ESA Climate Change Initiative Plus - Soil Moisture

Algorithm Theoretical Baseline Document (ATBD) Supporting Product Version v05.2

D2.1 Version 1

29-05-2020

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 http://www.esa-cci.org/.

Number of pages: 71 (inc. cover and preface)

Authors:		A. Pasik, T. Scanlon, W. Dorigo, R.A.M de Jeu, S. Hahn, R. van der Schalie, W. Wagner, R. Kidd, A. Gruber, L. Moesinger, W. Preimesberger	
Circulation:		Public Release **(after ESA review)**	
Issue	Date	Details	Editor
0.1	21/11/2019	Based upon CCI SM Phase 2: ATBD D2.1 Version 04.4 28/11/2018. Updated to product version 04.5 (data extension, public release). ATBD documents previously maintained separately for each of the datasets merged into a single document. Removal of "Active" sub-document (Active development is undertaken and reported in Eumetsat H-SAF Project). (previous doc ID Release 4.5.1 22/11/2019)	A. Pasik
0.2	09/12/2019	Review + corrections (previous doc ID Release 4.5.4 09/12/2019)	W. Dorigo, S.Hahn, L. Moesinger, W. Preimesberger, A. Pasik
0.3	20/02/2020	Updated to product version 04.7 (implementation of LPRMv6 for all passive sensors, temporal extension to 31-12-2019, public release). (previous doc ID Release 4.7.1 20/02/2020)	A. Pasik, W. Preimesberger, L. Moesinger
0.4	05/03/2020	Updated to product version 05.2 (Improved CDF matching and inclusion of SMAP data). (previous doc ID Release 5.2.1 05/03/2020)	A. Pasik, W. Preimesberger, L. Moesinger
0.5	30/03/2020	Revised Change Log (section 2) and Table 1 to v05.2. Revised figures and tables to include SMAP where applicable. Removed section 7.2.3 on future SMAP integration as this has been now implemented. Updated the AMSR2 scaling regime to match final v05.2 passive product version. Fixed page numbering. Introduced section 8.5.2.1 describing de-trended ASCAT dataset investigation. (previous doc ID Release 5.2.2 30/03/2020)	A. Pasik, L. Moesinger, W. Preimesberger
0.6	24/04/2020	Document Review. Revised Title Page, Deliverable ID and Document Issue ID. Revised Change Log to report changes from last approved ATBD. Updated header to	R. Kidd (EODC)

			provide Product Version, Document version. Update of all fields – resolved any cross-reference errors. Revised Header for change table in Section 2. To ESA for review.	
0.7	04/05/2020		RID: ESA_CCI_SM_RD_D2.1_v1_ATBD_v05.2_issue_0.6_RID. Addressed RID's 001, 002, 003. To TU Wien for completion	R. Kidd (EODC)
0.8	27/05/2020		[RID 004] – changed reference, [RID 005] revised Figure 1 to show only missions used in CCI. [RID 006] revised text to reflect "up to 1m". [RID 007] linked section 6 to Figure 1. [RID 008] revised text to reflect update of CCI based on update of H-SAF products, [RID 009] added reference to Figure 3 in section 7.1.6.[RID-010] document restructured in section 7.2 and 7.2.1. [RID-012] revised Figure 11 text to include depth of considered layer. Responded on all editorial corrections. For response on discussion on [RID-011] – please see RID response on CDF matching.	A. Pasik (TU Wien) RK (EODC)
1.0	29/05/2020		Accepted all reviewers' comments. Closed Document. To Pdf for delivery	RK

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)

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

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 4 and 5 provide a brief overview of the problem and of the ESA CCI SM production system respectively. Section 6 contains a brief description of soil moisture products from active microwave sensors used in in the ESA CCI Soil Moisture and points to the organizations responsible for their retrieval. Section 7 describes succinctly the VUA-NASA Land Parameter Retrieval Method (LPRM) for estimating soil moisture from passive microwave sensors, and section 8 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) 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 v05.2

This document forms deliverable 2.1 of CCI+ and provides an update for the ESA CCI SM v05.2 product expected to be publicly released in the Q3/2020. Changes between version v05.2 algorithm and the previously used version 04.4 algorithm (temporally extended at v04.5 and v04.7 without algorithmic changes) involve improved CDF-matching, inter-sensor scaling regime of AMSR2 and the inclusion of SMAP data from April 2015.

Version v05.2 provides data from 1978 (PASSIVE and COMBINED products) and 1991 (ACTIVE product) to the end of December 2019.

2.1.1   ATBD Document

- Updated for version v05.2. Tables and figures revised to include SMAP where applicable. SMAP and AMSR2 scaling regimes described (see section 8.3.2) and processing steps visualized (Figure 8).

2.2 Pre v05.2

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

Dataset Version	Release date	Public	Key Users	Project Partners	Major Changes Since Previous Versions	ATBD Version
v05.2	Q3/2020 (expected)	X			Inclusion of SMAP data from April 2015, improved CDF-matching and updated inter-sensor scaling regime of AMSR2.	5.2
v04.7	12/03/2020	X			No algorithm changes since v04.4. Temporal extension to 2019-12-31.	4.7
v04.5	2019-09-30	x			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	X			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
v04.3	2018-04-17			X	No algorithm changes since v04.1. Temporal extension to 2017-12-31. Not released, but used for State of the Climate BAMS report 2018.	4.3
v04.2	2018-01-12	X			No algorithm changes since v04.1.	4.2
v04.1	2017-08-02		X	X	Masking of unreliable retrievals is undertaken prior to merging.	-
v04.0	2017-03-20			X	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	-

Dataset Version	Release date	Public	Key Users	Project Partners	Major Changes Since Previous Versions	ATBD Version
					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.	
v03.3	2017-11-13	X			Temporal extension of ACTIVE, PASSIVE and COMBINED datasets to 2016-12-31.	3.3
v03.2	2017-02-14	X			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.	3.2
v03.1	2016-11-02		X	X	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.	-
v03.0	2016-04-25			X	Introduction of new weighted-average based merging scheme. Miras SMOS (LPRM) now integrated into the data products. Blending weights provided as ancillary data files.	-
v02.3	2016-02-08	X			Temporal extension to 2015-12-31. Valid_range in netCDF files now set to the packed data range.	-

Dataset Version	Release date	Public	Key Users	Project Partners	Major Changes Since Previous Versions	ATBD Version
v02.2	2015-12-17	X			No changes.	-
	2015-08-06		X	X	In ancillary files latitudes now goes from positive to negative values.	-
	2015-07-31		X	X	Email address added to metadata.	-
	2015-03-17			X	Temporal coverage extended (Nov-1978 to Dec-2014). Improvement in the flagging of the active data where extreme high and low values are filtered.	-
v02.1	2014-12-03			X	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	X			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.	-
v01.2	2014-03-03			X	All datasets updated to include days where no observations are available.	-
v01.1	2014-02-19			X	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	

Dataset Version	Release date	Public	Key Users	Project Partners	Major Changes Since Previous Versions	ATBD Version
					function. Improved rescaling algorithm. Data gaps in 2003-02-16 to 2006-12-31 filled with AMSR-E data.	
v 0.1	2012-06-18			X	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.	-

3   Reference Documents

The following references are of relevance to this document. Within the document, for the sake of clarity it has been sometimes necessary to provide sections of quoted texts taken from referenced documents, rather than just providing a reference to the document. In these cases, texts are "*presented in quotes as italic text*".

