ESA Climate Change Initiative Plus - Soil Moisture

Algorithm Theoretical Baseline Document (ATBD) Supporting Product Version 06.1

D2.1 Version 2

19-04-2021

Prepared by

Earth Observation Data Centre for Water Resources Monitoring (EODC) GmbH

in cooperation with

TU Wien, VanderSat, CESBIO and ETH Zürich

This document forms the deliverable D2.1 Algorithm Theoretical Basis Document (ATBD) and was compiled for the European Space Agency (ESA) Climate Change Initiative Plus Soil Moisture Project (ESRIN Contract No: 4000126684/19/I-NB" ESA CCI+ Phase 1 New R&D on CCI ECVS Soil Moisture").

For more information on the CCI programme of the ESA see https://climate.esa.int/en/.

Number of pages: 61 (including cover and preface)

Authors:	T. Scanlon, A. Pasik, W. Dorigo, R.A.M de Jeu, S. Hahn, R. van der Schalie, W. Wagner, R. Kidd, A. Gruber, L. Moesinger, W. Preimesberger		
Circulation:	Internal Review		
Issue	Date	Details	Editor
V1	29/05/2020	Update of document to support version 5.2 of the ESA CCI SM product. Included restructuring to a single document. Changes to the product included the addition of SMAP as well as updating the input data to LPRMv6 (PASSIVE and COMBINED) and the rescaling of AMSR2 to AMSRE (PASSIVE only). All products now use an updated CDF matching scheme. Temporal extent of product is to the end of 2019.	AP, RK, WP, WD, LM SH
V2 Issue 0.1	04/02/2021	Update to reflect version 6 of the ESA CCI SM product. Updates include: increased time period used for TMI, cross flagging implemented for frozen soils and improved SNR-VOD regression by splitting into land cover classes. All input passive sensor data has been updated to LPRMv6.1 and a new experimental ASCAT dataset is used. Temporal extent of product is to the end of 2020. Document updated to ensure all sections contain accurate and relevant information. Sections on known limitations updated. Scientific advancements taken as already provided in the Algorithm Development Plan (ADP). Section on reference documents folded into the references section at the end of the document. Table 1 has been updated to only include versions released to the public.	TS
V2 Issue 0.2	15/02/2021	Update post review by VanderSat	RvdS, TS
V2 Issue 0.3	02/03/2021	Review, Update ToC, Modification History. To ESA for Review	RK
V2 Issue 0.4	16/03/2021	Minor revision of text. Accepted all changes (from v0.3) retained reviewer's comments. Updated all section headers. To TU Wien for finalisation	RK
V2 Issue 0.5	16/03/2021	Changes in line with reviewer comments. Return document to EODC.	TS

V2 Issue 0.6	02/04/2021	Reviewed, Accepted all format changes, ToC, List of Figures, updated doc meta data. To ESA for approval	R. Kidd
V2 Issue 1.0	16/04/2021	Finalised	R. Kidd
V2 issue 1.1	19/04/2021	Revised v06.1 release date	R. Kidd

For any clarifications please contact Wouter Dorigo (Wouter.dorigo@geo.tuwien.ac.at).

Project Partners

Prime Contractor and project management	**EODC,** Earth Observation Data Centre for Water Resources Monitoring (Austria)
Earth Observation Partners	**TU Wien**, Vienna University of Technology (Austria) **VanderSat**, The Netherlands **CESBIO**, France
Climate Research Partners	**ETH**, Institute for Atmospheric and Climate Science, (Switzerland)

Table of contents

List of Figures

List of Tables

Definitions, acronyms and abbreviations

AMI	Active Microwave Instrument
AMSR2	Advanced Microwave Scanning Radiometer 2
AMSR-E	Advanced Microwave Scanning Radiometer-Earth Observing System
AMSU	Advanced Microwave Sounding Unit
ASAR	Advanced Synthetic Aperture Radar
ASCAT	Advanced Scatterometer (Metop)
CCI	Climate Change Initiative
CEOP	Coordinated Energy and Water Cycle Observations Project
CMORPH	Morphing Method of the Climate Prediction Centre
CPC	Climate Prediction Centre
DARD	Data Access Requirement Document
DMSP	Defense Meteorological Satellite Program
DTED	Digital Terrain Elevation Model
EASE	Equal-Area Scalable Earth
ECV	Essential Climate Variable
ENVISAT	Environmental Satellite
EO	Earth Observation
ERA-40	ECMWF ReAnalysis 40 data set
ERS	European Remote Sensing Satellite (ESA)
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FTP	File Transfer Protocol
GIMMS	Global Inventory Modeling and Mapping Studies
GLDAS	Global Land Data Assimilation System
GLWD	Global Lakes and Wetlands Database (GSPC/University of Kassel)
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project
GRACE	Gravity Recovery And Climate Experiment
GSWP	Global Soil Wetness Project
ISMN	International Soil Moisture Network
ITRDB	International Tree-Ring Data Bank
JAXA	Dokuritsu-gyosei-hojin Uchu Koku Kenkyu Kaihatsu Kiko, (Japan Aerospace Exploration Agency)
JPL	Jet Propulsion Laboratory (NASA)
METOP	Meteorological Operational Satellite (EUMETSAT)

NASA	National Aeronautics and Space Administration
NIMA	National Imagery and Mapping Agency
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center (radlab)
NWS	National Weather Service (NOAA)
QCM	Quantile Cumulative Matching
SAR	Synthetic Aperture Radar
SCAT	Scatterometer
SMAP	Soil Moisture Active and Passive mission
SMMR	Scanning Multichannel Microwave Radiometer
SMOS	Soil Moisture and Ocean Salinity (ESA)
SNR	Signal to Noise Ratio
SOW	Statement of Work
SSM	Surface Soil Moisture
SSM/I	Special Sensor Microwave Imager
TCA	Triple Collocation Analysis
TDR	Time Domain Reflectometry
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measuring Mission
TWS	Terrestrial Water Storage
USGS	United States Geological Survey
VIC	Variable Infiltration Capacity
VOD	Vegetation Optical Depth
WACMOS	Water Cycle Multimission Observation Strategy
WindSat	WindSat Radiometer

List of symbols

θ	Incidence angle (degree), generic
$\theta_{i,b}$	Observed incidence angle of beam $b \in \{f, m, a\}$ (fore-, mid-, aft-beam) of i-th record in the time series of the current GPI
$\varphi, \varphi_{i,b}$	Azimuth angle (degree), generic and observed
$\sigma^0, \sigma^0_{i,b}$	Radar cross-section, backscattering coefficient $\left(\frac{m^2}{m^2} \text{ or } dB\right)$, generic and observed
t, t_i	Time, generic and observed
$d = doy(t), d_i$	Day of year, $d \in \mathbb{N}, 1 \leq d \leq 366$, as function of t (t_i)
$\sigma^0(\theta, d)$	Backscatter, modelled as function of incidence angle, with the model depending on the day of year d (i.e., d indexes one instance of the model class)
$\sigma^0_{i,b}(\theta_i)$	Observed backscatter, represented in terms of the model
$\sigma'(\theta, d)$	First derivative of $\sigma^0(\theta, d)$
$\sigma'(\theta_{ref}, d)$	First derivative ('slope') at reference angle, parameter array
$\sigma''(\theta, d)$	Second derivative of $\sigma^0(\theta, d)$
$\sigma''(\theta_{ref}, d)$	Second derivative ('curvature') at reference angle, parameter array
$\overline{\sigma^0_i}(\theta_{ref})$	Normalised backscatter at reference angle, averaged over the beams, of the i-th record in the time series
θ_{dry}	Dry crossover angle
$\sigma^{dry}(\theta_{ref}, d)$	Dry reference at reference angle, parameter array
θ_{wet}	Wet crossover angle
$\sigma^{wet}(\theta_{ref}, d)$	Wet reference at reference angle, parameter array

1 Introduction

The Algorithm Theoretical Baseline Document (ATBD) provides a detailed description of the algorithms that are used within the ESA CCI Soil Moisture production system. The ESA CCI SM production system was initially developed within CCI Phases 1 & 2 and is continuously being updated within CCI+ to reflect the current state of the science driving the system. The aim of this document is to describe the algorithm development process for each of the ESA CCI SM products, as well as provide an executive summary setting them within framework for the CCI project and the ESA CCI SM production system.

The structure of this document reflects the distinct domains of the ATBD. Sections 3 and 4 provide a brief overview of the problem and of the ESA CCI SM production system respectively. Section 5contains a brief description of soil moisture products from active microwave sensors used in in the ESA CCI Soil Moisture and points to the organisations responsible for their retrieval. Section 6 describes succinctly the VUA-NASA Land Parameter Retrieval Method (LPRM) for estimating soil moisture from passive microwave sensors, and section 7 provides a description of the methodology adopted for merging the active and passive soil moisture products.

1.1 Purpose of the Document

The ATBD is intended to provide a detailed description of the scientific background and theoretical justification for the algorithms used to produce the ESA CCI soil moisture data sets. Furthermore, it describes the scientific advances and algorithmic improvements which are made within the CCI project. This document is complemented by (Dorigo et al., 2017) and Gruber et al. (2019) which provides detailed information on the product including a quality assessment which shows the evolution of the product between versions.

1.2 Targeted Audience

This document targets mainly:

1. Remote sensing experts interested in the retrieval and error characterisation of soil moisture from active and passive microwave data sets.
2. Users of the remotely sensed soil moisture data sets who want to obtain a more in-depth understanding of the algorithms and sources of error.

2 Change log

2.1 Current version 06.1

This document forms deliverable 2.1 of CCI+ and provides an update for the ESA CCI SM 06.1 product expected to be publicly released in the Q1/2021. Changes between version 06.1 algorithm and the previously used v05.2 algorithm (temporally extended at v05.3 without algorithmic changes) include:

- Inclusion of data from the new LPRMv6.1 datasets (for all passive sensors) as well as the addition of data from Global Precipitation Measurement (GPM) and FengYun-3B and -3D
    - Improved inter-calibration using FengYun-3B
    - Use of 37 GHz observations SMOS and SMAP temperature input
    - Improved flagging of snow / frozen soils in the passive (LPRM) datasets as well as cross-flagging within the CCI processor to exploit flags from all datasets
- Updated ACTIVE input from ASCAT-A and -B which addresses vegetation issues with a new cross-over angle optimisation and improved snow / frozen conditions flagging
- Extension of the time period used for TMI and updating the triplet used for TMI triple collocation analysis (for determining weights)
- SNR-VOD regression undertaken per land-cover class rather than globally
- Structural breaks from sensor changes detected and corrected and provided as an ancillary product
- Investigation of scaling references that are independent from land surface models (for implementation at a later product version)

Version 06.1 provides data from 1978 (PASSIVE and COMBINED products) and 1991 (ACTIVE product) to the end of December 2020.

2.1.1 ATBD Document

- Updated for version 06.1. Tables and figures revised where applicable.

2.2 Pre v06.1

The dataset and corresponding ATBD versions are summarised in Table 1. Further information can be found in the changelog provided with the data and the relevant documentation.

Table 1: Summary ESA CCI SM Versions Released to the Public

Dataset Version	Release date	Major Changes Since Previous Versions	ATBD Version
V06.1	2021-04-19	Inclusion of updated LPRMv6.1 data which includes improvements to inter-calibration and flagging as well as data being available for GPM and FY-3B. Inclusion of updated ASCAT datasets which includes improvements to the vegetation correction and snow / frozen conditions flagging. Updates to algorithm include implementation of cross-flagging, use of gap-filling the error characterisation per land cover class and extension of the TMI dataset to 2015.	6.1
v05.3	2021-02-08	Extension of the dataset until 2020-12-31. LPRMv6 data for SMAP generated from updated brightness temperatures (SPL3SMPv7).	5.3
v05.2	2020-09-08	Inclusion of SMAP data from April 2015, improved CDF-matching and updated inter-sensor scaling regime of AMSR2.	5.2
v04.7	2020-03-12	No algorithm changes since v04.4. Temporal extension to 2019-12-31.	4.7
v04.5	2019-09-30	No algorithm changes since v04.4. Temporal extension 2018-12-31. ATBD documentation previously maintained separately for each of the ESA CCI SM datasets merged into a single document. Removal of the Active ATBD.	4.5
v04.4	2018-11-12	No algorithm changes since v04.1. GLDAS 2.1 now used. Flagging of high VOD for SMOS and AMSR2 method changed. Temporal extension to 2018-06-30.	4.4
v04.2	2018-01-12	The combined product is now generated by merging all active and passive L2 products directly, rather than merging the generated active and passive products. Spatial gaps in TC-based SNR estimates now filled using a polynomial SNR-VOD regression. sm_uncertainties now available globally for all sensors except SMMR. The p-value based mask to exclude unreliable input data sets in the COMBINED product has been modified and is also applied to the passive product. Masking of unreliable retrievals is undertaken prior to merging.	4.2
v03.3	2017-11-13	Temporal extension of ACTIVE, PASSIVE and COMBINED datasets to 2016-12-31.	3.3
v03.2	2017-02-14	Introduction of new weighted-average based merging scheme. Miras SMOS (LPRM) now integrated into the data products. Blending weights provided as ancillary data files.	3.2

		Blending made more conservative concerning the inclusion of single low-accuracy observations (on the cost of temporal coverage). Integration of Metop-B ASCAT. Error estimates which are used for relative weight estimation now provided alongside with the merged soil moisture observations. SMOS temporal coverage extended. Uncertainty estimates for soil moisture now provided from 1991-08-05 onwards (ACTIVE), and from 1987-07-10 onwards (PASSIVE, and COMBINED). Two new quality flags introduced.	
v02.3	2016-02-08	Temporal extension to 2015-12-31. Valid_range in netCDF files now set to the packed data range.	-
v02.2	2015-12-17	Temporal coverage extended (Nov-1978 to Dec-2014). Improvement in the flagging of the active data where extreme high and low values are filtered. Email address added to metadata. In ancillary files latitudes now goes from positive to negative values. Change of product name to ESA CCI SM. Soil moisture values (flagged with values other than 0) are now set to NaN.	-
v02.0	2014-07-10	Combined product only including passive sensors (SMMR, SSM/I, TMI, AMSR-E; active: AMI-WS, ASCAT) with time span: 1978-11-01 to 2010-12-31. NetCDF-3 classic CF1.5 compliant. Active, passive and combined products made available. Dataset time span: 1978-11-01 to 2013-12-13 (passive and combined) and 1991-08-05 to 2013-12-13 (active). Using new land mask based on GSHHG 2.2.2. WindSat and preliminary AMSR2 included. ERS2 included in AMI-WS dataset. Active data resampled with Hamming window function. Improved rescaling algorithm. Data gaps in 2003-02-16 to 2006-12-31 filled with AMSR-E data. Provision of ancillary datasets (land mask, porosity map, soil texture data, AMSR-E VUA-NASA Vegetation Optical Depth averaged over the period 2002-2011, global topographic complexity and Global Wetland fraction. All datasets updated to include days where no observations are available.	-

3 Scope of ESA CCI Soil Moisture

3.1 Soil Moisture Becoming an ECV

Soil moisture is arguably one of the most important parameters for the understanding of physical, chemical and biological land surface processes (Legates et al., 2011). Therefore, it is essential for many geoscientific applications to know how much water is stored in the soil, and how it varies in space and time.

