analysis Algorithm, IEEE T. Geosci. Remote, 50, 2566–2582, https://doi.org/10.1109/TGRS.2011.2177667, 2012.

Ojo, E. R., Bullock, P., L. Heureux, J., Powers, J., McNairn, H., and Pacheco, A.: Calibration and Evaluation of a Frequency Domain Reflectometry Sensor for Real-Time Soil Moisture Monitoring, Vadose Zone J., 14, vzj2014.08.0114, https://doi.org/10.2136/vzj2014.08.0114, 2015.

Oliva, R., Daganzo, E., Kerr, Y. H., Mecklenburg, S., Nieto, S., Richaume, P., and Gruhier, C.: SMOS Radio Frequency Interference Scenario: Status and Actions Taken to Improve the RFI Environment in the 1400–1427-MHz Passive Band, IEEE T. Geosci. Remote, 50, 1427–1439, https://doi.org/10.1109/TGRS.2012.2182775, 2012.

Osenga, E. C., Arnott, J. C., Endsley, K. A., and Katzenberger, J. W.: Bioclimatic and Soil Moisture Monitoring Across Elevation in a Mountain Watershed: Opportunities for Research and Resource Management, Water Resour. Res., 55, 2493–2503, https://doi.org/10.1029/2018WR023653, 2019.

Osenga, E. C., Vano, J. A., and Arnott, J. C.: A community-supported weather and soil moisture monitoring database of the Roaring Fork catchment of the Colorado River Headwaters, Hydrol. Process., 35, e14081, https://doi.org/10.1002/hyp.14081, 2021.

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

Parinussa, R. M., Meesters, A. G. C. A., Liu, Y. Y., Dorigo, W., Wagner, W., and de Jeu, R. A. M.: Error Estimates for Near-Real-Time Satellite Soil Moisture as Derived From the Land Parameter Retrieval Model, IEEE Geosci. Remote Sens. Lett., 8, 779–783, https://doi.org/10.1109/LGRS.2011.2114872, 2011.

Pasik, A., Gruber, A., Preimesberger, W., De Santis, D., and Dorigo, W.: Uncertainty estimation for a new exponential-filter-based long-term root-zone soil moisture dataset from Copernicus Climate Change Service (C3S) surface observations, Geosci. Model Dev., 16, 4957–4976, https://doi.org/10.5194/gmd-16-4957-2023, 2023.

Paulik, C., Preimesberger, W., s scherrer, pstradio, Hahn, S., Baum, D., Plocon, A., Mistelbauer, T., tracyscanlon, Schmitzer, M., alegrub88, daberer, and teije01: TUW-GEO/pytesmo: v0.18.0, Zenodo [code], https://doi.org/10.5281/zenodo.14975386, 2025.

Pellarin, T., Laurent, J.-P., Cappelaere, B., Decharme, B., Descroix, L., and Ramier, D.: Hydrological modelling and associated microwave emission of a semi-arid region in South-western Niger, J. Hydrol., 375, 262–272, https://doi.org/10.1016/j.jhydrol.2008.12.003, 2009.

Petropoulos, G. P., Ireland, G., and Barrett, B.: Surface soil moisture retrievals from remote sensing: Current status, products & future trends, Phys. Chem. Earth, 83–84, 36–56, https://doi.org/10.1016/j.pce.2015.02.009, 2015.

Pham, H. T., Kim, S., Marshall, L., and Johnson, F.: Using 3D robust smoothing to fill land surface temperature gaps at the continental scale, Int. J. Appl. Earth Obs. Geoinf., 82, 101879, https://doi.org/10.1016/j.jag.2019.05.012, 2019.

Piles, M., Muñoz-Marí, J., Guerrero-Curieses, A., Camps-Valls, G., and Rojo-Álvarez, J. L.: Autocorrelation Metrics to Estimate Soil Moisture Persistence From Satellite Time Series: Application to Semiarid Regions, IEEE T. Geosci. Remote, 60, 1–17, https://doi.org/10.1109/TGRS.2021.3057928, 2022.

Preimesberger, W., Scanlon, T., Su, C.-H., Gruber, A., and Dorigo, W.: Homogenization of Structural Breaks in the Global ESA CCI Soil Moisture Multisatellite Climate Data Record, IEEE T. Geosci. Remote, 59, 2845–2862, https://doi.org/10.1109/TGRS.2020.3012896, 2021.