[RD-1]	ESA Climate Change Initiative Plus, Statement of Work, European Space Agency, ESA_CCI-EOPS-PRGM-SOW-18-0118.
[RD-2]	Technical Proposal (Part 3) in response to ESA Climate Change Initiative Plus AO/1-9322.18/I-NB , Vienna University of Technology.
[RD-3]	W. Dorigo, R. Kidd, R. De Jeu, S. Seneviratne, H. Mittelbach, J. Pulliainen, W.A. Lahoz, N. Dwyer, B. Barrett, Eva Haas, W. Wagner. ESA CCI Soil Moisture Data Access Requirements Document, v1.2
[RD-4]	Wagner, W., W. Dorigo, R. De Jeu, D. Fernandez, J. Benveniste, E. Haas, M. Ertl (2012) Fusion of active and passive microwave observations to create an Essential Climate Variable data record on soil moisture. Proceedings of the ISPRS Congress 2012, Melbourne, Australia, August 25-September 1, 2012.
[RD-5]	Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe, M. F., 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, doi:10.5194/hess-15-425-2011
[RD-6]	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, doi: 10.1016/j.rse.2012.03.014.
[RD-7]	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.
[RD-8]	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.

[RD-9]	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. 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
[RD-10]	R. Kidd, D. Chung, W. Dorigo, R. De Jeu (2013). ESA CCI Soil Moisture – Detailed Processing Model (DPM), Version 1.2, 26th November 2013.
[RD-11]	D. Chung R. Kidd, W. Dorigo (2013) ESA CCI Soil Moisture – System Prototype Description (SPD), Version 1.0, 4th July 2013.
[RD-12]	R. Kidd, D. Chung, W. Dorigo, R. De Jeu (2012). Input/ Output Data Definition Document (IODD), Version 1.0, 21st December 2012.

4   Scope of ESA CCI Soil Moisture

This Section is partially based on [RD-4].

4.1 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 for many geoscientific applications essential 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 15 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 201254 (Albergel et al. 2012). Data from NASA's Soil Moisture Active Passive (SMAP) have become 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.

4.2 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 any microwave technique, particularly in the low-frequency microwave region from 1 to 10 GHz, offers the opportunity to measure soil moisture in a relatively direct manner. Therefore, the CCI soil moisture project focuses at this stage on space-borne microwave systems operating at low-frequency bands. 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 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.
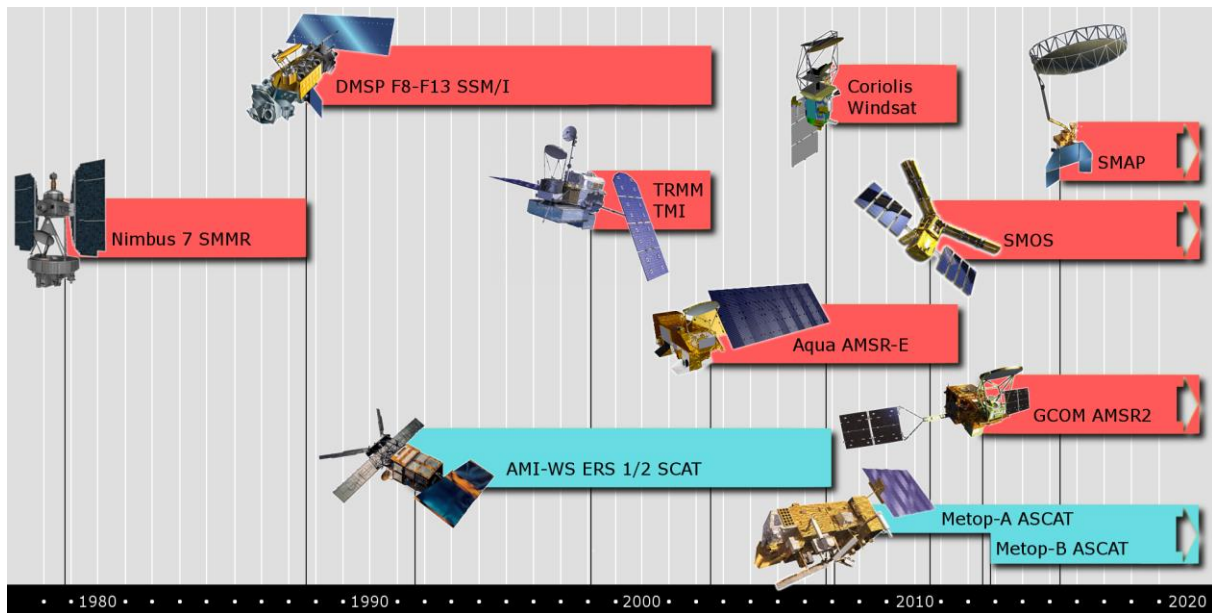
Figure 1: Active and passive microwave sensors used for the generation of the ESA CCI soil moisture data sets.

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 CCI Soil Moisture project thus aims to combine C-band scatterometers (e.g. ERS-1/2 scatterometer, METOP Advanced Scatterometers) and multi-frequency radiometers (e.g., SMMR, SSM/I, TMI, AMSR-E, Windsat, AMSR2, SMOS, SMAP) as these sensors are characterised by their high suitability for soil moisture retrieval and a long technological heritage (Figure 1). As specified in [RD-1], other microwave sensors suitable for soil moisture retrieval, including Synthetic Aperture Radars (SARs) and radar altimeters, are not considered in this phase of the CCI programme due to their recentness and/or their unfavourable spatio-temporal coverage. Nevertheless, the ESA CCI SM production system has been set up in such a way as to allow the integration of all these sensors in the future. A complete list and a detailed technical description of all data products used in the ESA CCI SM production system is provided in [RD-10 and RD-11].

4.3 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). Nevertheless, based on the requirements as specified in [RD-1], and drawing from the experiences of the use of the currently available satellite soil moisture data sets, a number of baseline requirements can be specified already at this stage.

4.3.1 Scientific Requirements

Thanks to the fact that several decade-long soil moisture data records have been released within the last few years 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:

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.

4.3.2 System Requirements

The generation of an ESA CCI SM data set is not a one-off activity, but should in fact be a long-term process where the ESA CCI SM product shall be continued and improved step by step with the active involvement of a broad scientific community. From a system point of view this requires that the ESA CCI SM Production System is modular 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, i.e. it allows adaptations to different data processing framework solutions

But not only modularity is a major requirement. The design and operations of the system should also be as lightweight as possible in order to be able 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 [RD-10] for further details on the soil moisture ESA CCI SM production system, detailing its components, their functions, and interfaces.

5 ESA CCI SM Production Approach

This Section is partly based on [RD-4].

5.1 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 one-can-do-it-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. It can e.g. be noted 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.