For many years, soil moisture was considered to be only an "emergent ECV" because the retrieval of soil moisture was deemed too difficult with existing satellite sensors. Therefore, in recognition of the strong need for global soil moisture data sets, the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA) each decided to develop a dedicated satellite mission operating at 1.4 GHz (L-band). The first mission is the Soil Moisture and Ocean Salinity (SMOS) satellite that was launched in November 2009 by ESA (Kerr et al., 2010). The second one is NASA's Soil Moisture Active Passive (SMAP) mission that was launched in January 2015 (Entekhabi et al., 2010a). But, as already noted by (Wagner et al., 2007): *"Besides these innovations in space technology, an initially less-visible revolution has taken place in algorithmic research. This revolution became possible thanks to the increasing availability of computer power, disk space, and powerful programming languages at affordable costs. This has allowed more students and researchers to develop and test scientific algorithms on regional to global scales than in the past. This has led to a greater diversity of methods and consequently more successful retrieval algorithms."*

In line with the above-described developments, several global and continental-scale soil-moisture datasets have been published and shared openly with the international community within the last 20 years. The very first remotely sensed global soil moisture dataset was published by the Vienna University of Technology (TU Wien) in 2002 and was based on nine years (1992-2000) of ERS C-band (5.6 GHz) scatterometer measurements (Scipal et al., 2002; Wagner et al., 2003). NASA released its first global soil moisture data retrieved from microwave radiometer measurements using the algorithms developed by Njoku et al. (2003) in the following year. Since then several other soil moisture data products mostly based on microwave radiometers (AMSR E, Windsat, etc.) have become freely available, notably the multi-sensor soil moisture datasets produced by Vrije Universiteit Amsterdam (VUA) in cooperation with NASA (Owe et al., 2008), and the WindSat soil moisture dataset produced by the US Naval Research Laboratory (Li et al., 2010).

The first operational near-real-time soil moisture service was launched by EUMETSAT in 2008 based on the METOP Advanced Scatterometer (ASCAT) and algorithms and software prototypes developed by TU Wien (Bartalis et al., 2007). Finally, SMOS Level 2 soil moisture data started to become available in 2010, with first validation results published in Albergel et al. (2012). Data from NASA's Soil Moisture Active Passive (SMAP) became available in the course of 2015, but unfortunately, only after 3 months of operation its radar failed thus impeding the continuation of the foreseen downscaled product.

Having a number of independent satellite soil moisture data sets does not mean that it is straight-forward to create long-term consistent time series suitable for climate change studies. In fact, for the assessment of climate change effects on soil moisture even subtle long-term trends must be detected reliably. This means that any potential influences of mission specifications, sensor degradation, drifts in calibration, and algorithmic changes must be carefully corrected for. Also, it must be guaranteed that the soil moisture data retrieved from the different active and passive microwave instruments are physically consistent.

3.2 Selected Satellite Sensors

Microwave remote sensing measurements of bare soil surfaces are very sensitive to the water content in the surface layer due to the pronounced increase in the soil dielectric constant with increasing water content (Ulaby et al. 1982). This is the fundamental reason why microwave techniques offer the opportunity to measure soil moisture in a relatively direct manner. For soil moisture studies the most important bands are: L-band (frequency f = 1 – 2 GHz, wavelength λ = 30 – 15 cm), C-band (f = 4 – 8 GHz, λ = 7.5 – 3.8 cm), and X-band (f = 8 – 12 GHz, λ = 3.8 – 2.5 cm).

In microwave remote sensing, one distinguishes active and passive techniques. Active microwave sensors (scatterometers) transmit an electromagnetic pulse and measure the energy scattered back from the Earth's surface. For passive sensors (radiometers), the energy source is the target itself, and the sensor is merely a passive receiver (Ulaby et al. 1982). Radiometers measure the intensity of the emission of the Earth's surface that is related to the physical temperature of the emitting layer and the emissivity of the surface.

Despite the different measurement processes, active and passive methods are closely linked through Kirchhoff's law which, applied to the problem of remote sensing of the Earth's surface, states that the emissivity is one minus the hemisphere integrated reflectivity (Schanda 1986). Therefore, both active and passive techniques deal, in principle, with the same physical phenomena, though the importance of different parameters on the measured signal may vary depending on the sensor characteristics.

Given that an ECV data record should be as long and complete as possible, it has to be based on both active and passive microwave observations. The ESA CCI SM product uses both C-band scatterometers (e.g. ERS-1/2 scatterometer, METOP Advanced Scatterometers (ASCAT)) and multi-frequency radiometers (e.g., SMMR, SSM/I, TMI, AMSR-E, Windsat, AMSR2, SMOS, SMAP, GPM and FengYun-3B). The coverage of these sensors is shown in Figure 1.

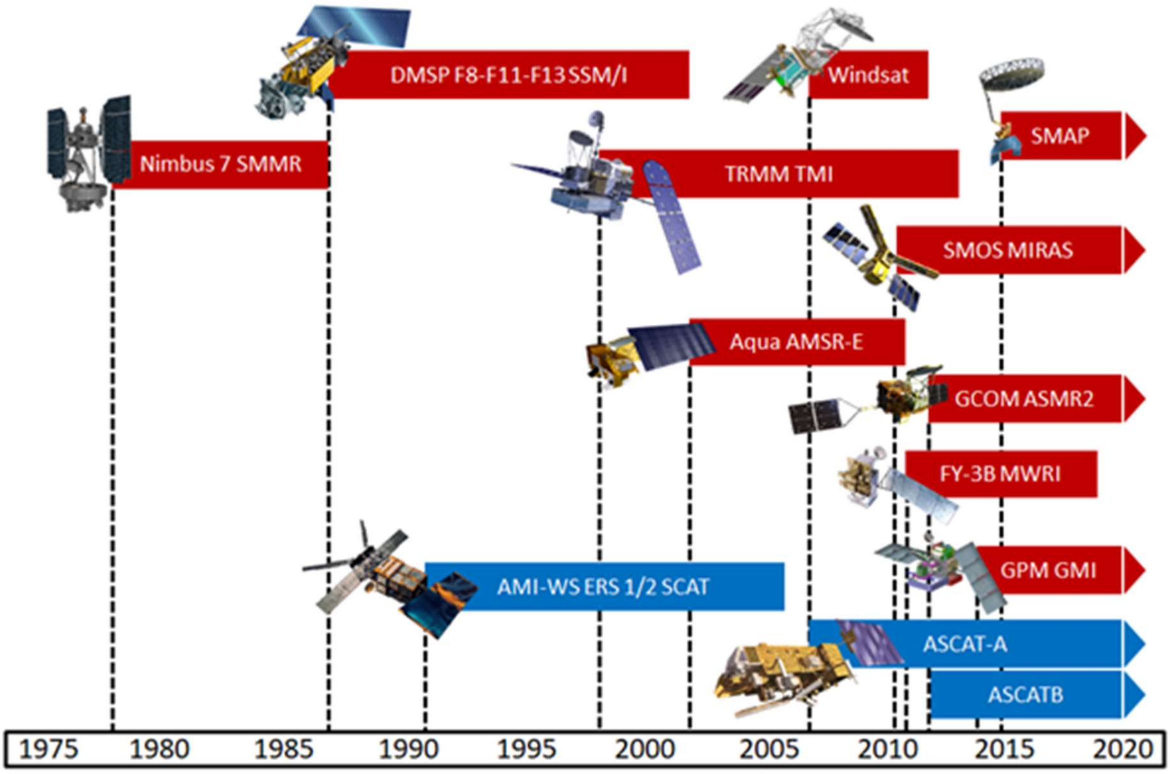


Figure 1: Active and passive microwave sensors used for the generation of the ESA CCI soil moisture data sets.

3.3 Baseline Requirements

As part of the CCI Soil Moisture project, a detailed assessment of the user requirements is carried out at regular intervals and reported in the User Requirement Document (URD) (Scanlon et al, 2020). Based on the URD and the requirements as specified in the SoW (ESA, 2018), and drawing from the experiences of the use of the currently available satellite soil moisture data sets, a number of baseline requirements are specified in the following sections.

3.3.1 Scientific Requirements

Due to the fact that several decade-long soil moisture data records have been released within the last decade the generic user requirements for ESA CCI soil moisture data records are already reasonably well understood. According to authors' experience from the cooperation with users of the TU Wien and VUA-NASA soil moisture data sets (de Jeu et al., 2008; Wagner et al., 2007), the most important of these are as listed below.

Note that these are suggested requirements and the ESA CCI SM product does not necessarily fulfil each requirement in full. Please see the URD (Scanlon et al., 2019) for further details.

1. Soil moisture is preferably expressed in volumetric soil moisture units (m^3m^{-3}). If soil moisture is expressed in a different unit, the conversion rule must be specified.
2. From an application point of view, the ESA CCI SM data should preferably represent the soil moisture content in deeper soil layers (up to 1 m), not just the thin (0.5-5 cm) remotely sensed surface soil layer. Nevertheless, expert users typically prefer to work with data that are as close to the sensor measurements as possible, making the conversion of the remotely sensed surface soil moisture measurements to profile estimates themselves.
3. When merging datasets coming from different sensors and satellites the highest possible degree of physical consistency shall be pursued.
4. Due to the long autocorrelation length of the atmosphere-driven soil moisture field (Entin et al., 2000) a spatial resolution of ≤50 km is sufficient for climate studies.
5. The temporal sampling interval depends on the chosen soil layer. For deeper soil layers (1 m) a sampling rate of 1 week is in general enough, but for the thin remotely sensed soil layer it is ≤1 day.
6. Having a good quantitative understanding of the spatio-temporal error field is more important than working under the assumption of arbitrarily selected accuracy thresholds (e.g. like the often cited 0.04 m^3m^{-3}).
7. Some soil moisture applications require a good accuracy (low bias), but for most applications it is in fact more important to achieve a good precision (Entekhabi et al., 2010b; Koster et al., 2009).
8. For climate change studies the drift in the bias and dynamic range of the soil moisture retrievals should be as small as possible.

3.3.2 System Requirements

The generation of an ESA CCI SM data set is not a one-off activity, but is in fact a long-term process where the ESA CCI SM product is continually improved step by step with the active involvement of a broad scientific community. A robust modular processing system has been developed so that:

- the system supports algorithm development and is most open to broad scientific participatory inputs
- algorithms can be improved while minimising reprocessing costs
- upgrades of any of its parts are facilitated without repercussions elsewhere
- the system can be moved to different operators if required

The design and operations of the system is also as lightweight as possible in order to:

- re-process ESA CCI SM data records on a frequent basis to account for Level 1 calibration- and Level 2 algorithmic updates

- update the ESA CCI SM datasets rapidly in case new Level 2 data sets become available
- test alternative error characterisation, matching and merging approaches
- keep operations and maintenance costs low

Please consult Kidd et al. (2013) for further details on the soil moisture ESA CCI SM production system, detailing its components, their functions, and interfaces.

4 ESA CCI SM Production Approach

4.1 Potential and drawbacks of merging Level 1 Microwave Observations

Probably the most straight-forward approach to generating an ESA CCI soil moisture data set would be to feed the Level 1 backscatter- and brightness temperature observations of all different active and passive microwave remote sensing instruments into one Level 2 soil moisture retrieval system, delivering as direct output a harmonised and consistent active-passive based ESA CCI surface soil moisture data set covering the complete period from 1978 to the present. As ideal as this approach may seem from a scientific point of view, there are some major practical problems:

- The technical specifications of the diverse active and passive microwave sensors suitable to soil moisture retrieval (ASCAT, AMSR-E, SMOS, SMAP, etc.) are so different that it appears hardly feasible to design a one-fits-all physical retrieval algorithm.
- The complexity of the retrieval algorithm and the requirements for high-quality ancillary data to constrain the retrieval process can be expected to increase drastically for a multi-sensor compared to a single-sensor Level 2 retrieval approach. This bears a certain risk of errors becoming less easily traceable. Also, the overall software system may not be scalable in terms of processing time and disk space.
- For much of the historic time period (1978-2007) the spatio-temporal overlap of suitable active and passive microwave measurements is minimal.
- Because the surface soil moisture content may vary within minutes to hours, combing measurements taken at different times of the day in multi-sensor approach may produce large errors. For example, that the measurements of ASCAT (9:30 and 21:30 local time), AMSR-E (1:30 and 13:30) and SMOS (6:00 and 18:00) are currently well spread over the complete day.

Each of these problems is serious enough to not consider an ESA CCI SM Production System based on the fusion of Level 1 microwave observations. Considered together one can conclude that such an ESA CCI SM Production system would neither be modular nor lightweight, which makes this approach technically intractable. Therefore, in the next section the fusion of Level 2 soil moisture retrievals is discussed.

4.2 Fusion of Level 2 Soil Moisture Retrievals

Prior to the establishment of the ESA CCI SM product, the possibility of generating a long-term soil moisture data set based on Level 2 soil moisture retrievals was already demonstrated within the WACMOS project funded by the European Space Agency (Su et al., 2010). The Level 2 fusion process of this early product involved first fusing all active, then passive datasets then merging those fused products to create a combined product.

In this approach the three important steps in the fusion process were:

1. error characterisation (Su et al., 2010)
2. matching to account for data set specific biases (Drusch et al., 2005; Reichle et al., 2004)
3. merging the bias-corrected datasets (Liu et al., 2011).

The major advantage of this approach is that it allows combining surface soil moisture data derived from different microwave remote sensing instruments with substantially different instrument characteristics. It is only required that the retrieved Level 2 surface soil moisture data pass pre-defined quality criteria. In this way it is guaranteed that no sensor is a priori excluded by this approach. It is thus straight-forward to further enhance the ESA CCI SM data set with Level 2 data from other existing and any new sensors.

In this approach, the ESA CCI SM Production System does not include the different Level 2 processors. In other words, the different Level 2 baseline data can be provided by the expert teams and organisations for the different sensor types (scatterometers, multi-frequency radiometers, SMOS, SMAP, etc.) and the ESA CCI SM Production System itself only has to deal with the fusion process, as described above. This design is modular and lightweight, meeting the requirements as discussed in Section 3.3.2.

The most serious concern related to this fusion approach is that Level 1 data processed with different Level 2 algorithms may not represent the same physical quantity. Fortunately, as an increasing number of validation and inter-comparison studies show (Albergel et al., 2012; Brocca et al., 2011; Gruhier et al., 2010; Rüdiger et al., 2009), the temporal soil moisture retrieval skills of many of the input datasets (including SMOS, ASCAT and AMSR-E) are often well comparable and of good quality in regions with sparse to moderate vegetation cover.

Therefore, after bias correction and, if necessary, a conversion of units, the different Level 2 soil moisture data sets can be merged. Nevertheless, to maximise physical consistency it is advisable to process all active microwave data sets with one algorithm, and all passive microwave data with another algorithm. For the ESA CCI SM product, the TU Wien change detection method is used for all active datasets and the LPRM algorithm is applied to all passive datasets.

However, as different algorithms are used for the active and passive soil moisture retrievals, the resulting data is not directly comparable. To account for these differences, the ESA CCI SM product delivers three products: ACTIVE – based only on scatterometer data, PASSIVE – based only on radiometer data and COMBINED which uses both. It is up to the user to decide which of these merged soil moisture data sets is best suited for their application.

The basic fusion concept developed within WACMOS and CCI still holds today, even though noticeable modifications were made over the years. The current status of the merging methodology is described in Section 7.

5 Soil Moisture Retrieval from Active Sensors

Active microwave soil moisture products (see Figure 1 for details) utilised in the generation of the ESA CCI SM ACTIVE and COMBINED datasets are obtained from external operational sources as follows:

- ERS-1 AMI surface soil moisture products have been generated at TU Wien (TU WIEN, 2013).
- ERS-2 AMI surface soil moisture data sets stem from reprocessing activities which have been carried out within ESA's SCIRoCCo project (Crapolicchio et al., 2016). The ERS-2 data set used in all ESA CCI SM versions is the ERS.SSM.H.TS 25 km soil moisture time series product (ESA, 2017).
- Metop ASCAT surface soil moisture data sets stem from the EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF, http://h-saf.eumetsat.int/). ESA CCI SM 06.1 uses both the H-SAF H119 Metop ASCAT SSM CDR v5 (H-SAF, 2019a). Each version of the ESA CCI SM dataset uses the most recent and updated Metop ASCAT CDR made available by H-SAF.