Preimesberger, W., Stradiotti, P., and Dorigo, W. A.: ESA CCI SM GAPFILLED Long-term Climate Data Record of Surface Soil Moisture from merged multi- satellite observations, TU Wien Research Data [data set], https://doi.org/10.48436/hcm6n-t4m35, 2024.

Raffelli, G., Previati, M., Canone, D., Gisolo, D., Bevilacqua, I., Capello, G., Biddoccu, M., Cavallo, E., Deiana, R., Cassiani, G., and Ferraris, S.: Local- and Plot-Scale Measurements of Soil Moisture: Time and Spatially Resolved Field Techniques in Plain, Hill and Mountain Sites, Water, 9, W10403, https://doi.org/10.3390/w9090706, 2017.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, B. Am. Meteorol. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Rosnay, P., Gruhier, C., Timouk, F., Baup, F., Mougin, E., Hiernaux, P., Kergoat, L., and LeDantec, V.: Multi-scale soil moisture measurements at the Gourma meso-scale site in Mali, J. Hydrol., 375, 241–252, https://doi.org/10.1016/j.jhydrol.2009.01.015, 2009.

Rubin, D. B.: Inference and missing data, Biometrika, 63, 581–592, 1976.

Rüdiger, C., Hancock, G., Hemakumara, H., Jacobs, B., Kalma, J., Martinez, C., Thyer, M., Walker, J., Wells, T., and Willgoose, G.: Goulburn River experimental catchment data set, Water Resour. Res., 43, W10403, https://doi.org/10.1029/2006WR005837, 2007.

Schaefer, G., Cosh, M., and Jackson, T.: The USDA natural resources conservation service soil climate analysis network (SCAN), J. Atmos. Ocean. Tech., 24, 2073–2077, https://doi.org/10.1175/2007JTECHA930.1, 2007.

Schlenz, F., dall'Amico, J. T., Loew, A., and Mauser, W.: Uncertainty Assessment of the SMOS Validation in the Upper Danube Catchment, IEEE T. Geosci. Remote, 50, 1517–1529, 2012.

https://doi.org/10.5194/essd-17-4305-2025

Earth Syst. Sci. Data, 17, 4305–4329, 2025

Shangguan, Y., Min, X., and Shi, Z.: Gap Filling of the ESA CCI Soil Moisture Data Using a Spatiotemporal Attention-Based Residual Deep Network, IEEE J. Sel. Top. Appl. Earth Obs., 16, 5344–5354, https://doi.org/10.1109/JSTARS.2023.3284841, 2023.

Shuman, D. I., Nayyar, A., Mahajan, A., Goykhman, Y., Li, K., Liu, M., Teneketzis, D., Moghaddam, M., and Entekhabi, D.: Measurement Scheduling for Soil Moisture Sensing: From Physical Models to Optimal Control, Proc. IEEE, 98, 1918–1933, https://doi.org/10.1109/JPROC.2010.2052532, 2010.

Smith, A., Walker, J., Western, A., Young, R., Ellett, K., Pipunic, R., Grayson, R., Siriwardena, L., Chiew, F., and Richter, H.: The Murrumbidgee Soil Moisture Monitoring Network data set, Water Resour. Res., 48, W07701, https://doi.org/10.1029/2012WR011976, 2012.

Srivastava, P. K.: Satellite soil moisture: Review of theory and applications in water resources, Water Resour. Manage., 31, 3161–3176, 2017.

Steven, M. D., Malthus, T. J., Baret, F., Xu, H., and Chopping, M. J.: Intercalibration of vegetation indices from different sensor systems, Remote Sens. Environ., 88, 412–422, https://doi.org/10.1016/j.rse.2003.08.010, 2003.

Stradiotti, P., Gruber, A., Preimesberger, W., and Dorigo, W.: Accounting for seasonal retrieval errors in the merging of multisensor satellite soil moisture products, Sci. Remote Sens., 12, 100242, https://doi.org/10.1016/j.srs.2025.100242, 2025.

Strang, G.: The Discrete Cosine Transform, SIAM Review, 41, 135–147, https://doi.org/10.1137/S0036144598336745, 1999.