5.2 Fusion of Level 2 Soil Moisture Retrievals

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 the following steps, based on available level 2 products (Figure 2):

1. Fusion of the active Level 2 data sets
2. Fusion of the passive Level 2 data sets
3. Fusion of the merged active and passive data sets from steps 2 and 3

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), and 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 enrich the ESA CCI SM data set with Level 2 data from other existing and forthcoming sensors (e.g. SMAP, radar altimeters, Aquarius).
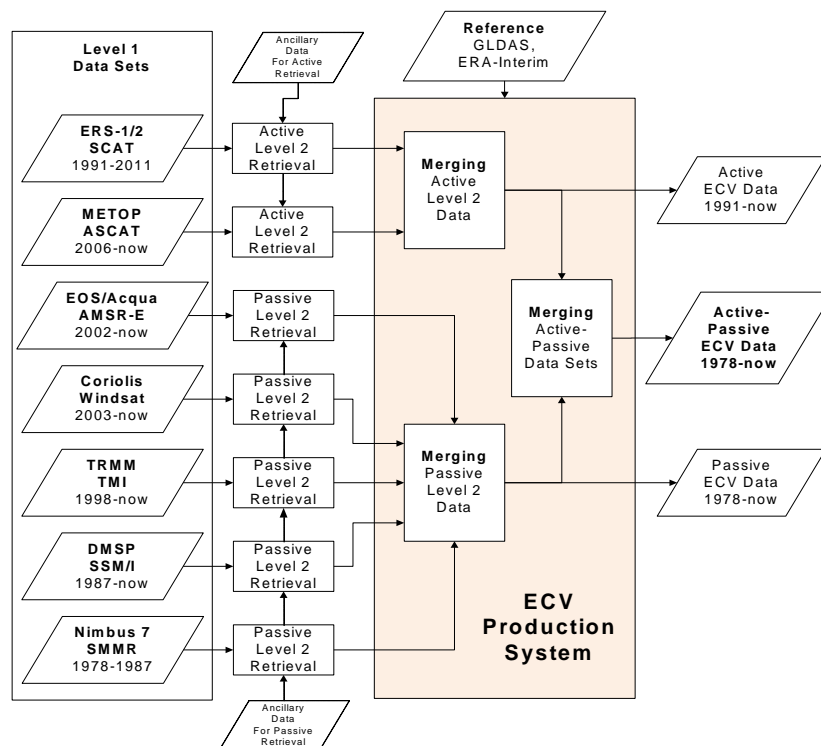
Figure 2: Flow chart of the ECV Production System as first proposed in the ESA funded WACMOS project (Liu et al. 2011; Su et al. 2010).

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 has to deal with

the fusion process only, as described above. This design is modular and lightweight, meeting the requirements as discussed in Section 4.3.2.

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 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. As a result, the combined active (scatterometer) and passive (multi-frequency radiometer) data sets may not always be directly comparable. Therefore, as illustrated in Figure 2 the ESA CCI SM Production System delivers, besides the fused and thus most complete active + passive (COMBINED) ESA CCI SM data set, the two active-only (ACTIVE) and passive-only (PASSIVE) ESA CCI SM data sets. It will be thus up to the user to decide, which of these merged soil moisture data sets is best suited for his or hers analyses.

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 8.

6 Description of soil moisture products from active microwave sensors used in the ESA CCI Soil Moisture

Active microwave soil moisture products (see Figure 1 for details) utilized in the generation of the CCI 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 v05.2 uses both the H SAF H115 Metop ASCAT SSM CDR v5 (H SAF, 2019a) and the H SAF H116 Metop ASCAT SSM CDR v5-Extension (H SAF, 2019b). Each version of the ESA CCI SM dataset uses the most recent and updated Metop ASCAT CDR made available by H SAF.

7 Methodological description on the retrieval of soil moisture from passive microwave 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).

7.1 Principles of the Land Parameter Retrieval Model

Brightness temperatures can be derived from several passive microwave sensors with different radiometric characteristics, i.e. Nimbus SMMR, the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) on-board the Soil Moisture and Ocean Salinity (SMOS) mission and the Advanced Microwave Scanning Radiometer (AMSR-E) on the AQUA Earth observation satellite and the radiometer instrument aboard Soil Moisture Active Passive (SMAP) [RD-5]. The observed brightness temperatures are converted to soil moisture values with the Land Parameter Retrieval Model (LPRM; van der Schalie et al. 2015).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 density in

combination with soil moisture and surface temperature. A result of this physical parameterization 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 [RD-7]. This important aspect makes LPRM suitable for the development of a long-term consistent soil moisture network within ESA's CCI soil moisture project.

The different processing steps of LPRM are described in detail in the next section, while Figure 3 presents a flowchart of the entire methodology.
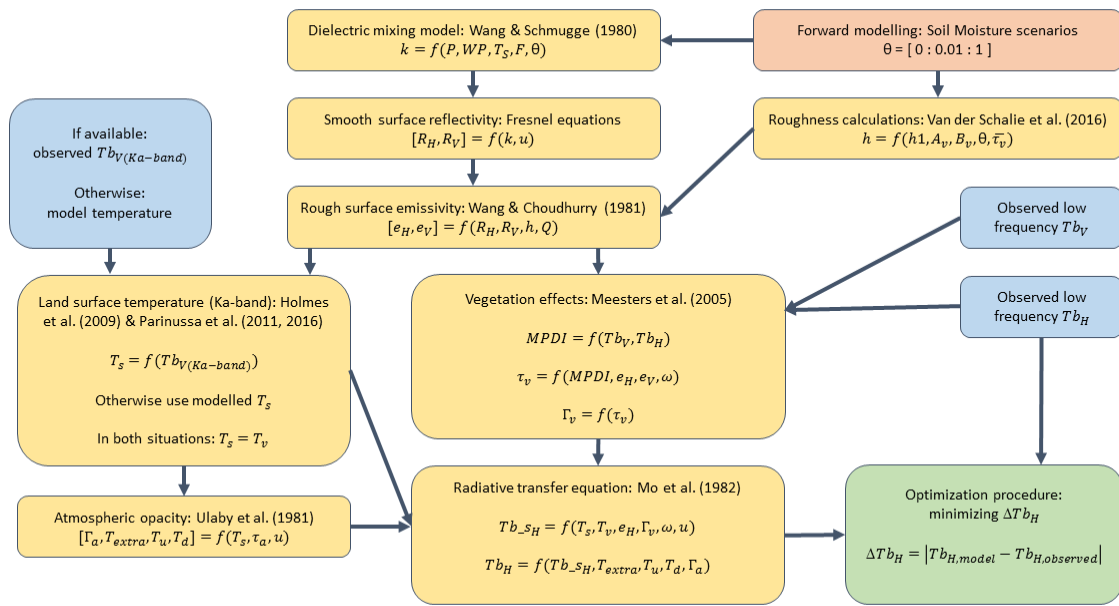
Figure 3: Flowchart of the main processes of the Land Parameter Retrieval Model (LPRM). Soil moisture is solved when the observed brightness temperature equals the modelled brightness temperature as derived by the radiative transfer.

7.1.1   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:

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

Eqn. 7-1

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 extraterrestrial brightness temperature and the 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) polarization.

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. 7-2

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. 7-3

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 Extraterrestrial 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. 7-4

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. 7-5

Where *Q* is the polarization mixing factor and *h* the roughness height. *h* is calculated using the related parameters h_1, A_v and B_v, see Eqn. 4-6, to take into account the effects of soil moisture (θ, m³ m⁻³) and vegetation cover ([RD-8], [RD-9]) on the *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. 7-6

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

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

$$r_{s,V} = \left| \frac{\varepsilon \cos u - \sqrt{\varepsilon - \sin^2 u}}{\varepsilon \cos u + \sqrt{\varepsilon - \sin^2 u}} \right|^2 \qquad \text{Eqn. 7-8}$$

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 to solve for the vegetation optical depth, 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 Polarization Difference Index (MPDI) to calculate τ_v, The MPDI is defined as:

$$MPDI = \frac{T_{b\_s,V} - T_{b\_s,H}}{T_{b\_s,V} + T_{b\_s,H}} \qquad \text{Eqn. 7-9}$$

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

$$\tau_v = \cos u \ln(ad + \sqrt{(ad)^2 + a + 1}) \qquad \text{Eqn. 7-10}$$

Where

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

Eqn. 7-6

And

$$d = \frac{1}{2}\frac{\omega}{(1-\omega)}$$

Eqn. 7-7

By using all these equations in combination with the dielectric mixing model, soil moisture could be solved in a forward model together with a parameterization 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:

$$T_s = 0.898T_{b\_37V} + 44.2$$

Eqn. 7-8

and for the night time (descending):

$$T_s = 0.893T_{b\_37V} + 44.8$$

Eqn. 7-9

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 FAO soil texture map (FAO, 2000), while All the other parameters were given a fixed value. Table 2 summarizes the values used for the different frequencies.