6 Soil Moisture Retrieval from Passive Sensors

Contrary to the active microwave soil moisture products, which are obtained from external operational sources, soil moisture products from passive microwave sensors are produced within the CCI project itself. They are derived from level 1 brightness temperature observations using the Land Parameter Retrieval Model (LPRM; van der Schalie et al., 2015, 2017, 2018).

6.1 Principles of the Land Parameter Retrieval Model

Brightness temperatures can be derived from passive microwave sensors with different radiometric characteristics. The observed brightness temperatures are converted to soil moisture values with the Land Parameter Retrieval Model (LPRM; van der Schalie et al., 2017).This model is based on a microwave radiative transfer model that links soil moisture to the observed brightness temperatures. A unique aspect of LPRM is the simultaneous retrieval of vegetation optical depth (VOD) in combination with soil moisture and surface temperature.

A result of this physical parameterisation is that any differences in frequency and incidence angle that exist among different satellite platforms are accounted for within the framework of the radiative transfer model based on global constant parameters (de Jeu et al., 2014). This important aspect makes LPRM suitable for the development of a long-term consistent soil moisture products which can be used in the ESA CCI SM products.

The different processing steps of LPRM are described in detail in the next section, while

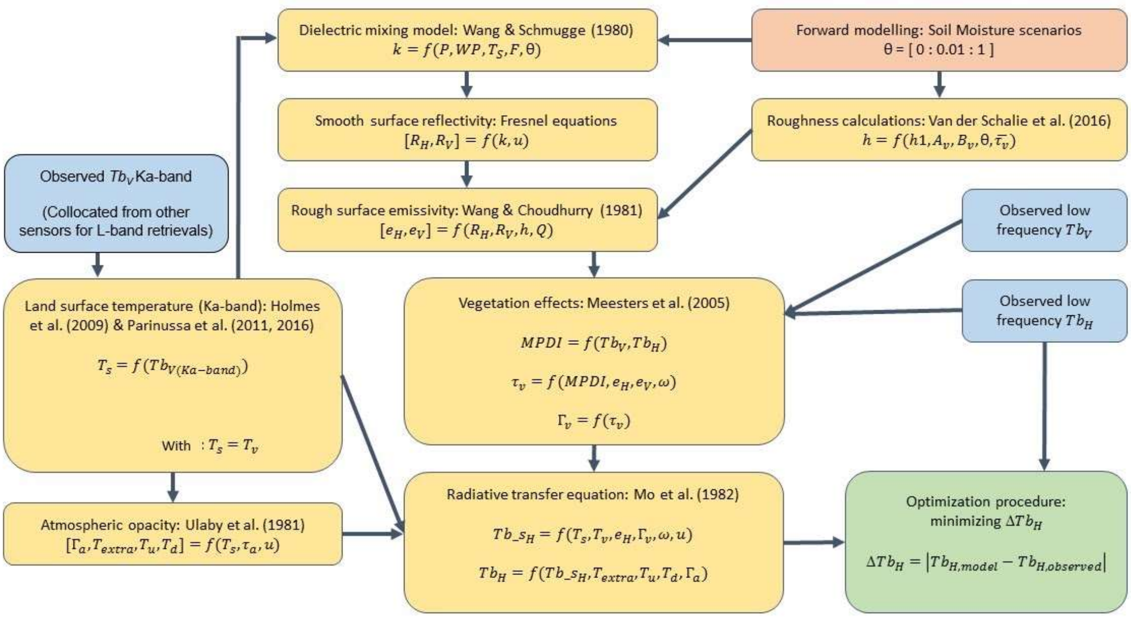


Figure 2 presents a flowchart of the entire methodology. The information presented here reflect the LPRMv6.1 product version which is used in the generation of ESA CCI SM v06.0 onwards.

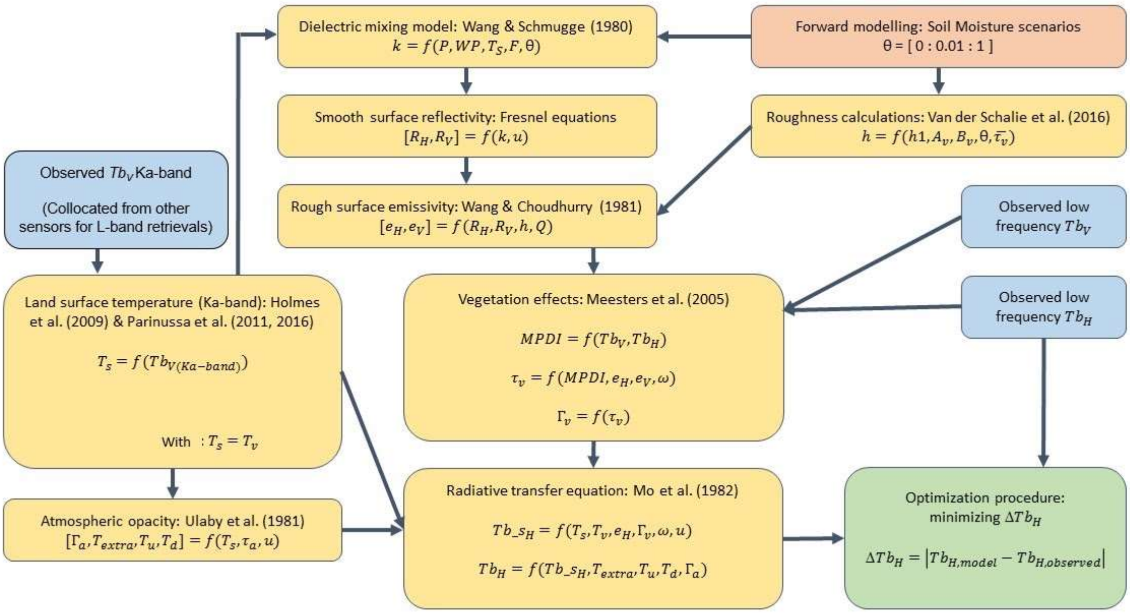


Figure 2: Flowchart of the main processes of the Land Parameter Retrieval Model (LPRM). Soil moisture is solved when the observed brightness temperature equals the modelled brightness temperature as derived by the radiative transfer.

6.1.1 Methodology

The thermal radiation in the microwave region is emitted by all natural surfaces, and is a function of both the land surface and the atmosphere. According to LPRM the observed brightness temperature (T_b) as measured by a space borne radiometer can be described as:

$$T_{b,p} = \Gamma_a (T_{b\_s,p} + (1 - e_{r,p})(T_{b\_d} + T_{b\_extra}\Gamma_a)\Gamma_v^2) + T_{b\_u} \quad \text{Eqn. 6-1}$$

Where Γ_a and Γ_v are the atmosphere and vegetation transmissivity respectively, T_{b_s} is the surface brightness temperature, e_r is the rough surface emissivity, T_{b_extra}, the extra-terrestrial brightness temperature and T_{b_u} and T_{b_d} are the upwelling and downwelling atmospheric brightness temperatures. The subscript p denotes either horizontal (H) or vertical (V) polarisation.

The vegetation/atmosphere transmissivity is further defined in terms of the optical depth, $\tau_{v/a}$, and satellite incidence angle, u, such that:

$$\Gamma_{v/a} = \exp(-\frac{\tau_{v/a}}{\cos u})$$ *Eqn. 6-2*

The upwelling brightness temperature from the atmosphere is estimated as (Bevis et al. 1992):

$$T_{b\_u,p} = 70.2 + 0.72 T_a (1 - \Gamma_a)$$ *Eqn. 6-3*

Were T_a is the atmospheric temperature. In LPRM the downwelling Temperature (T_d) is assumed to be equal to the upwelling temperature (T_u) and the extra-terrestrial temperature is set to 2.7 K (Ulaby et al. 1982).

The radiation from a land surface (T_{bp}) is described according to a simple radiative transfer (Mo et al. 1982):

$$T_{b\_s,p} = T_s e_{r,p} \Gamma_v + (1-\omega) T_v (1-\Gamma_v) + (1 - e_{r,p})(1-\omega) T_v (1-\Gamma_v)\Gamma_v$$ *Eqn. 6-4*

Where T_s and T_v are the thermodynamic temperatures of the soil and the vegetation, ω is the single scattering albedo.

LPRM uses the model of Wang and Choudhury (1981) to describe the rough surface emissivity as:

$$e_{r,p1} = 1 - Q(r_{s,p2} + (1-Q) r_{s,p1}) e^{-h\cos u}$$ *Eqn. 6-5*

Where Q is the polarisation mixing factor and h the roughness height. h is calculated using the related parameters h_1, A_v and B_v, see Eqn. 6-6, to take into account the effects of soil moisture (θ, m^3 m^{-3}) and vegetation cover (Van der Schalie et al. (2015, 2017)]) on h. $\overline{\tau_v}$ is an estimate of the vegetation density based on τ_v retrieved by calculating a primary LPRM run with A_v and B_v set to 1 and 0, with preferably a smoothing of ± 10 days applied to the τ_v to remove noise from the signal. The minimum h in LPRM is set to $h_1(B_v \overline{\tau_v})$.

$$h = h_1\,(A_v(1-2\theta) + B_v\,\overline{\tau_v})$$ *Eqn. 6-6*

r_s is the surface reflectivity and *p1* and *p2* are opposite polarisation (horizontal or vertical). The surface reflectivity are calculated from the Fresnel equations:

$$r_{s,H} = \left| \frac{\cos u - \sqrt{\varepsilon - \sin^2 u}}{\cos u + \sqrt{\varepsilon - \sin^2 u}} \right|^2$$ *Eqn. 6-7*

$$r_{s,V} = \left| \frac{\varepsilon \cos u - \sqrt{\varepsilon - \sin^2 u}}{\varepsilon \cos u + \sqrt{\varepsilon - \sin^2 u}} \right|^2$$ *Eqn. 6-8*

Where $r_{s,H}$ is the horizontal polarized reflectivity, and $r_{s,V}$ is the vertical polarized reflectivity and ε the complex dielectric constant of the soil surface ($\varepsilon=\varepsilon'+\varepsilon''i$). The dielectric constant is an electrical property of matter and is a measure of the response of a medium to an applied electric field. The dielectric constant is a complex number, containing a real (ε') and imaginary (ε'') part. The real part determines the propagation characteristics of the energy as it passes upward through the soil, while the imaginary part determines the energy losses (Schmugge et al. 1986). There is a large contrast in dielectric constant between water and dry soil, and several dielectric mixing models have been developed to describe the relationship between soil moisture and dielectric constant (Dobson et al. 1985; Mironov et al., 2004; Peplinski et al. 1995; Wang and Schmugge 1980). In 1998 Owe and Van de Griend compared the Dobson and Wang and Schmugge model and they concluded that the Wang and Schmugge model had better agreement with the laboratory dielectric constant measurements. Consequently, LPRM uses the Wang and Schmugge model, which requires information on the soil porosity (*P*) and wilting point (*WP*), observation frequency (*F*), T_S, and θ.

A special characteristic of LPRM is the internal analytical approach for solving for VOD, τ_V (Meesters et al., 2005). This unique feature reduces the required vegetation parameters to one, the single scattering albedo. LPRM makes use of the Microwave Polarisation Difference Index (MPDI) to calculate τ_V, The MPDI is defined as:

$$MPDI = \frac{T_{b\_s,V} - T_{b\_s,H}}{T_{b\_s,V} + T_{b\_s,H}} \qquad \textit{Eqn. 6-9}$$

When one assumes that τ and ω have minimal polarisation dependency at satellite scales, then the vegetation optical depth can be described as:

$$\tau_v = \cos u \ln(ad + \sqrt{(ad)^2 + a + 1}) \qquad \textit{Eqn. 6-10}$$

Where:

$$a = \frac{1}{2}\left[\frac{e_{r,V} - e_{r,H}}{MPDI} - e_{r,V} - e_{r,H}\right] \qquad \textit{Eqn. 6-6}$$

And:

$$d = \frac{1}{2}\frac{\omega}{(1-\omega)} \qquad \textit{Eqn. 6-7}$$

By using all these equations in combination with the dielectric mixing model, soil moisture can be solved in a forward model together with a parameterisation of the following parameters;

atmosphere, soil and vegetation temperature (T_a, T_s, T_c), the optical depth of the atmosphere (τ_a), the roughness parameters Q and h, soil wilting point (WP) and porosity (P), and the single scattering albedo (ω).

The temperatures were estimated using Ka-band (37 GHz) observations according to the method of Holmes et al. (2009).

For the day time (ascending) observations the following equation is used:

$$T_s = 0.898\, T_{b\_37V} + 44.2 \qquad \text{Eqn. 6-8}$$

and for the night time (descending):

$$T_s = 0.893\, T_{b\_37V} + 44.8 \qquad \text{Eqn. 6-9}$$

However, since the current L-band missions do not observe the Earth at the Ka-band frequency, they still require modelled T_S from land surface models as an input, which is something that will be improved in the near future to ensure an entirely model-independent soil moisture dataset.

The soil P and WP were derived from the ancillary 9km global soil attributes data set files from the SMAP mission (Das & O'Neill, 2020), which is a dataset that was specifically developed to support soil moisture retrievals from the SMAP mission.

Parameter	**Frequency**			
	L-band (~1.4 GHz)	**C-band (~6.9 GHz)**	**X-band (~10.8 GHz)**	**Ku-band (~19 GHz)**
τ_a	0	0.01	0.01	0.05
ω	0.12	0.075	0.075	0.06
h1 (h for Ku-band)	1.1 to 1.3	1.2	1.2	0.13
Q	0	0.115	0.127	0.14
A_V	0.7	0.3	0.3	n/a
B_V	2	2	2	n/a

Table 2: Values of the different parameters used in LPRM for the different frequencies

6.2 Known Limitations

The known limitations in deriving soil moisture from passive microwave observations are listed and described in detail in this section. These issues not only apply to the LPRM data used in the current CCI soil moisture dataset release (v06.1) but also to soil moisture retrievals from passive microwave observations in general.

6.2.1 Vegetation

Vegetation affects the microwave emission, and under a sufficiently dense canopy the emitted soil radiation will become completely masked by the overlaying vegetation. The simultaneously derived VOD can be used to detect areas with excessive vegetation, of which the boundary varies with observation frequency.

Figure 3 gives an example of the relationship between the analytical error estimate in soil moisture as described in the previous section and VOD. This figure shows larger error values in the retrieved soil moisture product for higher frequencies at similar vegetation optical depth values.

For example, for a specific agricultural crop (VOD=0.5), the error estimate for the soil moisture retrieval in the C-band is around 0.07 $m^3 \cdot m^{-3}$; in the X-band, this is around 0.11 $m^3 \cdot m^{-3}$, and in the Ku-band, this is around 0.16 $m^3 \cdot m^{-3}$. All relevant frequency bands show an increasing error with increasing vegetation optical depth. This is consistent with theoretical predictions, which indicate that, as the vegetation biomass increases, the observed soil emission decreases, and therefore, the soil moisture information contained in the microwave signal decreases (Owe et al., 2001).

In addition, retrievals from the higher frequency observations (i.e., X- and Ku-bands) show adverse influence by a much thinner vegetation cover.