Su, Z., Wen, J., Dente, L., van der Velde, R., Wang, L., Ma, Y., Yang, K., and Hu, Z.: The Tibetan Plateau observatory of plateau scale soil moisture and soil temperature (Tibet-Obs) for quantifying uncertainties in coarse resolution satellite and model products, Hydrol. Earth Syst. Sci., 15, 2303–2316, https://doi.org/10.5194/hess-15-2303-2011, 2011.

Sun, H. and Xu, Q.: Evaluating Machine Learning and Geostatistical Methods for Spatial Gap-Filling of Monthly ESA CCI Soil Moisture in China, Remote Sens., 13, 2848, https://doi.org/10.3390/rs13142848, 2021.

Tagesson, T., Fensholt, R., Guiro, I., Rasmussen, M., Huber, S., Mbow, C., Garcia, M., Horion, S., Sandholt, I., Holm-Rasmussen, B., Göttsche, F.-M., Ridler, M., Boke-Olén, N., Olsen, J., Ehammer, A., Madsen, M., Olesen, F., and Ardö, J.: Ecosystem properties of semi-arid savanna grassland in West Africa and its relationship to environmental variability, Glob. Change Biol., 21, 250–264, https://doi.org/10.1111/gcb.12734, 2014.

Tong, C., Wang, H., Magagi, R., Goïta, K., and Wang, K.: Spatial Gap-Filling of SMAP Soil Moisture Pixels Over Tibetan Plateau via Machine Learning Versus Geostatistics, IEEE J. Sel. Top. Appl. Earth Obs., 14, 9899–9912, https://doi.org/10.1109/JSTARS.2021.3112623, 2021.

Ulaby, F. T., Long, D. G., Blackwell, W. J., Elachi, C., Fung, A. K., Ruf, C., Sarabandi, K., Zebker, H. A., and van Zyl, J.: Microwave Radar and Radiometric Remote Sensing, University of Michigan Press, Ann Arbor, USA, ISBN: 978-0-472-11935-6, 2014.

Uranga, E., Llorente, A., González, J., de la Fuente, A., Oliva, R., Soldo, Y., and Jorge, F.: SMOS ESA RFI Monitoring and Information Tool: Lessons Learned, Remote Sens., 14, 5387, https://doi.org/10.3390/rs14215387, 2022.

van der Schalie, R., Parinussa, R., Renzullo, L., van Dijk, A., Su, C.-H., and de Jeu, R.: SMOS soil moisture retrievals using the land parameter retrieval model: Evaluation over the Murrumbidgee Catchment, southeast Australia, Remote Sens. Environ., 163, 70–79, https://doi.org/10.1016/j.rse.2015.03.006, 2015.

van der Schalie, R., Kerr, Y., Wigneron, J., Rodríguez-Fernández, N., Al-Yaari, A., and Jeu, R.: Global SMOS Soil Moisture Retrievals from The Land Parameter Retrieval Model, Int. J. Appl. Earth Obs. Geoinf., 45, 125–134, https://doi.org/10.1016/j.jag.2015.08.005, 2016.

van der Schalie, R., van der Vliet, M., Rodríguez-Fernández, N., Dorigo, W. A., Scanlon, T., Preimesberger, W., Madelon, R., and de Jeu, R. A. M.: L-Band Soil Moisture Retrievals Using Microwave Based Temperature and Filtering. Towards Model-Independent Climate Data Records, Remote Sens., 13, 2480, https://doi.org/10.3390/rs13132480, 2021.

van der Schalie, R., van der Vliet, M., Albergel, C., Dorigo, W., Wolski, P., and de Jeu, R.: Characterizing natural variability in complex hydrological systems using passive microwave-based climate data records: a case study for the Okavango Delta, Hydrol. Earth Syst. Sci., 26, 3611–3627, https://doi.org/10.5194/hess-26-3611-2022, 2022.

van der Vliet, M., van der Schalie, R., Rodriguez-Fernandez, N., Colliander, A., de Jeu, R., Preimesberger, W., Scanlon, T., and Dorigo, W.: Reconciling Flagging Strategies for Multi-Sensor Satellite Soil Moisture Climate Data Records, Remote Sens., 12, 3439, https://doi.org/10.3390/rs12203439, 2020.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Meth., 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Vreugdenhil, M., Dorigo, W., Broer, M., Haas, P., Eder, A., Hogan, P., Bloeschl, G., and Wagner, W.: Towards a high-density soil moisture network for the validation of SMAP in Petzenkirchen, Austria, in: 2013 IEEE International Geoscience and Remote Sensing Symposium – IGARSS, 1865–1868, https://doi.org/10.1109/IGARSS.2013.6723166, 2013.