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

7.1.2 Soil Moisture Uncertainties

An uncertainty analysis for soil moisture retrievals as derived from passive microwave observations according to the Land Parameter Retrieval Model (LPRM; Owe et al. 2008, van der Schalie et al. 2015) was presented in Parinussa et al. (2011). Their methodology was based on standard error propagation, as can be found in general statistical text books (Bevington and Robinson 2002), and provides information about how the uncertainty in each of the input parameters propagate to the soil moisture output.

$$\Sigma y = U \Sigma x U^I$$

Eqn. 7-10

where U is the partial derivative matrix. When the errors and the internal correlations between the input parameters are known, the accuracy of soil moisture can be calculated.

LPRM is a zero order radiative transfer based model. Several input parameters are affected by instrumentation uncertainties, data acquisition, reduction limitations, methodology and environmental factors. Each of these errors will introduce an uncertainty in the final soil moisture product as derived from LPRM. In general, we are not able to determine the actual error in the result if no considered true data are available for evaluation of the experimental model. Therefore, we need to develop a consistent error model for uncertainty determination. This error model informs us about the random errors, but not the biases and it does not tell us whether the model itself is correct or wrong. So the main task of an uncertainty analyses is to quantify the random error in the output of a model under the assumption that the model itself is physically correct.

7.1.2.1 Analytical Derivation

Because of the high computational costs of statistical methods (e.g. Monte Carlo simulations), it's not feasible to apply such techniques on a global and (sub-) daily scale. A possible solution is proposed in the following part, where the radiative transfer equation was rewritten and an analytical solution of the quantitative uncertainty for passive microwave remote sensing of soil moisture product was derived.

The basis of the analytical solution to calculate the error in the soil moisture product lies in the use of the most basic error propagation methodology presented in most statistical textbooks; for example the function $x = f(u,v,...)$.

$$\sigma_x^2 \cong \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + ... + 2r_{uv}\sigma_u\sigma_v\left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + ...$$

Eqn. 7-11

The methodology is adapted here to determine the variance σ_k^2 of the dielectric constant (k), using the variances of several input parameters. After the determination of the variance in the dielectric constant, a dielectric mixing model (Wang and Schmugge 1980) was used to calculate the uncertainty in soil moisture. The challenge in using the basic error propagation methodology is to define the partial derivatives.

To define the partial derivatives, we used the Jacobian matrix. The Jacobian matrix is a matrix containing the first order partial derivatives of the radiative transfer equation with respect to each variable. In our case the Jacobian matrix (J) can be described as

$$J = \begin{pmatrix} \dfrac{\partial T_{bH}}{\partial \Gamma} & \dfrac{\partial T_{bH}}{\partial k} & \dfrac{\partial T_{bH}}{\partial T_{LS}} & \dfrac{\partial T_{bH}}{\partial \omega} & \dfrac{\partial T_{bH}}{\partial h} \\ \dfrac{\partial T_{bV}}{\partial \Gamma} & \dfrac{\partial T_{bV}}{\partial k} & \dfrac{\partial T_{bV}}{\partial T_{LS}} & \dfrac{\partial T_{bV}}{\partial \omega} & \dfrac{\partial T_{bV}}{\partial h} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Eqn. 7-12

After applying the land surface temperature assumption, one is able to rewrite the radiative transfer equation (Parinussa et al. 2011) after putting T_{LS} outside brackets, for convenience we drop subscript 'P' for polarization.

$$T_b = T_{LS}\left[e_r\Gamma + (1-\omega)(1-\Gamma) + (1-e_r)(1-\omega)(1-\Gamma)\Gamma\right]$$

Eqn. 7-13

This can be rewritten to

$$T_b = T_{LS}\left[e_r(\Gamma - (1-\omega)(1-\Gamma)\Gamma) + (1-\omega)(1-\Gamma^2)\right]$$

Eqn. 7-14

For convenience we define the expressions $F(\Gamma,\omega)$ Eqn. 7-15 and $G(\Gamma,\omega)$ Eqn. 7-16 to rewrite equation Eqn. 7-12, resulting in Eqn. 7-13

$$F(\Gamma,\omega) = \Gamma - (1-\omega)(1-\Gamma)\Gamma$$

Eqn. 7-15

$$G(\Gamma,\omega) = (1-\omega)(1-\Gamma^2)$$

Eqn. 7-16

$$T_b = T_{LS}\left[F(\Gamma,\omega)e_r(k,h) + G(\Gamma,\omega)\right]$$

Eqn. 7-17

The rough surface emissivity $e_{r(P)}$ follows from Eqn. 7-17, wherein the horizontal (H) and vertical (V) polarization are reintroduced. This equation was written to calculate the rough surface emissivity in horizontal polarization. To calculate the rough surface emissivity at vertical polarization the (H) and (V) sign for polarization should be swapped. Q is the roughness parameter known as the cross polarization, h is the roughness and k refers to the dielectric constant.

$$e_{r(H)}(k,h) = 1 - \left[Q(1 - e_{s(V)}(k)) + (1-Q)(1 - e_{s(H)}(k)) \right] \chi(h)$$

Eqn. 7-18

where the last term refers to

$$\chi(h) = \exp(-h \cdot \cos(u))$$

Eqn. 7-19

The smooth surface emissivity was calculated using equation 7.22 and 7.23, for convenience we drop subscript 's' from smooth emissivity.

$$e_H = 1 - \left(\frac{\cos(u) - \Delta}{\cos(u) + \Delta} \right)^2$$

Eqn. 7-20

$$e_V = 1 - \left(\frac{k \cdot \cos(u) - \Delta}{k \cdot \cos(u) + \Delta} \right)^2$$

Eqn. 7-21

where the Δ term refers to

$$\Delta = \sqrt{k - \sin^2 u}$$

Eqn. 7-22

The following derivatives will be needed

$$\frac{\partial F}{\partial \Gamma} = 1 - (1 - \omega)(1 - 2\Gamma)$$

Eqn. 7-23

$$\frac{\partial G}{\partial \Gamma} = -2(1 - \omega)\Gamma$$

Eqn. 7-24

$$\frac{\partial e_H}{\partial k} = \frac{2\cos(u)}{\Delta} \frac{\cos(u) - \Delta}{(\cos(u) + \Delta)^3}$$

Eqn. 7-25

$$\frac{\partial e_V}{\partial k} = 2\cos(u) \left(\frac{k}{\Delta} - 2\Delta \right) \frac{k \cdot \cos(u) - \Delta}{(k \cdot \cos(u) + \Delta)^3}$$