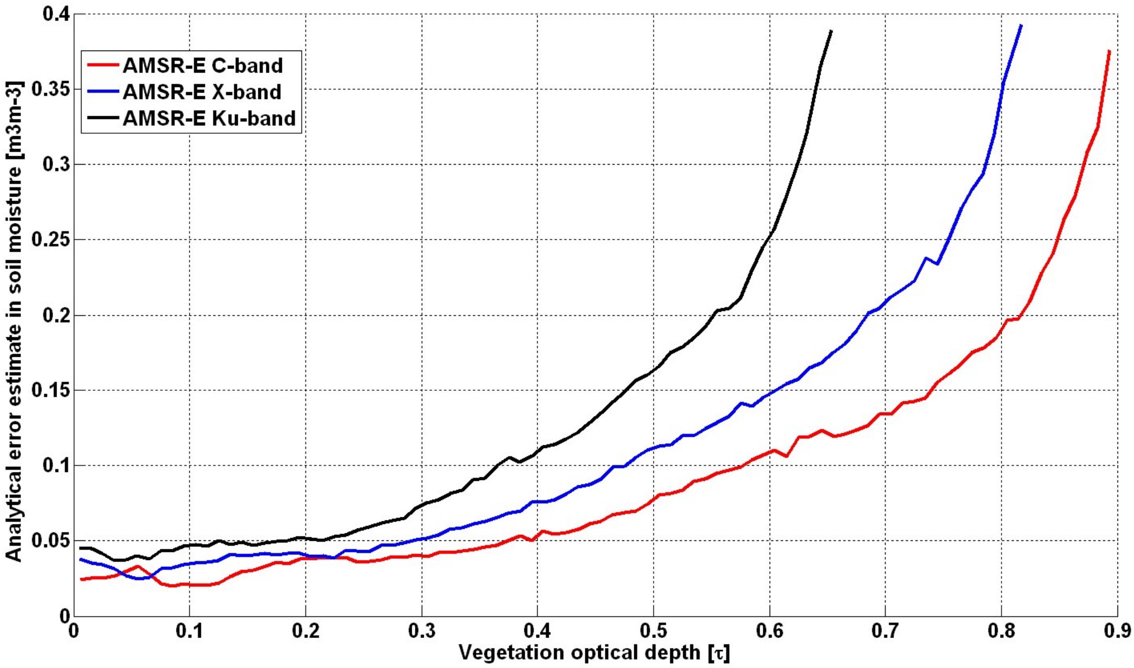
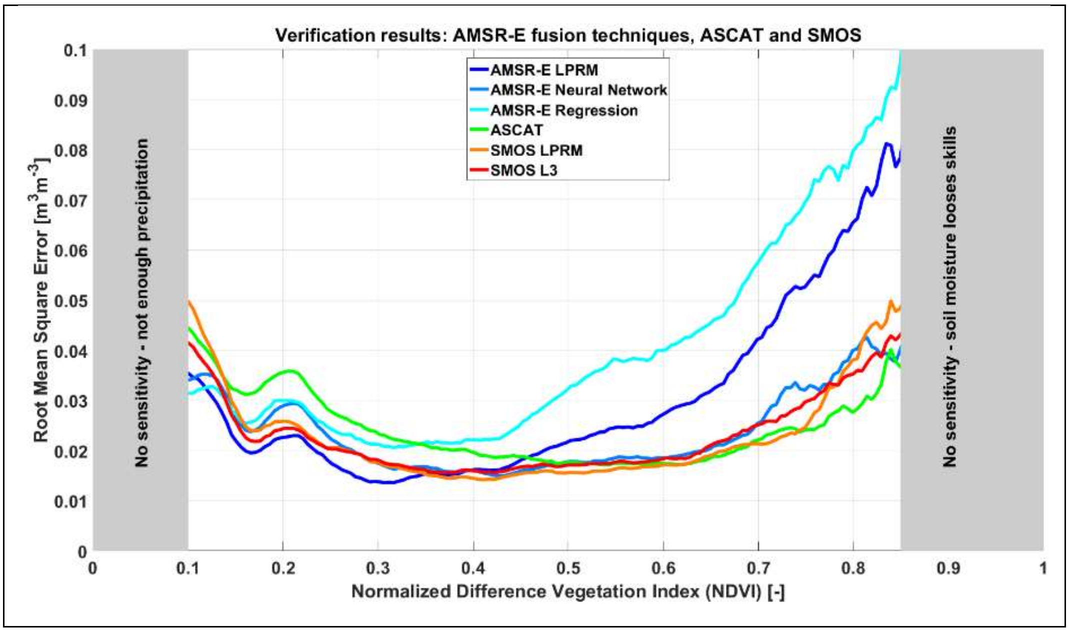


Figure 3: Error of soil moisture as related to the vegetation optical depth for 3 different frequency bands (from Parinussa et al., 2011).

For the L-band based retrievals from SMOS, the vegetation influence is less as compared to the C-, X- and Ku-band retrievals, which can be seen from the R_{value} and Triple Collocation Analysis (TCA) results in Figure 4. In Figure 4, the SMOS LPRM and AMSR-E LPRM (based on C-band) are included and shows more stable results over dense vegetation, i.e. NDVI values of over 0.45.



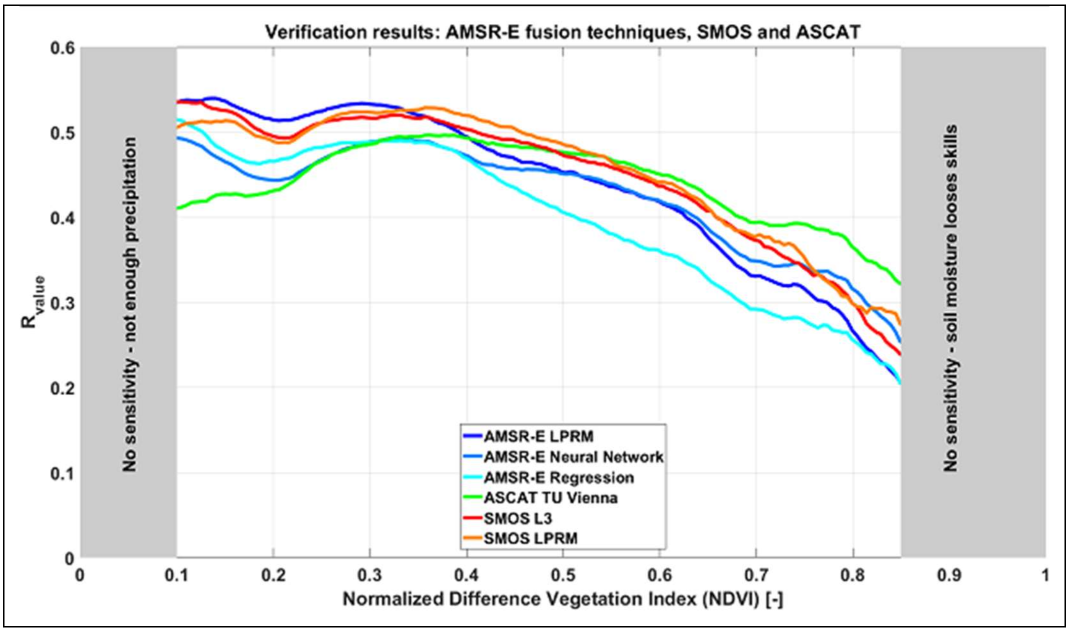


Figure 4: Triple collocation analysis (TCA: top) and R_{value} results (bottom) for several soil moisture datasets, including SMOS LPRM and AMSR-E LPRM, for changing vegetation density (NDVI). Based on (van der Schalie et al., 2018).

6.2.2 Frozen Surfaces and Snow

Under frozen surface conditions the dielectric properties of the water changes dramatically. As snow cover, ice, and frozen conditions were demonstrated to have a big impact on data quality and availability within the current Passive product, a uniform satellite driven flagging strategy was designed by Van der Vliet et al. (2020). Prior to this, all pixels where the surface temperature is observed to be at or below 274.15 K are assigned with an appropriate frozen data flag, this was determined using the method of Holmes et al. (2009). However, as this methodology is insufficiently accurate in detecting the transition to snow and frozen conditions, the new methodology was introduced by Van der Vliet et al. (2020), which uses three frequencies (Ku-, K- and Ka-band) to properly flag these conditions.

6.2.3 Water Bodies

Water bodies within the satellite footprint can strongly affect the observed brightness temperature due to the high dielectric properties of water. Especially when the size of a water body changes over time they can dominate the signal. LPRM uses a 5 % water body threshold based on MODIS observations and pixels with more than 5 % surface water are masked (Owe et al., 2008).

6.2.4 Rainfall

Rainstorms during the satellite overpass can strongly affect the brightness temperature observations, and are therefore should be flagged in LPRM. Ongoing investigations are done to define a proper filtering mechanism derived from the passive microwave observations themselves. Currently, only strong events are removed due to its effect on the retrieved temperature from Ka-band, which then drops below 274.15K.

6.2.5 Radio Frequency interference

Natural emission in several low frequency bands are affected by artificial sources, so called Radio Frequency Interference (RFI). As a diagnostic for possible errors an RFI index is calculated according to De Nijs et al. (2015). Most passive microwave sensors that are used for soil moisture retrieval observe in several frequencies. This allows LPRM to switch to higher frequencies in areas affected by RFI.

The new methodology that is used since LPRMv6 for RFI detection uses the estimation of the standard error between two different frequencies. It uses both the correlation coefficient between two observations and the individual standard deviation to determine the standard error in Kelvin. A threshold value of 3 Kelvin is used to detect RFI. This method does not produce false positives in extreme environments and is more sensitive to weak RFI signals in relation to the traditional methods (e.g. Li et al., 2004).

As the currently integrated SMOS mission does not have multiple frequencies to apply this method, here we base the filtering on the RFI probability information that is supplied by in the SMOS Level 3 data. SMAP, by using different channels around 1.4GHz, already has an internal mitigation of RFI that removes almost all occurrence of RFI, therefore no extra filtering is needed for use with the ESA CCI SM.

6.2.6 *Updated temperature input from Ka-band observations*

The land surface temperature plays a unique role in solving the radiative transfer model and therefore directly influences the quality of the soil moisture retrievals. The current linear regression to link Ka-band measurements to the effective soil temperature has been adjusted and optimized by Parinussa et al. (2016) for day-time observations. This is done using an optimisation procedure for soil moisture retrievals through a quasi-global precipitation-based verification technique, the so-called R_{value} metric. In this optimisation, different biases were locally applied to the existing linear regression and final results have been used to create an updated global linear regression. The focus on this study was to improve the skill to capture the temporal dynamics of the soil moisture. After the updated linear regression for the land surface temperature, the R_{value} increased on average with 16.5% and the triple collocation analysis showed an average reduction in RMSE of 15.3%. This shows an improved skill in daytime retrievals from LPRM and giving way to using both daytime and night-time retrievals together in the future.

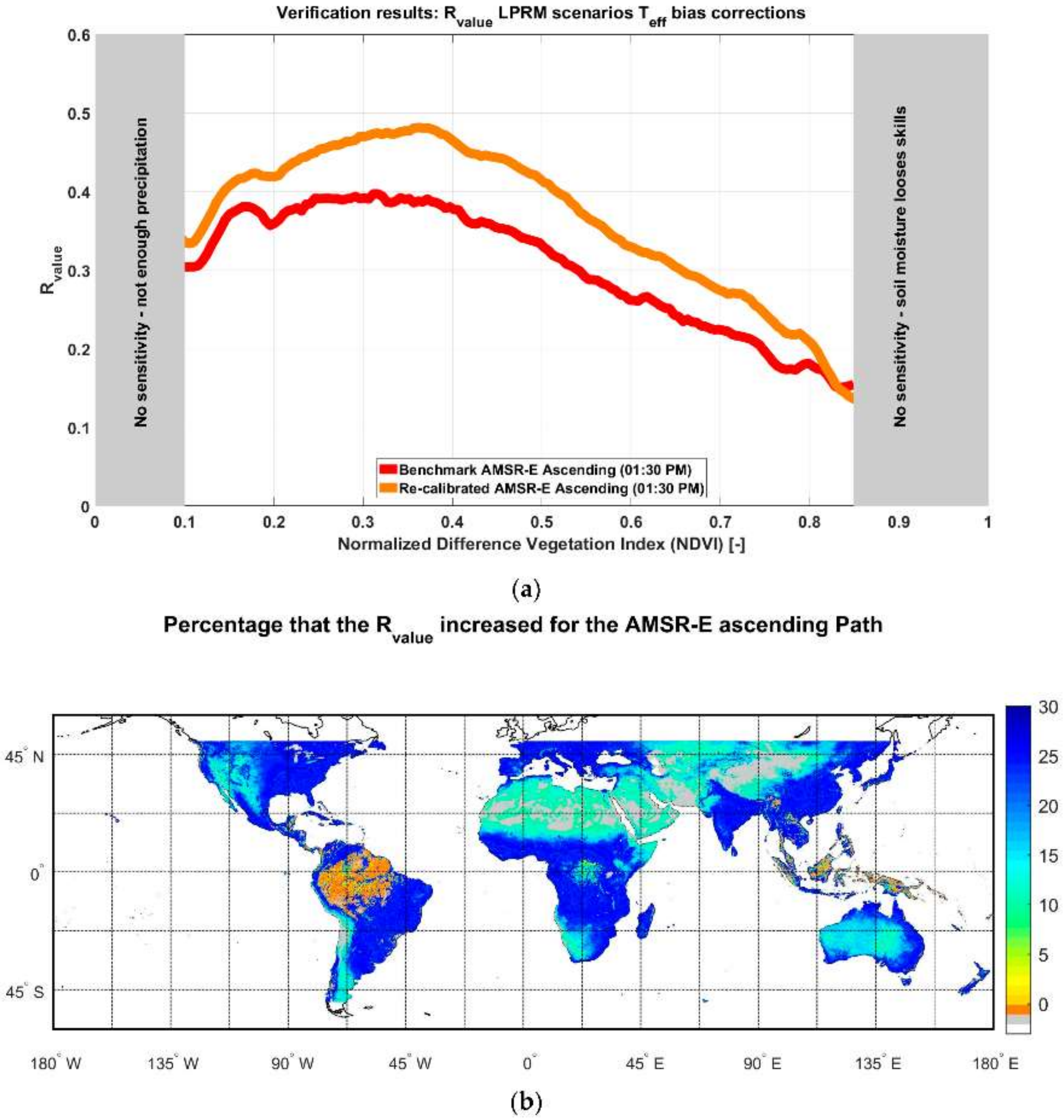


Figure 5: (up) comparison of R_{value} with the old and new daytime land surface temperature binned over NDVI, (down) the difference in R_{value} compared to the old temperature parameterisation in [%].

This explorative work showed the high impact temperature has on the quality of the LPRM retrievals. Secondly, there are issues with (seasonal) bias in the effective temperature derived from Ka-band using a linear regression method, which are caused by the seasonal changes in soil moisture, vegetation cover and atmospheric composition. Therefore upcoming research will focus on a temperature retrieval that, similar to soil moisture, uses a radiative transfer model and includes corrections for the before mentioned issues.

Also, in order to remove model dependency for the L-band soil moisture retrievals, we have collocated Ka-band observations from other satellites, to SMAP and SMOS. This way, the temperature input to the L-band missions are done similar as the other missions, while simultaneously be able to apply the new filtering methodology developed by Van der Vliet et al. (2020).

7 ESA CCI SM Merging Algorithm

7.1 Principles of the Merging Process

The ESA CCI SM project provides three soil moisture datasets:

1. ACTIVE – a dataset generated from active scatterometers using the TU Wien change detection algorithm, with the ASCAT datasets provided through H-SAF (1991 – 2020)
2. PASSIVE – a dataset generated from passive radiometers using the LPRM algorithm (1978 – 2020)
3. COMBINED – a dataset incorporating all of the sensors included in the ACTIVE and PASSIVE datasets (1978-2020)

Details of the sensors used in the generation of the ESA CCI SM datasets are provided in Table 3 and Table 4with a summary provided in

Figure 6.

The homogenised and merged products represent surface soil moisture with a global coverage and a spatial resolution of 0.25°. The time period spans the entire period covered by the individual sensors, i.e. 1978 – 2020, while measurements are provided at a 1-day sampling.