Vreugdenhil, M., Greimeister-Pfeil, I., Preimesberger, W., Camici, S., Dorigo, W., Enenkel, M., van der Schalie, R., Steele-Dunne, S., and Wagner, W.: Microwave remote sensing for agricultural drought monitoring: Recent developments and challenges, Front. Water, 4, 1045451, https://doi.org/10.3389/frwa.2022.1045451, 2022.

Wagner, W., Lemoine, G., and Rott, H.: A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data, Remote Sens. Environ., 70, 191–207, https://doi.org/10.1016/S0034-4257(99)00036-X, 1999.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldana, J., De Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., and Steinnocher, K.: The ASCAT

soil moisture product: A review of its specifications, validation results, and emerging applications, Meteorol. Z., 22, 5–33, https://doi.org/10.1127/0941-2948/2013/0399, 2013.

Wagner, W., Lindorfer, R., Hahn, S., Kim, H., Vreugdenhil, M., Gruber, A., Fischer, M., and Trnka, M.: Global Scale Mapping of Subsurface Scattering Signals Impacting ASCAT Soil Moisture Retrievals, IEEE T. Geosci. Remote, 62, 1–20, https://doi.org/10.1109/TGRS.2024.3429550, 2024.

Wang, G., Garcia, D., Liu, Y., de Jeu, R., and Johannes Dolman, A.: A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations, Environ. Model. Softw., 30, 139–142, https://doi.org/10.1016/j.envsoft.2011.10.015, 2012.

Wang, T., Yu, P., Wu, Z., Lu, W., Liu, X., Li, Q. P., and Huang, B.: Revisiting the Intraseasonal Variability of Chlorophyll-a in the Adjacent Luzon Strait With a New Gap-Filled Remote Sensing Data Set, IEEE T. Geosci. Remote, 60, 1–11, https://doi.org/10.1109/TGRS.2021.3067646, 2022.

Wang, X., Lü, H., Crow, W. T., Corzo, G., Zhu, Y., Su, J., Zheng, J., and Gou, Q.: A reduced latency regional gap-filling method for SMAP using random forest regression, iScience, 26, 105853, https://doi.org/10.1016/j.isci.2022.105853, 2023.

Wigneron, J.-P., Dayan, S., Kruszewski, A., Aluome, C., Al-Yaari, A., Fan, L., Guven, S., Chipeaux, C., Moisy, C., Guyon, D., and Loustau, D.: The Aqui Network: Soil Moisture Sites in the "Les Landes" Forest and Graves Vineyards (Bordeaux Aquitaine Region, France), in: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, 3739–3742, https://doi.org/10.1109/IGARSS.2018.8517392, 2018.

Xaver, A., Zappa, L., Rab, G., Pfeil, I., Vreugdenhil, M., Hemment, D., and Dorigo, W. A.: Evaluating the suitability of the consumer low-cost Parrot Flower Power soil moisture sensor for scientific environmental applications, Geosci. Instrum. Method. Data Syst., 9, 117–139, https://doi.org/10.5194/gi-9-117-2020, 2020.

Yang, K., Qin, J., Zhao, L., Chen, Y., Tang, W., Han, M., Zhu, l., Chen, Z., Lv, N., Ding, B., Wu, H., and Lin, C.: A Multi-Scale Soil Moisture and Freeze-Thaw Monitoring Network on the Third Pole, B. Am. Meteorol. Soc., 94, 1907–1916, https://doi.org/10.1175/BAMS-D-12-00203.1, 2013.

Yang, X., Zhang, C., Zhaoyun, C., Fan, Y., Wang, J., and Han, Y.: Filling method for soil moisture based on BP neural network, J. Appl. Remote Sens., 12, 042806, https://doi.org/10.1117/1.JRS.12.042806, 2018.