Eqn. 7-26

From these derivations it follows that the Jacobian matrix Eqn. 7-12 can be calculated analytically

$$J_{11} = T_{LS}\left(\frac{\partial F}{\partial \Gamma}e_{r,H} + \frac{\partial G}{\partial \Gamma}\right)$$

Eqn. 7-27

$$J_{21} = T_{LS}\left(\frac{\partial F}{\partial \Gamma}e_{r,V} + \frac{\partial G}{\partial \Gamma}\right)$$

Eqn. 7-28

$$J_{12} = T_{LS}F\left(Q\frac{\partial e_V}{\partial k} + (1-Q)\frac{\partial e_H}{\partial k}\right)\chi$$

Eqn. 7-29

$$J_{22} = T_{LS}F\left(Q\frac{\partial e_H}{\partial k} + (1-Q)\frac{\partial e_V}{\partial k}\right)\chi$$

Eqn. 7-30

$$J_{13} = Fe_{r,H} + G$$

Eqn. 7-31

$$J_{23} = Fe_{r,V} + G$$

Eqn. 7-32

$$J_{14} = T_{LS}\left[(1-\Gamma)\Gamma e_{r,H} - 1 + \Gamma^2\right]$$

Eqn. 7-33

$$J_{24} = T_{LS}\left[(1-\Gamma)\Gamma e_{r,V} - 1 + \Gamma^2\right]$$

Eqn. 7-34

$$J_{15} = T_{LS}F\left[Q(1-e_V) + (1-Q)(1-e_H)\right]\chi\cos(u)$$

Eqn. 7-35

$$J_{25} = T_{LS}F\left[Q(1-e_H) + (1-Q)(1-e_V)\right]\chi\cos(u)$$

Eqn. 7-36

From LPRM, it follows that variations in the observed parameters $T_{bH}, T_{bV}, T_{LS(obs)}$, ω^{est} and h_{est} are related to variations in the unknown model parameters Γ, k, T_{LS} , ω and h. Combining this with the inverse Jacobian matrix results in the following expression:

$$
\begin{pmatrix}
\delta\Gamma \\
\delta k \\
\delta T_{LS} \\
\delta\omega \\
\delta h
\end{pmatrix}
= J^{-1}
\begin{pmatrix}
\delta T_{bH} \\
\delta T_{bV} \\
\delta T_{LS(obs)} \\
\delta\omega_{est} \\
\delta h_{est}
\end{pmatrix}
$$

Eqn. 7-37

The second line in this equation holds the result:

$$
\sigma_k^2 = ((J^{-1})_{21})^2 \sigma_{TbH}^2 + ((J^{-1})_{22})^2 \sigma_{TbV}^2 +
$$
$$
2((J^{-1})_{21})((J^{-1})_{22})r\sigma_{TbH}\sigma_{TbV}
$$
$$
+ ((J^{-1})_{23})^2 \sigma_{LS(obs)}^2 + ((J^{-1})_{24})^2 \sigma_\omega^2 + ((J^{-1})_{25})^2 \sigma_h^2
$$

Eqn. 7-38

Herein, the correlation between the errors in T_{bH} and T_{bV} is expressed in r.

Figure 4 presents the global average error for AMSR-E C-band observation over 2008 resulting from the analytical error propagation analysis. It clearly shows standard deviation values below 0.06 m^{3}m^{-3} for all the dry and semi-arid regions and higher value up to 0.1 m^3 m^{-3} and beyond for the more densely vegetated regions.
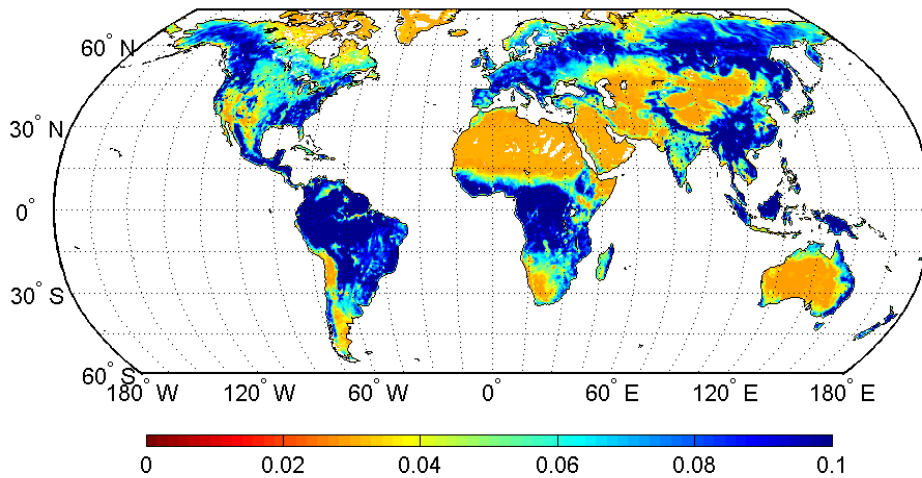
Figure 4: Average estimated standard deviation of AMSR-E C-band soil moisture for 2008 as derived from the analytical error propagation analysis proposed by Parinussa et al., (2011).

7.1.3   Known Limitations

The known limitations in deriving soil moisture from passive microwave observations are listed and described in detail in this section. These issues do not only apply to the current CCI soil moisture dataset release (v05.2) but also to soil moisture retrievals from passive microwave observations in general.

7.1.4 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 vegetation optical depth can be used to detect areas with excessive vegetation, of which the boundary varies with observation frequency. Figure 5 gives an example of the relationship between the analytical error estimate in soil moisture as described in the previous section and vegetation optical depth. 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. Soil Moisture retrievals with a soil moisture error estimates beyond 0.2 $m^3 m^{-3}$ are considered to be unreliable and are masked out
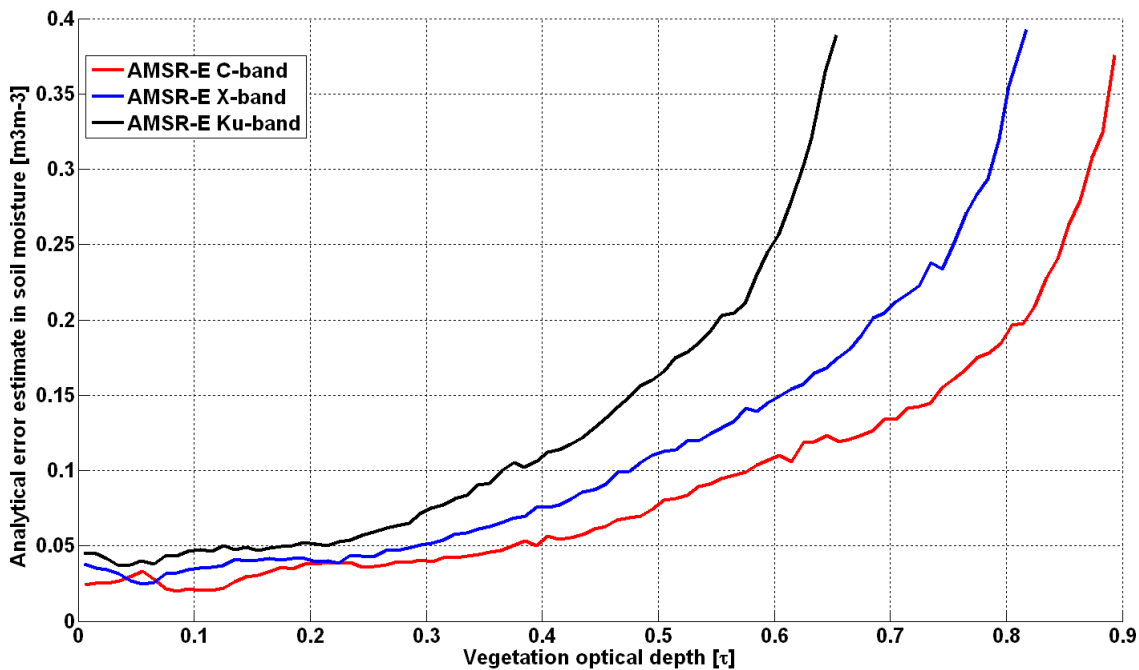
Figure 5: Error of soil moisture as related to the vegetation optical depth for 3 different frequency bands (from Parinussa et al., 2011).