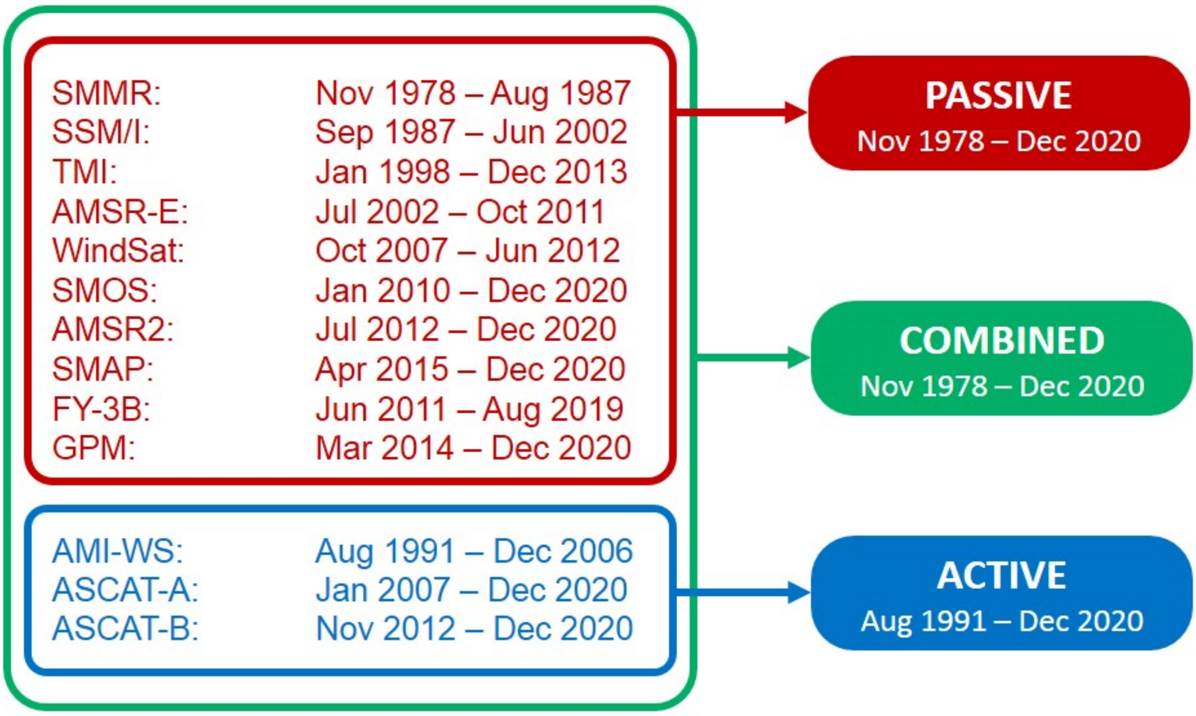


Figure 6: Overview of the time periods used for each of the products of ESA CCI SM v06.1.

Table 3: Major characteristics of passive microwave instruments used in ESA CCI SM

	Passive microwave products									
	SMMR	**SSM/I**	**TMI**	**AMSR-E**	**AMSR2**	**Windsat**	**MIRAS**	**SMAP**	**GMI**	**MWRI**
Platform	Nimbus 7	DMSP	TRMM	Aqua	GCOM-W1	Coriolis	SMOS	SMAP	GPM	FY-3B
Product	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)	LPRM (VanderSat)
Product Source	NASA, Tape derived	NASA EarthData, XCAL calibrated with GPM	NASA EarthData, XCAL calibrated with GPM	JAXA, G-portal	JAXA, G-portal	Bespoke order	CATDS	NASA EarthData	NASA EarthData, XCAL calibrated with GPM	nsmc
Algorithm Product version	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)	LPRM v06.1 (2)
Time period used	11/1978–8/1987	09/1987–12/2007	01/1998–12/2013	07/2002–10/2011	05/2012–12/2019	10/2007–7/2012	01/2010–12/2020	04/2015-12/2020	03/2014 – 12/2020	06/2011 – 08/2019
Channel used for soil moisture	6.6 GHz	19.3 GHz	10.7 GHz	6.9/10.7 GHz	6.925/10.65 GHz	6.8/10.7 GHz	1.4 GHz	1.4GHz	10.7 GHz	10.7 GHz
Original spatial resolution(1) (km^2)	150×150	69 × 43	59 × 36	76 × 44	35 x 62	25 x 35	40 km	38 x 49	19x32	51 x 85
Spatial coverage	Global	Global	N40° to S40°	Global	Global	Global	Global	Global	N70° to S70°	Global
Swath width (km)	780	1400	780/897 after boost in Aug 2001	1445	1450	1025	600	1000	931	1400
Equatorial crossing time	Descending: 0:00	Descending: 06:30	Varies (non polar-orbiting)	Descending: 01:30	Descending 01:31	Descending 6:03	Ascending 6:00	Descending 06:00	Varies (non polar-orbiting)	Descending: 01:30
Unit	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}	m^3m^{-3}

(1) For passive microwave instruments, this stands for the footprint spatial resolution.
(2) LPRM v6.1 references: van der Schalie et al. (2015, 2017, 2018), van der Vliet (2020)

Table 4: Major characteristics of active microwave instruments and model products used in ESA CCI SM

	Active microwave products				Model product	
	AMI-WS	AMS-WS	ASCAT	ASCAT	GLDAS-2-Noah	GLDAS-2-Noah
Platform	ERS1/2	ERS2	Metop-A	Metop-B	---	---
Product	SSM Product (TU WIEN, 2013)	SSM Product (Crapolicchio et al., 2016)	H 115/116 (H-SAF 2019a and 2019b)	H 115/116 (H-SAF 2019a and 2019b)	---	---
Algorithm Product version	TU WIEN Change Detection [2]	TU WIEN Change Detection [3]	TU WIEN Change Detection [3]		V2.0	V2.1
Time period used	7/1991–12/2006	5/1997–2/2007	1/2007–12/2020	11/2012–12/2020	1/1948– 12/2010	1/2000–12/2020
Channel used for soil moisture	5.3 GHz	5.3 GHz	5.3 GHz	5.3 GHz	---	---
Original spatial resolution[1] (km^2)	50 × 50	25 x 25	25 × 25	25 × 25	25 × 25	25 × 25
Spatial coverage	Global	Global	Global	Global	Global	Global
Swath width (km)	500	500	1100 (550×2)	1100 (550×2)	---	---
Equatorial crossing time	Descending: 10:30	Descending 10:30	Descending: 09:30	Descending: 09:30	---	---
Unit	Degree of saturation (%)	Degree of saturation (%)	Degree of saturation (%)	Degree of saturation (%)	kg m^{-2}	kg m^{-2}

(1) For active microwave instruments, this stands for the footprint spatial resolution.
(2) TU Wien change detection algorithm references for AMI-WS: Wagner et al. (1999)
(3) H-SAF references for ASCAT: H-SAF (2019a, 2019b)

7.2 Algorithm Description

The level 2 surface soil moisture products derived from the active and passive remotely sensed data undergo a number of processing steps in the merging procedure (see Figure 7 for an overview):

1. Spatial resampling and temporal resampling (including flagging and cross-flagging of observations)
2. Rescaling passive and active level 2 observations into radiometer and scatterometer climatologies (for the ACTIVE and PASSIVE product), and separately rescaling all level 2 observations into a common model-based climatology (for the COMBINED product)
3. Triple collocation analysis (TCA)-based error characterisation of all rescaled level 2 products
4. Polynomial regression between VOD and error estimates to fill spatial gaps where errors could not be reliably retrieved i.e., where TCA is deemed unreliable
5. Merging rescaled passive and active time series into the PASSIVE, ACTIVE, and COMBINED products, respectively.

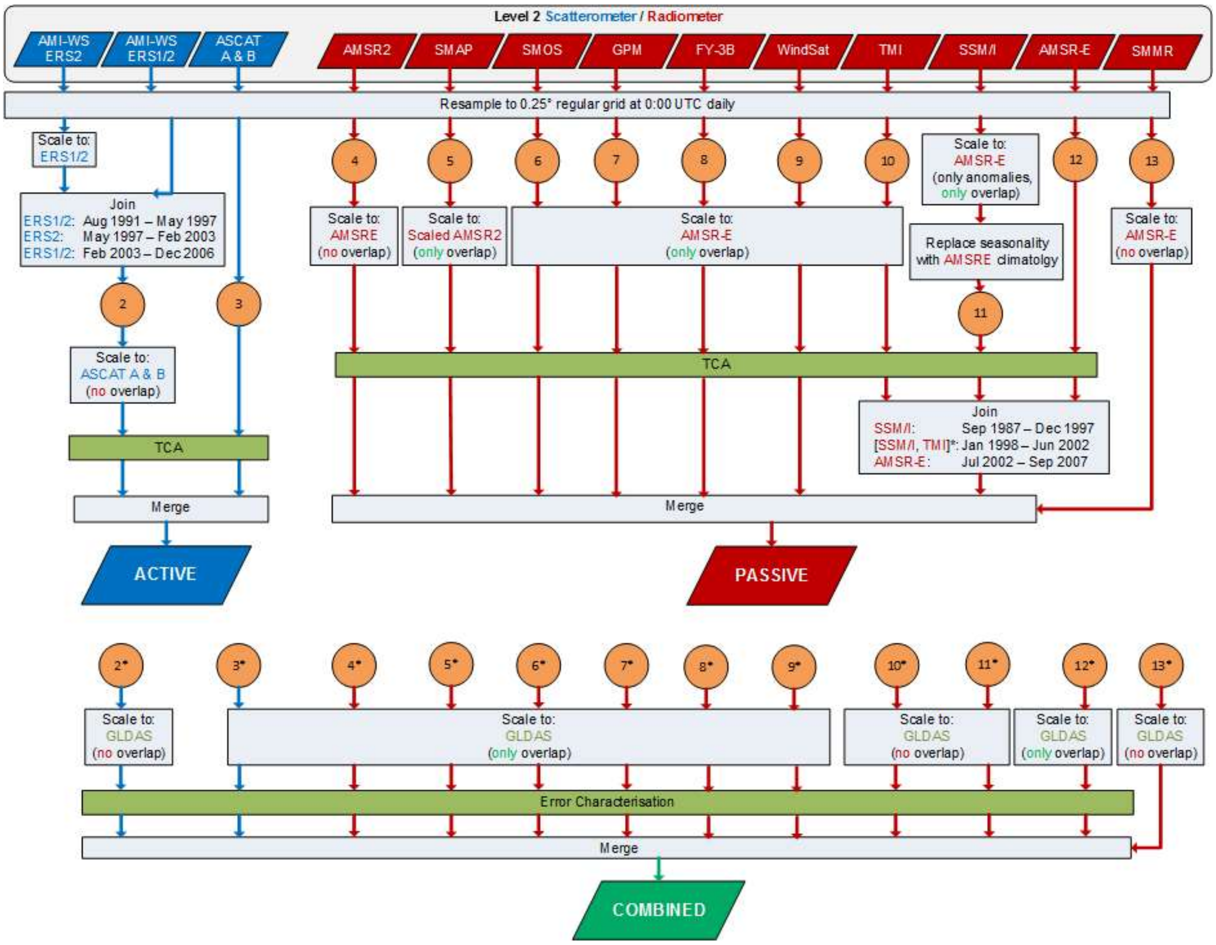


*Figure 7: Overview of the processing steps in the ESA CCI SM product generation (v06.1): The merging of two or more data sets is done by weighted averaging and involves overlapping time periods, whereas the process of joining data sets only concatenates two or more data sets between the predefined time periods. The join process is performed on datasets of each lines and on datasets separated by comma within the rectangular process symbol. *The [SSM/I, TMI] period is specified not only by the temporal, but also by the spatial latitudinal coverage (see Figure 9).*

In this section the algorithms of the scaling and merging approach are described. Notice that several algorithms, e.g. rescaling, are used in various steps of the process, but will be described only once.

7.2.1 Resampling

The sensors used for the different merged products have different technical specifications (Table 3 and Table 4) with clear differences in spatial resolution and crossing times. Both elements need to be brought into a common reference before the actual merging can take place.

Spatial Resampling

The final CCI SM merged products are provided on a regular grid with a spatial resolution of 0.25° in both latitude and longitude extension. This is a trade-off between the higher resolution scatterometer data and the generally coarser passive microwave observations without leading to any under-sampling. The same resolution is often adopted by land surface models.

For the LPRM passive data, nearest neighbour resampling is performed on the radiometer input data sets to bring them into the common regular grid. Following this resampling technique each grid point in the reference (regular grid) data set is assigned to the value of the closest grid point in the input dataset. In general, the nearest neighbour resampling algorithm can be applied to data set with regular degree grid.

For the active microwave data sets, where equidistant grid points are defined by the geo-reference location of the observation, the hamming window function is used to resample the input data to a 0.25° regular grid. The search radius is a function of latitude of the observation location, as the distance between two regular grid points reduces as the location tends towards the poles. In contrast, the active microwave data set uses the discrete global grid (DGG), where the distance between every two points is the same. This main difference between the DGG (active) and the targeted regular degree grid is rectified by using a hamming window with search radius dependent on the latitude for the spatial resampling of the active microwave data.

Temporal Resampling

The temporal sampling of the merged product is 1 day. The reference time for the merged dataset is set at 0:00 UTC. For each day starting from the time frame center at 0:00 UTC observations within ±12 hours are considered.

The temporal resampling strategy firstly searches for the valid observation that is closest to the reference time. In the case that there are only invalid observations, which are flagged other than “0” (zero), within this time frame, the closest measurement among these invalid observations is selected. In the event that there are no measurements available at all within a time frame, no value is assigned to that day. This strategy results in data gaps when no observations within ±12 hours from the reference time are available.

Flagging

During the temporal resampling stage, flagging is applied to the datasets where relevant information is available. The key flags set during this process are ‘frozen’, ‘high VOD’ and ‘Other’ and these flags are propagated through the entire processing chain to the final product.

The ASCAT and ERS products include a Surface State Flag (SSF) which effectively encodes information about whether or not the surface is frozen or snow-covered. In the ESA CCI SM product, those soil moisture values where the SSF is greater than 2^1 (i.e. 2) are used to flag the observation as frozen. The ASCAT and ERS products do not provide information on high VOD.

The LPRMv6.1 products (which are used in the production of v06.1) provide a FLAGS field which provides information on high VOD, frozen conditions and the performance of the LPRM algorithm. The thresholds above which VOD is considered 'high' are set based on the saturation point in the VOD signal for each sensor and band. This is the point at which the VOD value is considered to equal 100% vegetation signal. Secondly, the frozen/snow flag was applied using the new approach by Van der Vliet et al. (2020), which derives the frozen/snow conditions from Ku-, K- and Ka-band observations.

New at v06.1 is the implementation of cross-flagging. This means that any frozen flags provided in any of the datasets are effectively transferred to all of the datasets. It works by reading in all of the flag data for all of the datasets, determining if the frozen flag is set in any of them and then if it is, applying it to all those observations in all of the sensors.

7.2.2 Merging ASCAT

In the time period from 6 November 2012 to 2020-12-31 Metop-A ASCAT and Metop-B ASCAT data are available. These two datasets are merged by applying the arithmetic average for locations, where both observations are available, otherwise either one of the two is then used.

7.2.3 Rescaling

Rescaling Methodology

Due to different observation frequencies, observation principles, and retrieval techniques, the contributing soil moisture datasets are available in different observation spaces. Therefore, before merging can take place, the datasets need to be rescaled into a common climatology.

Scaling is performed using cumulative distribution function (CDF) matching which is a well-established method for calibrating datasets with deviating climatologies (Drusch et al., 2005; Liu et al., 2007; Liu et al., 2011; Reichle et al., 2004, Moesinger et al., 2020). CDF-matching is applied for each grid point individually and based on piece-wise linear matching. This variation of the CDF-matching technique proved to be robust also for shorter time periods (Liu et al., 2011). CDF-matching is performed in the following way:

1. For the time-collocated data points CDFs are computed.
2. If more than 400 time-collocated data points exist, for each CDF curve the 0, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95 and 100 percentiles are identified. Otherwise, evenly

spaced percentile bins are generated such that each of them contains at least 20 observations.