Young, R., Walker, J., Yeoh, N., Smith, A., Ellett, K., Merlin, O., and Western, A.: Soil moisture and meteorological observations from the Murrumbidgee catchment, Department of Civil and Environmental Engineering, The University of Melbourne, https://www.researchgate.net/profile/Andrew-Western/publication/267832777_Soil_Moisture_and_Meteorological_Observations_From_the_Murrumbidgee_Catchment/links/557a496c08aeacff2003d2a9/Soil-Moisture-and-Meteorological-Observations-From-the-Murrumbidgee-Catchment.pdf (last access: 20 August 2025), 2008.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., and Vereecken, H.: A Network of Terrestrial Environmental Observatories in Germany, Vadose Zone J., 10, 955–973, https://doi.org/10.2136/vzj2010.0139, 2011.

Zappa, L., Forkel, M., Xaver, A., and Dorigo, W.: Deriving Field Scale Soil Moisture from Satellite Observations and Ground Measurements in a Hilly Agricultural Region, Remote Sens., 11, 2596, https://doi.org/10.3390/rs11222596, 2019.

Zappa, L., Woods, M., Hemment, D., Xaver, A., and Dorigo, W.: Evaluation of Remotely Sensed Soil Moisture Products using Crowdsourced Measurements, Eighth International Conference on Remote Sensing and Geoinformation of Environment, SPIE, Cyprus, https://doi.org/10.1117/12.2571913, 2020.

Zaussinger, F., Dorigo, W., Gruber, A., Tarpanelli, A., Filippucci, P., and Brocca, L.: Estimating irrigation water use over the contiguous United States by combining satellite and reanalysis soil moisture data, Hydrol. Earth Syst. Sci., 23, 897–923, https://doi.org/10.5194/hess-23-897-2019, 2019.

Zhang, L., Liu, Y., Ren, L., Teuling, A. J., Zhang, X., Jiang, S., Yang, X., Wei, L., Zhong, F., and Zheng, L.: Reconstruction of ESA CCI satellite-derived soil moisture using an artificial neural network technology, Sci. Total Environ., 782, 146602, https://doi.org/10.1016/j.scitotenv.2021.146602, 2021.

Zhao, T., Shi, J., Lv, L., Xu, H., Chen, D., Cui, Q., Jackson, T. J., Yan, G., Jia, L., Chen, L., L., Zhao, K., Zheng, X., Zhao, L., Zheng, C., Ji, D., Xiong, C., Wang, T., Li, R., Pan, J., Wen, J., Yu, C., Zheng, Y., Jiang, L., Chai, L., Lu, H., Yao, P., Ma, J., Lv, H., Wu, J., Zhao, W., Yang, N., Guo, P., Li, Y., Hu, L., Geng, D., and Zhang, Z.: Soil moisture experiment in the Luan River supporting new satellite mission opportunities, Remote Sens. Environ., 240, 111680, https://doi.org/10.1016/j.rse.2020.111680, 2020.

Zheng, J., Zhao, T., Lü, H., Shi, J., Cosh, M. H., Ji, D., Jiang, L., Cui, Q., Lu, H., Yang, K., Wigneron, J.-P., Li, X., Zhu, Y., Hu, L., Peng, Z., Zeng, Y., Wang, X., and Kang, C. S.: Assessment of 24 soil moisture datasets using a new in situ network in the Shandian River Basin of China, Remote Sens. Environ., 271, 112891, https://doi.org/10.1016/j.rse.2022.112891, 2022.

Zotta, R.-M., Moesinger, L., van der Schalie, R., Vreugdenhil, M., Preimesberger, W., Frederikse, T., de Jeu, R., and Dorigo, W.: VODCA v2: multi-sensor, multi-frequency vegetation optical depth data for long-term canopy dynamics and biomass monitoring, Earth Syst. Sci. Data, 16, 4573–4617, https://doi.org/10.5194/essd-16-4573-2024, 2024.

Zreda, M., Desilets, D., Ferré, T., and Scott, R.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, Geophys. Res. Lett., 35, L21402, https://doi.org/10.1029/2008GL035655, 2008.

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.: COSMOS: the COsmic-ray Soil Moisture Observing System, Hydrol. Earth Syst. Sci., 16, 4079–4099, https://doi.org/10.5194/hess-16-4079-2012, 2012.