For the new 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 6. In Figure 6, 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. A complete analysis on the error for L-band soil moisture, comparable to the results from Figure 5, are planned in the near future.
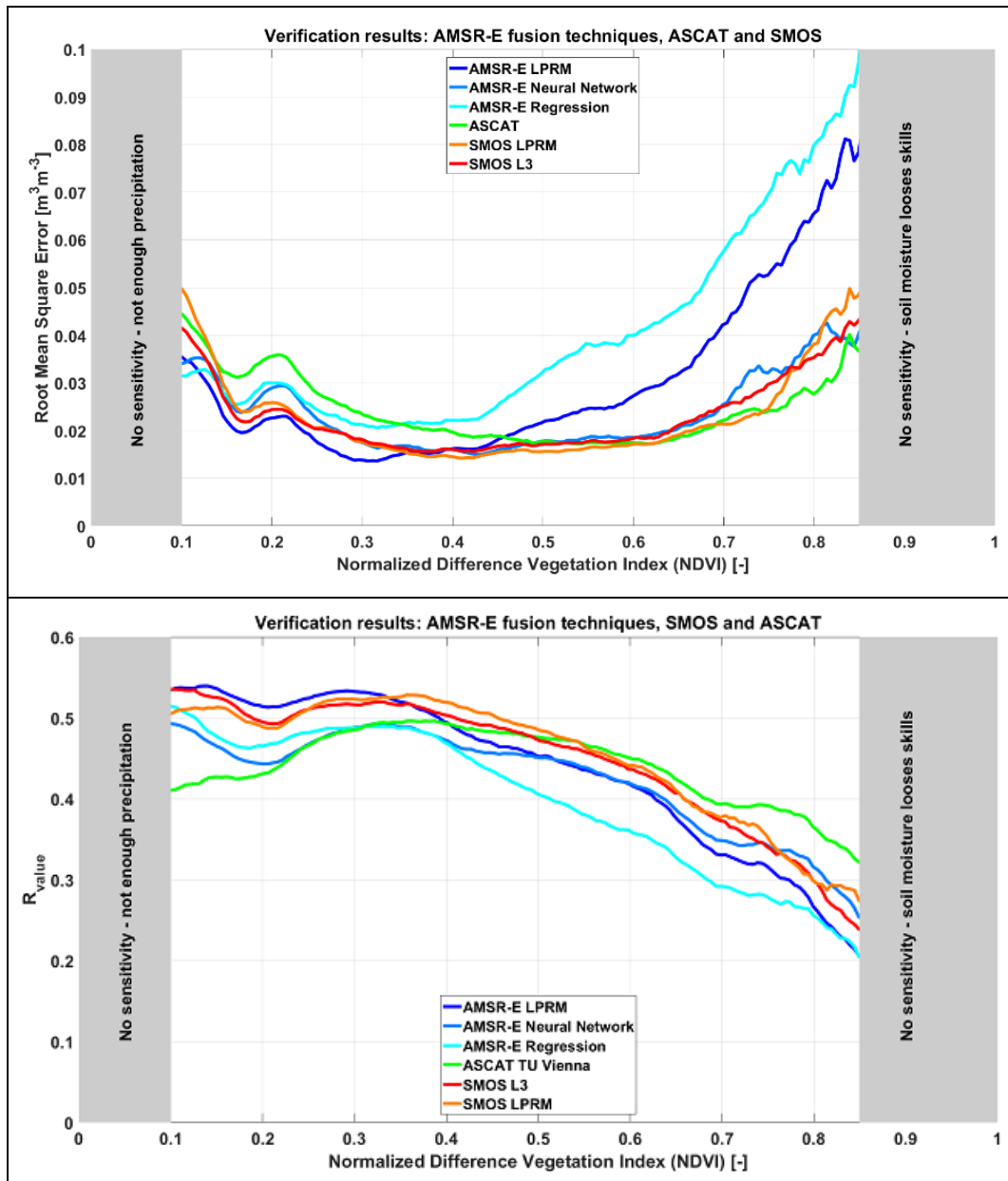
Figure 6: Triple collocation analysis (TCA: top) and R_{value} results (bottom) for several soil moisture datasets, including SMOS LPRM and AMSR-E LPRM, for changing vegetation density (NDVI). Based on (van der Schalie et al. 2018).

7.1.5 Frozen surfaces and snow

Under frozen surface conditions the dielectric properties of the water changes dramatically and therefore all pixels where the surface temperature is observed to be at or below 273 K are assigned with an appropriate data flag, this was determined using the method of Holmes et al. (2009).

7.1.6 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).

7.1.7 Rainfall

Rainstorms during the satellite overpass affect the brightness temperature observation, and are therefore flagged in LPRM. The flagging system for active rain is based on the rainfall index of Seto et al., 2005. This method makes use of the vertical polarized 36.5 GHz and 19 GHz observations to detect a rain event. Index values of 5 and beyond are used to identify an active rainstorm. Soil moisture retrievals with these index values are flagged.

7.1.8 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 now used 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.

8 Methodological description on the merging process of soil moisture data sets

8.1 Principle of the merging process

The generation of the long-term (40 years) soil moisture data sets involve three steps (Figure 7):

(1) merging the original passive microwave soil moisture products into one product,

(2) merging the original active microwave soil moisture products into one product, and

(3) merging all original active and passive microwave soil moisture products into one dataset.

The input datasets considered for the generating and validating the merged soil moisture product v05.2 are:

- Scatterometer-based soil moisture products
  - ERS-1 AMI surface soil moisture products generated at TU Wien (TU WIEN, 2013).
  - ERS-2 AMI WS soil moisture from ESA (ESA, 2017).
  - Metop-A ASCAT, Metop-B ASCAT soil moisture from H SAF 115 (H SAF 2019a) and H SAF 116 (H SAF 2019b).
  - Time span: 1991 – 2019-12-31
- Radiometer-based soil moisture products
  - SMMR, SSM/I, TRMM, AMSR-E, AMSR2, Windsat, SMOS and SMAP produced within ESA CCI.
  - Retrieval method: VUA-NASA LPRM v6 model inversion package
  - Time span: 1978 – 2019-12-31
- Modelled 0 – 10 cm soil moisture from the Noah land surface model of the Global Land Data Assimilation System (GLDAS; (Rodell et al. 2004)).
  - v2.1: Time span: 2000 – 2019-12-31 (0.25 degree resolution)
  - v2.0: Time span: 1948 – 2000 (0.25 degree resolution)
- In situ measurements:
  - Various networks: ESA/TU Wien International Soil Moisture Network (http://ismn.geo.tuwien.ac.at)
  - Time period: variable depending on station
  - Probes and depths: variable depending on station