3. Use the percentiles of the CDF curves to define n-1 segments.
4. The n percentile values from the AMSR-E and ASCAT CDF curves are plotted against those of Noah and scaling linear equations (e.g., slope and intercept) between two consecutive percentiles are computed.

$$slope_i = \frac{pref_{i+1} - pref_i}{psrc_{i+1} - psrc_i}$$

$$intercept_i = pref_i - (psrc_i * slope_i)$$

where i=1..12, is the number of the segments, and $pref$ is the percentile of the GLDAS-Noah data (reference), and $psrc$ is the percentile of either AMSR-E or ASCAT data (source) respectively.

5. An exception are the first and last segment. Instead of using the first and last percentile for interpolation, the slope is derived using least squares regression. This is more robust to outliers.
6. The obtained linear equations are used to scale all observations of the target data set (i.e., also the time steps that do not have a corresponding observation in the reference data set) to the climatology of the reference data set.

$$sm_r = slope_i * sm + intercept_i$$

where sm_r is the rescaled soil moisture and sm is the original soil moisture value. $slope_i$ and $intercept_i$ are chosen depending on the sm value and its corresponding i-percentile.

The AMSR-E and ASCAT values outside of the range of CDF curves can also be properly rescaled, using the linear equation of the closest value.

Rescaling of Active Datasets

Different sensor specifications between ERS1/2 and ERS2 (e.g. spatial resolution) need to be compensated for using scaling. The CDF curves for ERS2 are calculated based on the overlap with ERS1/2. Rescaling ERS2 against ERS1/2 and then joining them generates the AMI-WS active data set, which is subsequently scaled to the Metop-A ASCAT data (ACTIVE product) or the GLDASv2.1 data (COMBINED product) (see Figure 7).

For the ACTIVE product, the limited overlap between AMI-WS ERS1/2 and Metop-A ASCAT in time (i.e., a few months) rules out the global adjustment method based on the information of their overlapping period. However, as retrievals from Metop-A ASCAT and AMI-WS capture similar seasonal cycles (Liu et al., 2011), we assume that their dynamic ranges are identical and therefore, can use non-overlapping observations for the rescaling (i.e. the entire time period for each sensor).

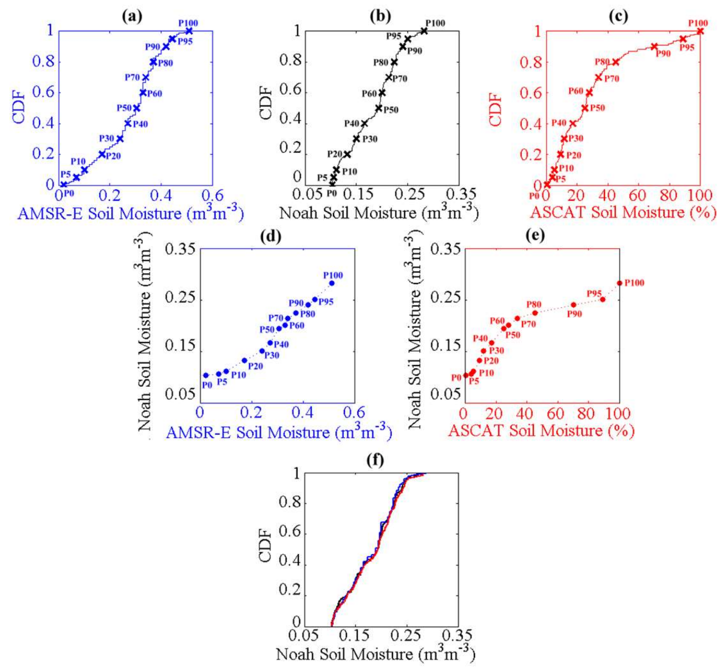
Rescaling of Passive Datasets

The seasonal cycle associated with the SSM/I dataset is deemed to be unreliable and therefore, for all CCI products, the SSM/I seasonal cycle is replaced with that from AMSR-E. The high frequency variations (anomalies) associated with SSM/I are scaled to those from AMSR-E prior to recombining the decomposed signal. An example of the SSM/I decomposition and rescaling is shown in Figure 8.

For the PASSIVE product, all datasets with the exception of SMAP are rescaled to AMSR-E. Where sufficient overlap is available, this is utilised; for all other cases (except AMSR2), the entire time period of AMSR-E and the sensor being scaled is utilised. For AMSR2, data in the last three years of AMSR-E and the first three years of ASMR2 are used, i.e. 2008-10-04 to 2015-07-01. SMAP is rescaled to AMSR2 which has already been rescaled to AMSR-E.

Rescaling in the COMBINED product

For generating the combined product, all passive and active level 2 data sets are rescaled against GLDASv2.1, with the exception of ERS1/2, ASCAT and SSMI which are discussed above.



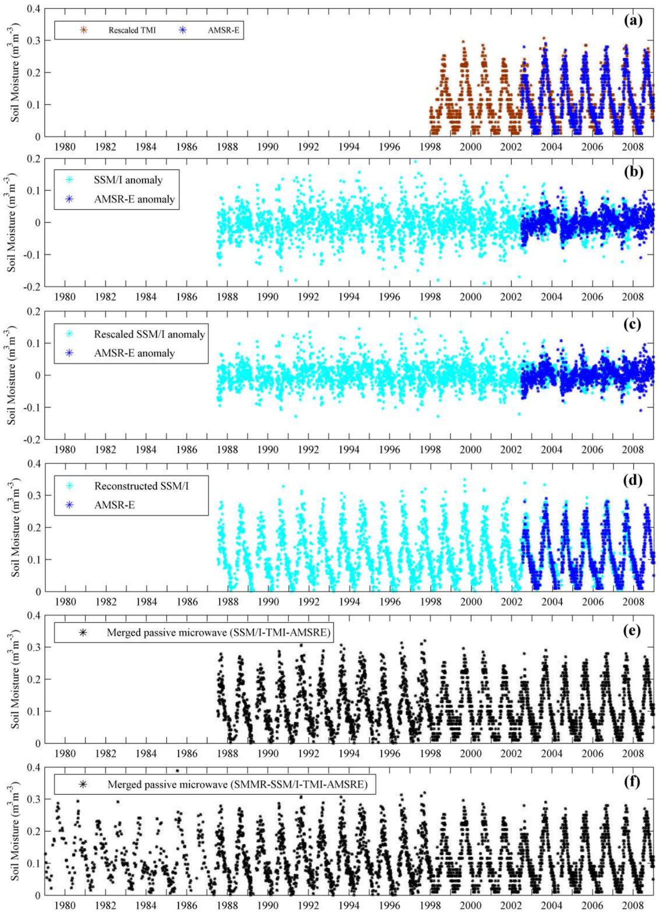


Figure 8: Example illustrating how (a) the TMI was rescaled against AMSR-E, (b-e) the SSM/I anomalies were rescaled against AMSRE-E anomalies, reconstructed and merged with rescaled TMI and AMSR-E, and (e) the SMMR was rescaled and merged with the others. The grid cell is centred at 13.875°N, 5.875°W (Image courtesy Liu et al., 2012).

7.2.4 Error characterisation

Errors in the individual active and passive products are characterized by means of triple collocation analysis (TCA). These errors are used both for estimating the merging parameters and for characterising the errors of the merged product (see section 7.2.5).

TCA is a statistical tool that allows estimating the individual random error variances of three data sets without assuming that any of them are acting as supposedly accurate reference (Gruber et al., 2016). This method requires the errors of the three data sets to be uncorrelated, therefore triplets always comprise of (i) an active data set, (ii) a passive data set, and (iii) the GLDAS-Noah land surface model, which are commonly assumed to fulfil this requirement (Dorigo et al., 2010). Error variance estimates are obtained as:

$$\sigma^2_{\varepsilon_a} = \sigma^2_a - \frac{\sigma_{ap}\sigma_{am}}{\sigma_{pm}}$$
$$\sigma^2_{\varepsilon_p} = \sigma^2_p - \frac{\sigma_{pa}\sigma_{pm}}{\sigma_{am}} \qquad \textit{Eqn. 7-1}$$

where σ^2_ε denotes the error variance; $\sigma^2$and σ denote the variances and covariances of the data sets; and the superscripts denote the active (a), the passive (p), and the modelled (m) data sets, respectively. For a detailed derivation see Gruber et al. (2016). Note that these error estimates represent the average random error variance of the entire considered time period, which is commonly assumed to be stationary. Furthermore, the soil moisture uncertainties of the three products (ACTIVE, PASSIVE, and COMBINED) are determined by the above equations, hence the uncertainty estimates provided with the product apply to an entire merging period and do not vary with every timestamp.

7.2.5 Error Gap-Filling

TCA does not provide reliable error estimates in all regions, mainly if there is no significant correlation between all members of the triplet, which often happens for example in high-latitude areas or in desert areas. TCA error estimates are therefore disregarded where the Pearson correlation between any of the data sets is deemed insignificant (i.e. p-value < 0.05).

In these areas, error estimates are derived by deriving an SNR-VOD regression model per land cover class and using this to determine the SNR based on the VOD at each location where SNR could not be retrieved:

$$SNR_x = \sum_{\{i=0\}}^{N} a_i VOD_x^i \qquad \textit{Eqn. 7-2}$$

Where the subscript denotes the spatial location; and the parameters a_i are derived from the polynomial regression. For TMI and WINDSAT third order polynoms (N=3) are used and for all other sensors second order polynoms (N=2) are used, which was empirically found to provide the best regression results.

7.2.6 Merging

The merging procedure is undertaken separately for each of the ESA CCI SM products (ACTIVE, PASSIVE and COMBINED) from the rescaled L2 products. The merging periods used in each product are shown in Figure 9 and details of each merging period are listed in Table 5 (ACTIVE0,) Table 6 (PASSIVE) and (COMBINED).

Considering the covering period of each microwave instrument we divided the entire time period (1978 – 2020-12-31) into eleven segments. Table 5, Table 6 and Table 7 list these time periods, and Figure 9c illustrates also the spatial sensor usage at global scale.

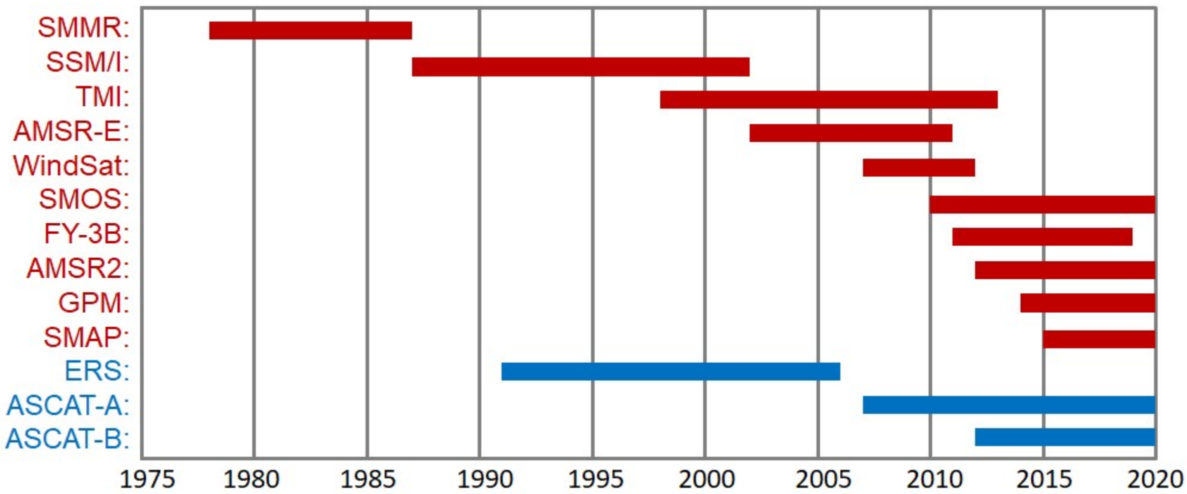


Figure 9: Spatial and temporal coverage of different sensors in the CCI SM v06.1 products. All sensors contribute to the COMBINED product, those in red contribute to the PASSIVE product and those in blue contribute to the ACTIVE product. Note that data from the TMI sensor is only available between -37° and +37°.

Table 5 Active sensors used in the ACTIVE products

Time Periods	Active Sensors
05/08/1991 – 19/05/1997	ERS1/2 (AMI-WS)
20/05/1997 – 17/02/2003	ERS2 (AMI-WS)
18/02/2003 – 31/12/2006	ERS1/2 (AMI-WS)
01/01/2007 – 05/11/2012	Metop-A ASCAT
06/11/2012 – 2020-12-31	Metop-A ASCAT, Metop-B ASCAT

Table 6 Passive sensors in the PASSIVE product. Note: a = ascending, d = descending.

Time Period	Passive Sensors
01/11/1978 – 31/07/1987	SMMR (d)

01/09/1987 – 31/12/1997	SSM/I (a)
01/01/1998 – 18/06/2002	SSM/I (a) [90N – 37N], [90S – 37S], TMI (a/d) [37N – 37S]
19/07/2002 – 30/09/2007	AMSR-E (d)
01/10/2007 – 14/01/2010	AMSR-E (d), Windsat (d)
15/01/2010 – 04/10/2011	AMSR-E (d), WindSat (d), SMOS (a)
05/10/2011 – 30/06/2012	WindSat (d), SMOS (a)
01/07/2012 – 2015/03/30	SMOS (a), AMSR2 (d)
2015/03/31 – 2020-12-31	SMOS (a), AMSR2 (d), SMAP (d)

Table 7 Sensors used in the COMBINED product in individual time periods. Note: a = ascending, d = descending.

Time Periods	Active Sensors	Passive Sensors
01/11/1978 – 31/08/1987	N/A	SMMR
01/09/1987 – 04/08/1991	N/A	SSM/I
05/08/1991 – 31/12/1997	AMI-WS	SSM/I
01/01/1998 – 18/06/2002	AMI-WS	SSM/I [90N-37N], [90S-37S], TMI [37N-37S]
19/06/2002 – 31/12/2006	AMI-WS	AMSR-E
01/01/2007 – 30/09/2007	Metop-A ASCAT	AMSR-E
01/10/2007 – 30/06/2010	Metop-A ASCAT	AMSR-E, WindSat
15/07/2010 – 04/10/2011	Metop-A ASCAT	AMSR-E, WindSat, SMOS
05/10/2011 – 30/06/2012	Metop-A ASCAT	WindSat, SMOS
01/07/2012 – 30/03/2015	Metop-A ASCAT, Metop-B ASCAT	AMSR2, SMOS
31/03/2015 – 2020-12-31	Metop-A ASCAT, Metop-B ASCAT	AMSR2, SMOS, SMAP

Weight estimation

The merging is performed by means of a weighted average which takes into account the error properties of the individual data sets that are being merged:

$$\Theta_m = \sum_{i=1}^{N} w_i \cdot \Theta_i \qquad \text{Eqn. 7-2}$$

Where Θ_m denotes the merged soil moisture product; Θ_i are the soil moisture products that are being merged, and $\mathrm{w_i}$ are the merging weights.

Per definition, the optimal weights for a weighted average are determined by the error variances of the input data sets and write as follows:

$$w_i = \frac{\sigma_{\varepsilon_i}^{-2}}{\sum_{j=1}^{N} \sigma_{\varepsilon_j}^{-2}} \qquad \text{Eqn. 7-3}$$

where the superscripts denote the respective data sets; i is the data set for which the weight is being calculated; and N is the total number of data sets which are being averaged. The required error variances are calculated using Eqn. 7-3. Notice that error covariance terms are neglected as they cannot be estimated reliably.