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 – 2019-12-31, while measurements are provided at a 1-day sampling.
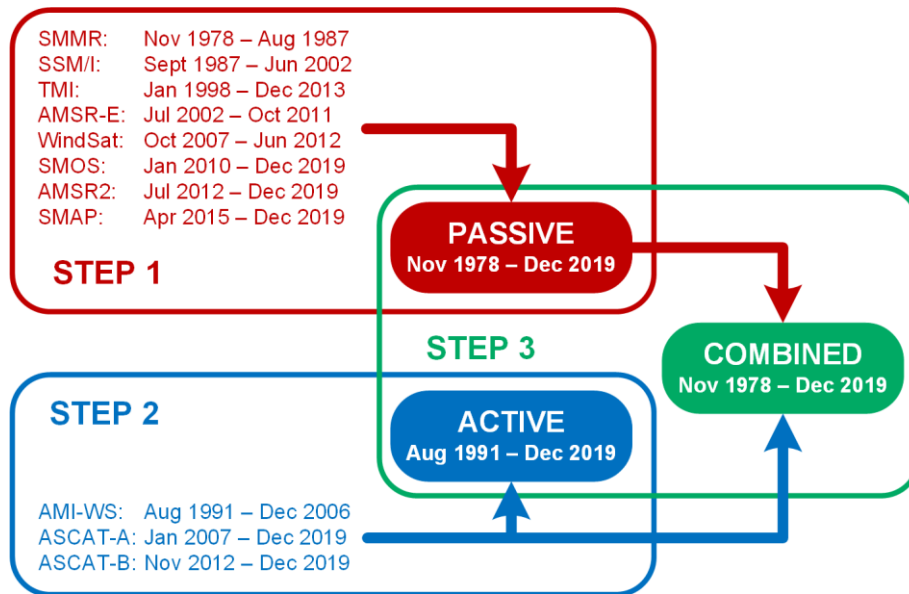
Figure 7: Overview of the three-step blending approach from the level 2l products to the final blended active & passive microwave soil moisture product for ESA CCI SM v05.2. (Adapted from Liu et al. 2012).

8.2   Overview of processing steps

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 8 for an overview):

1. Spatial Resampling
2. Temporal Resampling (including flagging of observations)
3. 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 climatology (for the COMBINED product)
4. Triple collocation analysis (TCA)-based error characterisation of all rescaled level 2 products
5. Polynomial regression between VOD and error estimates
6. Derivation of error estimates from the VOD regression in regions where they were not available after (4), i.e., where TCA is deemed unreliable
7. Merging rescaled passive and active time series into the PASSIVE, ACTIVE, and COMBINED product, respectively.
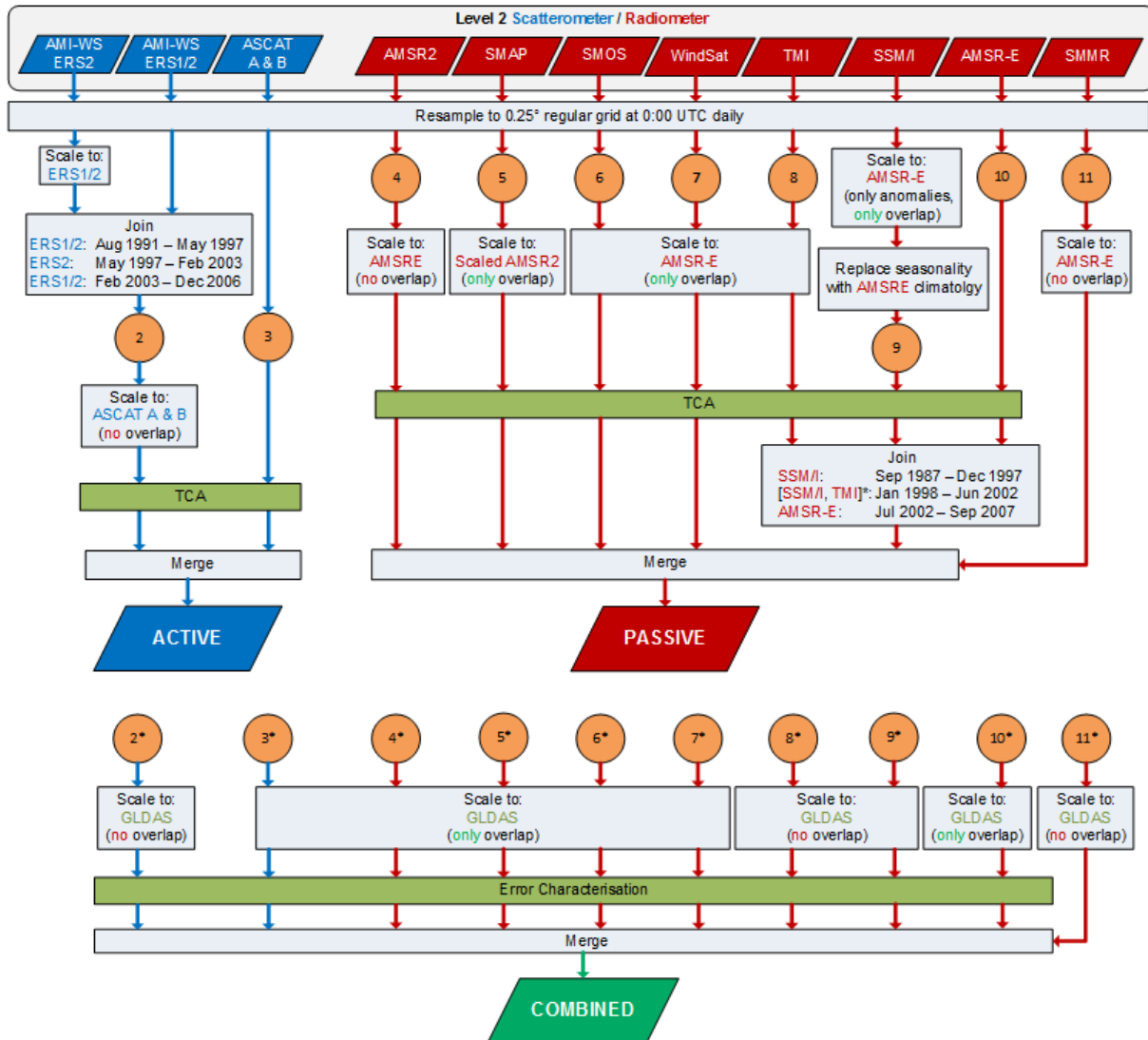
*Figure 8: O*verview of the processing steps in the ESA CCI SM product generation (vv05.2): The merging of two or more data sets is done by weighted averaging and involves overlapping time periods, whereas the process of joining data sets only concatenates two or more data sets between the predefined 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 14*).

8.3 Description of Algorithms

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.

8.3.1   *Resampling*

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

8.3.1.1 *Spatial Resampling*

The 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 undersampling. The resolution of the products is often adopted by land surface models. 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 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.