It should be mentioned that the above definition of the weights based on error variances assumes all data sets to be in the same data space. However, data sets usually vary in their signal variability due to algorithmic differences, varying signal frequencies, etc. Therefore, conceptually, it is more appropriate to define relative weights in terms of the data sets SNR properties rather than of their error variance (Gruber et al., 2017). Nevertheless, the actual merging requires a harmonisation of the data sets into a common data space, which in the case of the CCI SM data set is done using the CDF matching approach described in Section 7.2.2. Therefore, the calculation of the weights using Eqn. 7-3 suffices, keeping in mind that they represent rescaled error variances of rescaled data sets.

Notice that soil moisture estimates of the various sensors are not available every day, hence there are certain dates during the overlapping periods on which not all data sets provide a valid estimate to calculate the weighted average. In such cases, the weights are re-distributed amongst the remaining data sets, again based on their relative SNR properties.

However, this re-distribution of weights could significantly worsen data quality on these days because of the increasing contribution of measurements which initially would have had a low weight due to their (relatively) low SNR. Therefore, soil moisture estimates in the merged product on days where not all data sets provide valid estimates are set to NaN values (Not a Number), if the sum of the initial weight of the remaining data sets is lower than 1/(2N) where N is the total number of data sets that are potentially available for the corresponding merging period. This threshold has been derived empirically to provide a good trade-off between temporal measurement density and average data quality.

Similar to the generation of the PASSIVE product, relative weights at each time step are derived from the TCA- or VOD-regression based error estimates for each individual sensor. Depending on how many sensors are available within a particular period, a (1/2N) threshold for the minimum weight of a particular sensor was applied if not all sensors provide a soil moisture estimate at that day.

7.2.7 *Break detection and correction*

At ESA CCI SM v06, a new ancillary product is provided in addition to the main ESA CCI SM products. This is a version of the COMBINED product which has been subject to break detection and correction methods using reanalysis soil moisture as the reference data set.

Breaks may occur as a result of merging different sensor combinations over time, as shown in Figure 10. Such breakpoints may therefore appear between periods with different input sensors. Structural inhomogeneities may affect statistics such as trends and changes in extreme values (percentiles) and therefore should not only be detected but also corrected.

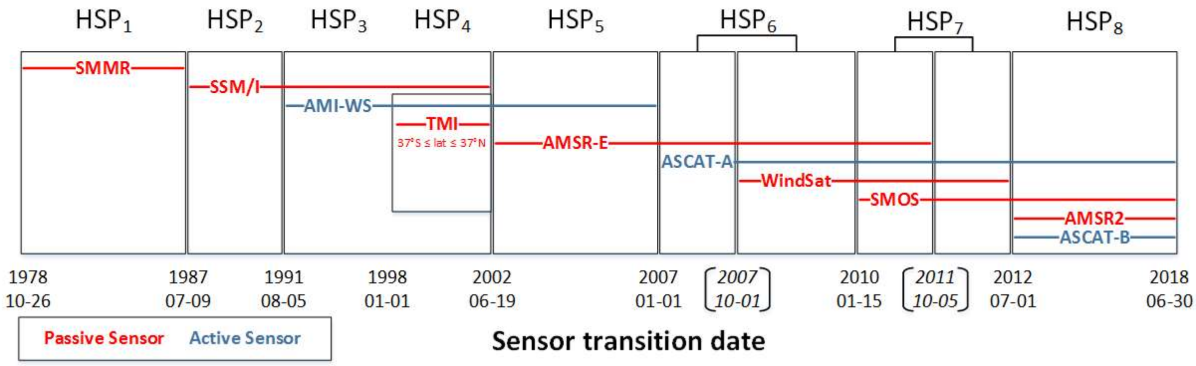


Figure 10: Potential break times in the ESA CCI SM v04.4 (COMBINED) product- corresponding to changeovers in the blended sensors, building the homogeneous (sensor) sub-periods (HSP).

Based on the work of Su et al. (2016), a procedure has been developed at TU Wien to relatively test for potential inhomogeneities in the ESA CCI SM Climate Data Record (CDR) using the Fligner-Killeen test for homogeneity of variances and Wilcoxon rank-sums test for shifts in population mean ranks (Preimesberger et al., 2020). For the product provided with ESA CCI SM v06.1, the reference dataset used is ERA-5.

To adjust detected breaks in the data set, Quantile Category Matching (QCM) is used. This method uses split-fitted differences in empirical CDFs of the ESA CCI SM and reference SM (between quantile categories, i.e. average SM within a number of quantile ranges) values before and. after a break are used to find corrections for quantiles of ESA CCI SM before the break.

Adjustment is performed iteratively, with the goal that across each detected break, changes in ESA CCI SM means and variances are matched to follow changes within the reference data set (relative bias correction) and homogenised observation series (with respect to the reference data set) are derived

The results of the correction performed on v04.4 of the CCI dataset are shown in Figure 11. Figure 12 shows the longest homogenous period of available data both before (top) and after (bottom) correction using the QCM method (which leads to the lowest number of re-detected breaks after correction from the three described methods).

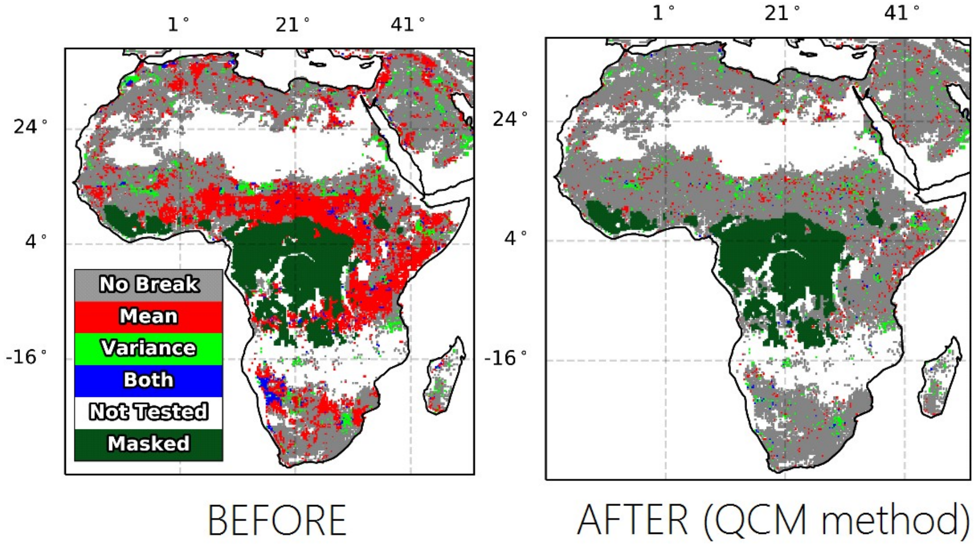
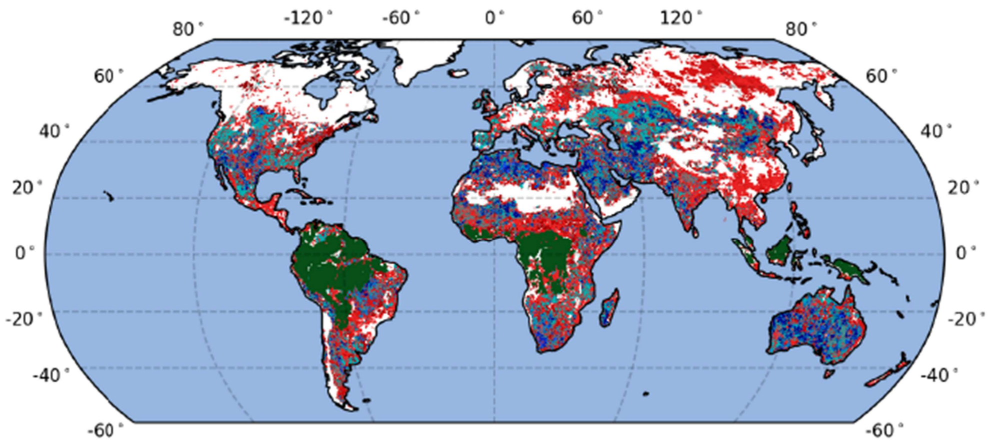


Figure 11: Results of the inhomogeneity testing (between HSP_3 and HSP_4) before any correction methods have been applied (top) with the results of the testing after the correction methods are applied (for each method as indicated) (bottom). Adapted from (Preimesberger et al., 2021)



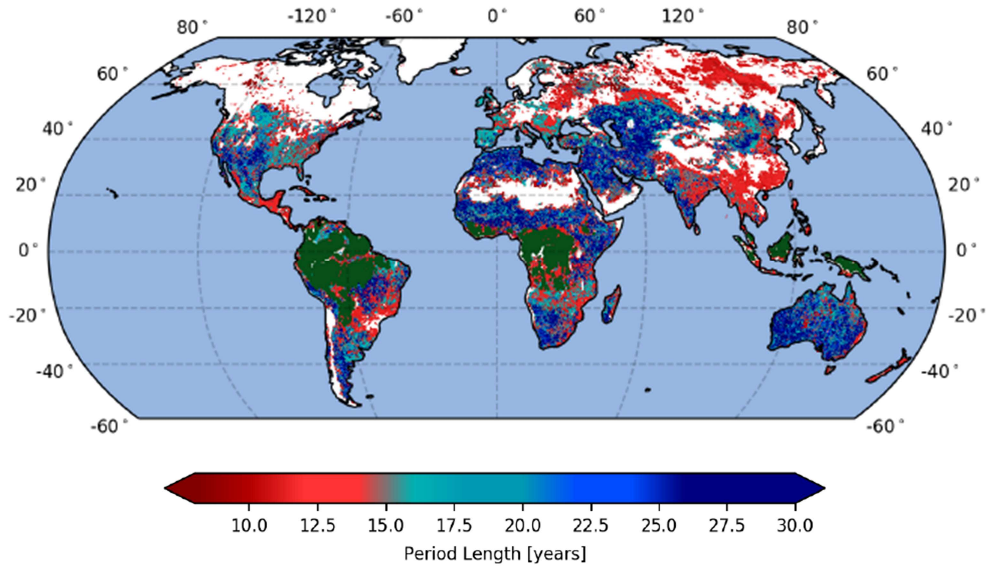


Figure 12: Longest homogenous period in ESA CCI SM v04.4 (COMBINED) before adjustment (top) and after adjustment (bottom) using the QCM method. Taken from (Preimesberger et al., 2021).

7.3 Quality Control and Validation

With the production of each new ESA CCI SM dataset, a series of quality assessments are undertaken to ensure the best quality product is being delivered to the data users. These assessments include spatial / temporal completeness assessments for all soil moisture and flag fields provided in the dataset, assessment of the soil moisture anomalies and analysis of trends with comparison against modelled data. These assessments are not only run on the final product, but with each incremental update to the product during the development stage.

In addition to these detailed quality controls, validation against in-situ and modelled datasets is undertaken to ensure the high level of accuracy of the product is maintained. This is undertaken by an independent organisation, ETH Zurich, and the results are provided in the PVIR (Hirschi et al., 2020). In addition, the newest datasets are provided on the QA4SM portal to allow users to run their own validations (https://qa4sm.eu/).

7.4 Known Limitations

The production of the ESA CCI SM product is a continuous process with new versions released each year. With each new major product version (i.e. v05 vs. v06), improvements are made to both the input datasets and the merging algorithm. This section discusses the current limitations of the ESA CCI SM.

7.4.1 Active Input Datasets

Metop-C was successfully launched on 7th November 2018 from French Guiana. Currently (v06), the ESA CCI SM product does not include observations from the ASCAT instrument on-board Metop-C. The use of these additional observations will allow improved characterisation of the daily ASCAT data used in ESA CCI SM dataset.

In addition to the use of ASCAT-C observations, other changes may be implemented in the H-SAF project which will benefit the ESA CCI SM products. For example, in future versions of the H-SAF ASCAT SSM CDRs, potential ways to correct sub-surface scattering effects will be tested. Morrison et al. (2019) found in an experimental setup that sub-surface scattering effects are caused by reflecting features in shallow soils under extremely dry conditions as e.g. in deserts. These anomalies can appear temporary or continuously in ASCAT observations, dominate the received backscatter signal and cause large fluctuations after rainfall events.

In addition, the use of bootstrapping and/or Monte Carlo methods to estimate the error budget of ASCAT SSM retrieval will be explored. Bootstrapping uses a large number of subsamples drawn from a set of measurements. This way it extrapolates to the population that the original measurements come from and allow to make estimates about the distribution of the errors based on the (resampled) measurements themselves.

7.4.2 Inter-calibration of ERS and ASCAT

The generation of ESA CCI SM ACTIVE product is based on the individual time series of ERS and ASCAT. The inter-calibration of ERS and ASCAT brightness temperatures (Level 1) would improve the quality of the individual measurements as well as the robustness of the calculation of the dry and wet references.

This could, for example, be undertaken using the method of Reimer (2014) who introduced a model-based inter-calibration methodology that accounts for temporal calibration biases within a specific scatterometer mission and subsequently considers temporal invariant inter-calibration biases between various scatterometer missions. This approach employs a number of natural calibration targets (rainforests) supposed to result in a more robust estimation of inter-calibration biases.

As the merging algorithm of ESA CCI SM starts from Level 2 products, such an undertaking is outside of the scope of the product generation, but ESA CCI SM would benefit from such inter-calibration work being undertaken for the L2 input products currently provided by H-SAF.

7.4.3 *Day time observations from passive sensors*

The ESA CCI Soil Moisture is currently based on night-time overpasses. This is because near surface land surface temperature gradients are regarded to be reduced at night leading to more robust retrievals (Owe et al., 2008). However, studies (Brocca et al., 2011) suggest that for specific land cover types day-time observations may provide more robust retrievals than night-time observations, although the exact causes are still unknown. It is expected that introducing daytime overpasses as well will result in a significant improvement in data coverage over the entire product period.

7.4.4 Improved Flagging

The integration of spurious Level 2 soil moisture retrievals into ESA CCI Soil Moisture may adversely affect its quality in different ways. While RFI may lead to increased random error, frozen soils, snow cover, and open water will have a more systematic impact on the soil moisture levels, e.g. leading to a bias.

7.4.5 *Decomposition into Climatologies and Anomalies*

Currently the SNR-based merging scheme applies a relative weighting of data sets based on their relative error characteristics. However, studies have shown that different frequency components may be subject to different error magnitudes (Brocca et al., 2011 , Su et al., 2015, Draper et al., 2015).

7.4.6 Independency from land surface models

Soil Moisture simulations from NASA's GLDAS Noah model (Rodell et al., 2004) are currently used as the scaling reference to harmonise L2 input data for the combined product prior to estimating uncertainties for merging (Gruber et al., 2019). This leads to the ESA CCI SM (COMBINED) observations remaining in the value domain of GLDAS Noah SM afterwards. Features in the satellite observations (e.g. impact of irrigation) are potentially attenuated in this process. Independence from model SM is therefore desired.

Harmonised L-band observations from SMAP and SMOS could be used to create an alternative scaling reference. The comparably short time periods of available L-band SM and effects such as radio frequency interference (RFI) in this frequency domain must be considered as they could negatively affect the creation of a scaling reference.

7.4.7 Uncertainty Analysis

Currently, "sm_uncertainty" field provided with the ESA CCI SM products is based on the TCA results only and does not include errors associated with any other processing steps, nor the input data.

7.4.8 *Spatial and Temporal Gap-Filling*

The number of sensors used to generate the soil moisture product varies temporally and spatially; this directly translates to the number of valid observations available in the end product which can adversely impact end-user applications.

8 References

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., & W., W. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment, 118*, 215-226

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

Bevington, P.R., & Robinson, D.K. (2002). *Data reduction and error analysis for the physical sciences*. (3th ed.). Boston: McGraw-Hill Science/Engineering/Math

Bevis, M., Businger, S., Herring, T., Rocken, C., Anthes, R., & Ware, R. (1992). GPS Meteorology: Remote Sensing of Atmospheric Water Vapor Using the Global Positioning System. *Journal of Geophysical Research, 97*, null-15801

BIPM (2008). *Evaluation of measurement data—guide for the expression of uncertainty in measurement. JCGM 100: 2008*.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., & Bittelli, M. (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study accross Europe. *Remote Sensing of Environment, 115*, 3390-3408

Chung, D., Kidd, R., Dorigo, W. (2013) ESA CCI Soil Moisture – System Prototype Description (SPD), Version 1.0, 4th July 2013.

Crapolicchio, R., A. Bigazzi, G. De Chiara, X. Neyt, A. Stoffelen, M. Belmonte, W. Wagner, C. Reimer (2016) The scatterometer instrument competence centre (SCIRoCCo): Project's activities and first achievements, Proceedings European Space Agency Living Planet Symposium 2016, 9-13 May 2016, Prague, Czech Republic, 9-13.

Das, N.N., and O'Neill, P. (2020). Soil Moisture Active Passive (SMAP) Ancillary Data Report; Soil Attributes. JPL D-53058, VERSION-B.

de Jeu, R., Wagner, W., Holmes, T., Dolman, H., van de Giesen, N.C., & Friesen, J. (2008). Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surveys in Geophysics, 29*, 399-420

de Jeu R.A.M, T.R.H. Holmes, R. M. Parinussa, M Owe (2014). A spatially coherent global soil moisture product with improved temporal resolution, Journal of Hydrology 516, 284-296.

de Nijs, A., Parinussa, R., de Jeu, R., Schellekens, J., Holmes, T. (2015) A Methodology to Determine Radio-Frequency Interference in AMSR2 Observations. *IEEE Transactions on Geoscience and Remote Sensing 53(9):5148-5159.* DOI: 10.1109/TGRS.2015.2417653

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., & Lecomte, P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*

Dorigo, W.A., Scipal, K., Parinussa, R.M., Liu, Y.Y., Wagner, W., de Jeu, R.A.M., & Naeimi, V. (2010). Error characterisation of global active and passive microwave soil moisture datasets. *Hydrology and Earth System Sciences, 14*, 2605-2616

Dobson, M.C., Ulaby, F.T., Hallikainen, M.T., & Elrayes, M.A. (1985). Microwave dielectric behavior of wet soil. Part II: Dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing, 23*, 35-46

Draper, C., & Reichle, R. (2015). The impact of near-surface soil moisture assimilation at subseasonal, seasonal, and inter-annual timescales. *Hydrology and Earth System Sciences, 19*, 4831-4844

Drusch, M., Wood, E., & Gao, H. (2005). Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. *Geophysical Research Letters, 32*, L15403

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellog, 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., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., & Van Zyl, J. (2010a). The Soil Moisture Active Passive (SMAP) mission. *Proceedings of the IEEE, 98*, 704-716

Entekhabi, D., Reichle, R.H., Koster, R.D., & Crow, W.T. (2010b). Performance metrices for soil moisture retrievals and application requirements. *Journal Of Hydrometeorology, 11*, 832-840

Entin, J.K., Robock, A., Vinnikov, K.Y., Hollinger, S.E., Liu, S., & Namkhai, A. (2000). Temporal and spatial scales of observed soil moisture variations in the extratropics. *Journal of Geophysical Research, 105*, 11865-11877

ESA (2017) ERS-2 SCATTEROMETER Surface Soil Moisture Time Series in High Resolution - ERS.SSM.H.TS (25 km Time-Series product), SCI-MAN-16-0047-v02 https://earth.esa.int/documents/700255/3799027/scirocco-pum-ts.pdf/cb893b39-b7db-441a-a821-9328436aa8f5

ESA (2018), ESA Climate Change Initiative Plus, Statement of Work, European Space Agency, ESA_CCI-EOPS-PRGM-SOW-18-0118, 25/05/2018.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture Climate Data Records and their underlying merging methodology. *Earth System Science Data 11, 717-739,* https://doi.org/10.5194/essd-11-717-2019.

Gruber, A., Dorigo, W., Crow, W., & Wagner, W. (2017). Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Transactions of Geoscience and Remote Sensing*

Gruhier, C., de Rosnay, P., Hasenauer, S., Holmes, T., de Jeu, R., Kerr, Y., Mougin, E., Njoku, E., Timouk, F., Wagner, W., & Zribi, M. (2010). Soil moisture active and passive microwave products: intercomparison and evaluation over a Sahelian site. *Hydrology and Earth System Sciences, 14*, 141-156

Gruber, A., Su, C.H., Zwieback, S., Crowd, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation, 45*, 200-211

H-SAF (2019a) ASCAT Surface Soil Moisture Climate Data Record v5 12.5 km sampling - Metop (H115), EUMETSAT SAF on Support to Operational Hydrology and Water Management, DOI: 10.15770/EUM_SAF_H_0006. http://dx.doi.org/10.15770/EUM_SAF_H_0006

H-SAF (2019b) ASCAT Surface Soil Moisture Climate Data Record v5 Extension 12.5 km sampling - Metop (H116), EUMETSAT SAF on Support to Operational Hydrology and Water Management. https://navigator.eumetsat.int/product/EO:EUM:DAT:METOP:H116

Hirschi, M., Nicolai-Shaw, N. , Preimesberger, W., Scanlon, T., Dorigo, W., Rodriguez Fernandez, N., Thevenon, H., Kidd, R. (2020), Product Validation and Intercomparison Report (PVIR) Supporting Product version v04.7 Deliverable ID: D4.1 Version 1, 29th May 2020.

Holmes, T. R. H., R. A. M. De Jeu, M. Owe, and A. J. Dolman. "Land surface temperature from Ka band (37 GHz) passive microwave observations." Journal of Geophysical Research: Atmospheres 114, no. D4 (2009).

Kerr, Y., 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., Martin-Neira, M., & Mecklenburg (2010). The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE, 98*, 666-687

Kidd, R., Chung, D., Dorigo, W., De Jeu, R. (2013). ESA CCI Soil Moisture – Detailed Processing Model (DPM), Version 1.2, 26th November 2013.

Koster, R.D., Guo, Z., Yang, R., Dirmeyer, P.A., Mitchell, K., & Puma, M.J. (2009). On the Nature of Soil Moisture in Land Surface Models. *Journal of Climate, 22*, 4322-4335

Legates, D.R., Mahmood, R., Levia, D.F., DeLiberty, T.L., Quiring, S.M., Houser, C., & Nelson, F.E. (2011). Soil moisture: A central and unifying theme in physical geography. *Progress in Physical Geography, 35*, 65-86

Li, L., Gaiser, P.W., Gao, B.-C., Bevilacqua, R.M., Jackson, T.J., Njoku, E.G., Rüdiger, C., Calvet, J.-C., & Bindlish, R. (2010). WindSat global soil moisture retrieval and validation. *IEEE Transaction on Geoscience and Remote Sensing, 48*, 2224-2241

Li, L., Njoku, E.G., Im, E., Chang, P.S., & Germain, K.S. (2004). A preliminary survey of radio-frequency interference over the US in Aqua AMSR-E data. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 380-390

Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., De Jeu, R.A.M., Wagner, W., McCabe, M.F., Evans, J.P., & Van Dijk, A.I.J.M. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment, 123*, 280-297

Liu, Y.Y., Parinussa, R.M., Dorigo, W.A., de Jeu, R.A.M., Wagner, W., van Dijk, A., McCabe, F.M., & Evans, J.P. (2011). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences, 15*, 425-436

Liu, Y., de Jeu, R.A.M., van Dijk, A.I.J.M., & Owe, M. (2007). TRMM-TMI satellite observed soil moisture and vegetation density (1998-2005) show strong connection with El Nino in eastern Australia. *Geophysical Research Letters, 34*, Art. No. L15401

Meesters, A., De Jeu, R.A.M., & Owe, M. (2005). Analytical derivation of the vegetation optical depth from the microwave polarisation difference index. *IEEE Geoscience and Remote Sensing Letters, 2*, 121-123

Mironov, V.L., Dobson, M.C., Kaupp, V.H., Komarov, S.A., & Kleshchenko, V.N. (2004). Generalized refractive mixing dielectric model for moist soils. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 773-785

Mo, T., Choudhury, B.J., Schmugge, T.J., Wang, J.R., & Jackson, T.J. (1982). A model for microwave emission from vegetation-covered fields. *Journal of Geophysical Research-Oceans and Atmospheres, 87*, 1229-1237

Moesinger, L. and Dorigo, W. and de Jeu, R. and van der Schalie, R. and Scanlon, T. and Teubner, I. and Forkel, M. (2020). The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA). Earth System Science Data, 12, 177-196

K. Morrison and W. Wagner, "Explaining Anomalies in SAR and Scatterometer Soil Moisture Retrievals From Dry Soils With Subsurface Scattering," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 3, pp. 2190-2197, March 2020, doi: 10.1109/TGRS.2019.2954771.

Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., & Nghiem, S.V. (2003). Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing, 41*, 215-229

Owe, M., de jeu, R., & Holmes, T. (2008). Multisensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research-Earth Surface, 113*, F01002

Owe, M., de Jeu, R., Walker, J. (2001) A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarisation difference index. *IEEE Transactions on Geoscience and Remote Sensing, 39(8), 1643-1654*

Parinussa, R., de Jeu, R., van der Schalie, R., Crow, W., Lei, F., & Holmes, T. (2016). A Quasi-Global Approach to Improve Day-Time Satellite Surface Soil Moisture Anomalies through the Land Surface Temperature Input. *Climate, 4*, 50

Parinussa, R.M., Meesters, A., Liu, Y.Y., Dorigo, W., Wagner, W., & de Jeu, R.A.M. (2011). Error Estimates for Near-Real-Time Satellite Soil Moisture as Derived From the Land Parameter Retrieval Model. *IEEE Geoscience and Remote Sensing Letters, 8*, 779-783

Peplinski, N.R., Ulaby, F.T., & Dobson, M.C. (1995). Dielectric properties of soils in the 0.3-1.3 GHz range. *IEEE Transactions on Geoscience and Remote Sensing, 33*, 803-807

Piles, M., Schalie, R.v.d., Gruber, A., Munoz-Mari, J., Camps-Valls, G., Mateo-Sanchis, A., Dorigo, W., & De Jeu, R. (2018). Global Estimation of Soil Moisture Persistence with L And C-Band Microwave Sensors. In, *IGARSS 2018*. Valencia: IEEE.

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, *in IEEE Transactions on Geoscience and Remote Sensing,* vol. 59, no. 4, pp. 2845-2862, April 2021, doi: 10.1109/TGRS.2020.3012896.

Reichle, R.H., Koster, R.D., Dong, J., & Berg, A.A. (2004). Global Soil Moisture from Satellite Observation, Land Surface Models, and Ground Data: Implications for Data Assimilation. *Journal Of Hydrometeorology, 5*, 430-442

Reimer, C. (2014). Calibration of space-borne Scatterometers: Towards a consistent climate data record for Soil Moisture Retrieval. (p. 118). Vienna: Vienna University of Technology.

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K.,Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D.,Toll, D. (2004). The global land data assimilation system. *Bull. Am. Meteor. Soc. 85,381–394.* https://doi.org/10.1175/BAMS-85-3-381.

Rüdiger, C., Holmes, T., Calvet, J.-C., de Jeu, R., & Wagner, W. (2009). An intercomparision of ERS-Scat and AMSR-E soil moisture observations with Model Simulations over France. *Journal Of Hydrometeorology, 10*, 431-447

Scanlon, T., Dorigo W., Preimesberger, W., Kidd, R., Van der Schalie, R., de Jeu, R., Rodrigues Fernandez, N., Thevenon, H., van der Vliet, M., User Requirements Document (URD) D1.1 Version 2.0, 19-11-2020.

Schanda, E. (1986). *Physical fundamentals of remote sensing*. Berlin Heidelberg New York Tokyo: Springer Verlag

Schmugge, T.J., O'Neill, P.E., & Wang, J.R. (1986). Passive microwave soil moisture research. *IEEE Transaction on Geoscience and Remote Sensing, GE-24*, 12-22

Scipal, K., Wagner, W., Trommler, M., & Naumann, K. (2002). The global soil moisture archive 1992-2000 from ERS scatterometer data: First results. Igarss 2002: Ieee International Geoscience and Remote Sensing Symposium and 24th Canadian Symposium on Remote Sensing, Vols I-Vi, Proceedings, 1399-1401

Su, C. H., Ryu, D., Dorigo, W., Zwieback, S., Gruber, A., Albergel, C., Reichle, R. H., Wagner, W. (2016). Homogeneity of a global multiscale soil moisture climate data record. *Geophysical Research Letters, 43(21), 11,245-11,252*

Su, C.H., & Ryu, D. (2015). Multi-scale analysis of bias correction of soil moisture. *Hydrology and Earth System Sciences, 19*, 17-31

Su, Z., Dorigo, W., Fernández-Prieto, D., Helvoirt, M.V., Hungershoefer, K., R. de Jeu, R.P., Timmermans, J., Roebeling, R., Schröder, M., Schulz, J., Tol, C.V.d., Stammes, P., Wagner, W., Wang, L., Wang, P., & Wolters, E. (2010). Earth observation Water Cycle Multi-Mission Observation Strategy (WACMOS). *Hydrology and Earth System Sciences Discussions, 7*, 7899-7956

TU Wien (2013). ERS AMI-WS (ESCAT) Surface Soil Moisture Product generated from E1/2-SZ-WNF/UWI-00 dataset. Department of Geodesy and Geoinformation, TU Wien, 2013.

Ulaby, F.T., Moore, B., & Fung, A.K. (1982). *Microwave Remote Sensing - Active and Passive, Vol. II: Radar Remote Sensing and Surface Scattering and Emission Theory.* Norwood: Artech House

Van der Schalie, R., Kerr, Y.H., Wigneron, J.P., Rodriguez-Fernandez, N.J., Al-Yaari, and De Jeu, R.A.M. (2015). Global SMOS Soil Moisture Retrievals from the Land Parameter Retrieval Model. *Int. J. Appl. Earth Observ. Geoinf.*, doi: http://dx.doi.org/10.1016/j.jag.2015.08.005.

Van der Schalie, R., De Jeu, R.A.M., Kerr, Y.H., Wigneron, J.-P., Rodriguez-Fernandez, N.J., Al-Yaari, A., Parinussa, R.M., Mecklenburg, S., and Drusch, M. (2017) The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E. *Remote Sens. Environ.*, 189 (2017), pp. 180-193. doi.org/10.1016/j.rse.2016.11.026

Van der Schalie, R., de Jeu, R., Parinussa, R., Rodríguez-Fernández, N., Kerr, Y., Al-Yaari, A., Wigneron, J.-P., & Drusch, M. (2018). The Effect of Three Different Data Fusion Approaches on the Quality of Soil Moisture Retrievals from Multiple Passive Microwave Sensors. *Remote Sensing, 10*, 107

Van der Vliet, Mendy, Robin van der Schalie, Nemesio Rodriguez-Fernandez, Andreas Colliander, Richard de Jeu, Wolfgang Preimesberger, Tracy Scanlon, and Wouter Dorigo (2020). Reconciling Flagging Strategies for Multi-Sensor Satellite Soil Moisture Climate Data Records. *Remote Sensing* 12, no. 20: 3439.

Wagner, W., Naeimi, V., Scipal, K., de Jeu, R., & Martinez-Fernandez, J. (2007). Soil moisture from operational meteorological satellites. *Hydrogeology Journal, 15*, 121-131

Wagner, W., Scipal, K., Pathe, C., Gerten, D., Lucht, W., & Rudolf, B. (2003). Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *Journal of Geophysical Research-Atmospheres, 108*

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sensing of Environment, 70*, 191-207

Wang, J.R., & Choudhury, B.J. (1981). Remote sensing of soil moisture content over bare field at 1.4 GHz frequency. *Journal of Geophysical Research-Oceans and Atmospheres, 86*, 5277-5282

Wang, J.R., & Schmugge, T.J. (1980). An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing, 18*, 288-295