Table 3: Major characteristics of passive and active microwave instruments and model product

	Passive microwave products								Active microwave products				Model product	
	SMMR	SSM/I	TMI	AMSR-E	AMSR2	Windsat	SMOS	SMAP	AMI-WS	AMS-WS	ASCAT	ASCAT	GLDAS-2-Noah	GLDAS-2-Noah
Platform	Nimbus 7	DMSP	TRMM	Aqua	GCOM-W1	Coriolis	SMOS	SMAP	ERS1/2	ERS2	Metop-A	Metop-B	---	---
Product	VUA NASA	VUA NASA	VUA NASA	VanderSat NASA	VanderSat NASA	VUA NASA	VanderSat NASA	VanderSat NASA	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	LPRM v06 (van der Schalie et al. 2015)	LPRM v06 (van der Schalie et al. 2015)	LPRM v06 (van der Schalie et al. 2015)	LPRM v06 (van der Schalie et al. 2015)	LPRM v06 (van der Schalie et al. 2015)	LPRM v06 (van der Schalie et al. 2015)	LPRM v06 (van der Schalie et al. 2015)	LPRM v06 (van der Schalie et al. 2015)	TU WIEN Change Detection (Wagner et al. 1999)	TU WIEN Change Detection (Wagner et al. 1999)	TU WIEN Change Detection (H SAF in prep.)	TU WIEN Change Detection (H SAF in prep.)	V2.0	V2.1
Time period used	1/1979–8/1987	9/1987–12/2007	1/1998–12/2013	7/2002–10/2011	5/2012–12/2019	10/2007–7/2012	1/2010–12/2019	4/2015–12/2019	7/1991–12/2006	5/1997–2/2007	1/2007–12/2019	11/2012–12/2019	1/1948–12/2010	1/2000–12/2019
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	5.3 GHz	5.3 GHz	5.3 GHz	5.3 GHz	---	---
Original spatial resolution* (km²)	150x150	69 × 43	59 × 36	76 × 44	35 × 62	25 x 35	40 km	38 x 49	50 × 50	25 × 25	25 × 25	25 × 25	25 × 25	25 × 25
Spatial coverage	Global	Global	N40° to S40°	Global	Global	Global	Global	Global	Global	Global	Global	Global	Global	Global
Swath width (km)	780	1400	780/897 after boost in Aug 2001	1445	1450	1025	600	1000	500	500	1100 (550×2)	1100 (550×2)		
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	Descending: 10:30	Descending 10:30	Descending; 09:30	Descending; 09:30	---	---
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}	Degree of saturation (%)	Degree of saturation (%)	Degree of saturation (%)	Degree of saturation (%)	$kg\ m^{-2}$	$kg\ m^{-2}$

*For passive and active microwave instruments, this stands for the footprint spatial resolution.

8.3.1.2 *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 elaborated temporal resampling strategy firstly searches for the valid observation that is closest to the reference time. In case there are only invalid observations, which are flagged other than "0" (zero), within a certain 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 action is taken. This strategy results in data gaps when no observations within ±12 hours from the reference time are available. For the modelled soil moisture datasets, no resampling is required as they already include the reference time stamp of 0:00 UTC. The LPRM (passive) soil moisture estimates based on night-time (often the descending mode) observations are more reliable than those obtained during the day (often the ascending mode). This is mainly caused by the complexity to derive accurate estimates of the effective surface temperature during the day. For this reason, only night-time soil moisture observations from radiometers are used for the merged product.

During the temporal resampling stage, flagging is applied to the datasets where relevant information is available. For the LPRM products, the data is flagged for high VOD using the VOD fields provided in the data product. At vv05.2 of the ESA CCI SM, LPRM v6 is used for all passive sensors and 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.

8.3.2 *Rescaling*

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 at either level, the datasets need to be rescaled into a common climatology. All soil moisture observations of each product are rescaled to the climatology of a different reference, namely AMSRE, ASCAT or GLDAS for the passive, active or combined product respectively.

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). The matching is shown by means of an example for a grid point centred at 41.375ºN, 5.375ºW. Figure 9 shows for this location the time series of soil moisture estimates from

GLDAS-Noah, AMSR-E and ASCAT, respectively. CDF-matching for these time series is performed in the following way:

1. For the time-collocated data points CDFs are computed (Figure 9: a-c). In the passive product AMSR2 is scaled using the parameters derived from the last 3 years of AMSRE and first 3 years of AMSR2. SMAP is then scaled to the scaled AMSR2.

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

3. Use the npercentiles of the CDF curves to define n-1 segments. The CDF curves of these circled values are shown in Figure 10: a, b and c.

4. The n percentile values from the AMSR-E and ASCAT CDF curves are plotted against those of Noah (Figure 10: d and e) 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 (Figure 10: f).

$$sm_r = slope_i * sm + intercept_i$$

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.
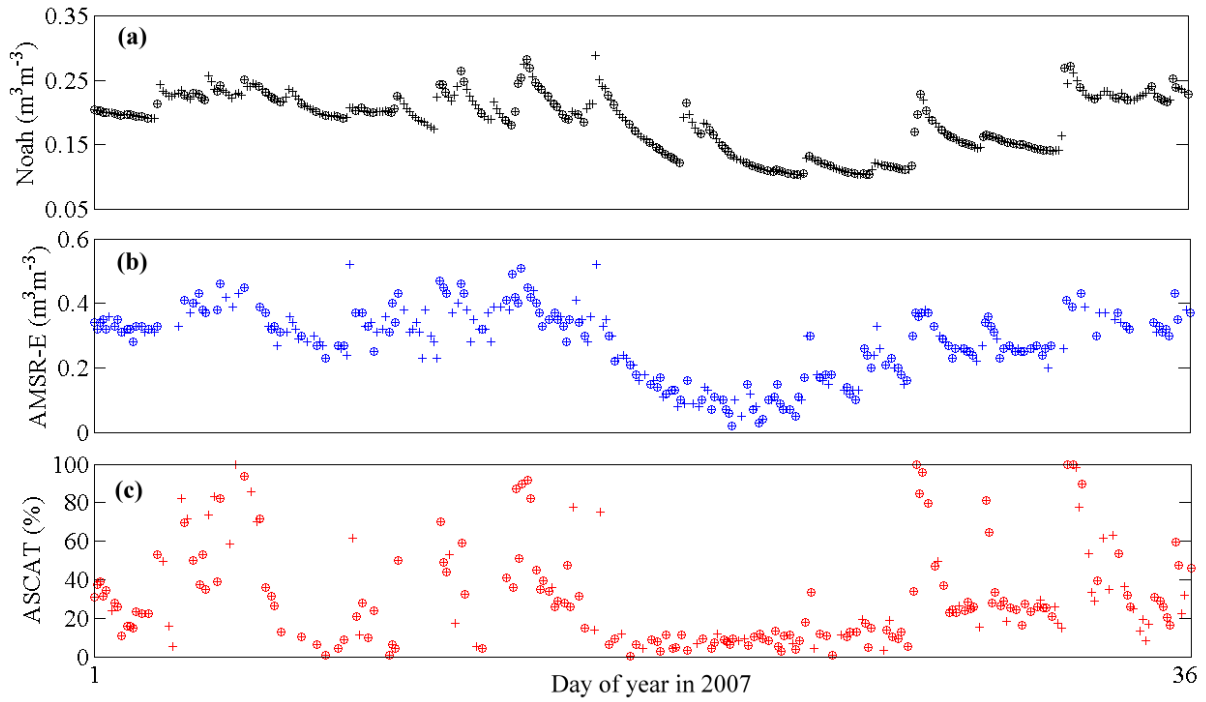
Figure 9: Time series of surface soil moisture estimates from (a) GLDAS-Noah, (b) AMSR-E and (c) ASCAT for a grid cell (centered at 41.375° N, 5.375° W) in 2007. Circles represent days when Noah, AMSR-E and ASCAT all have valid estimates (Figure taken from Liu et al. 2011).
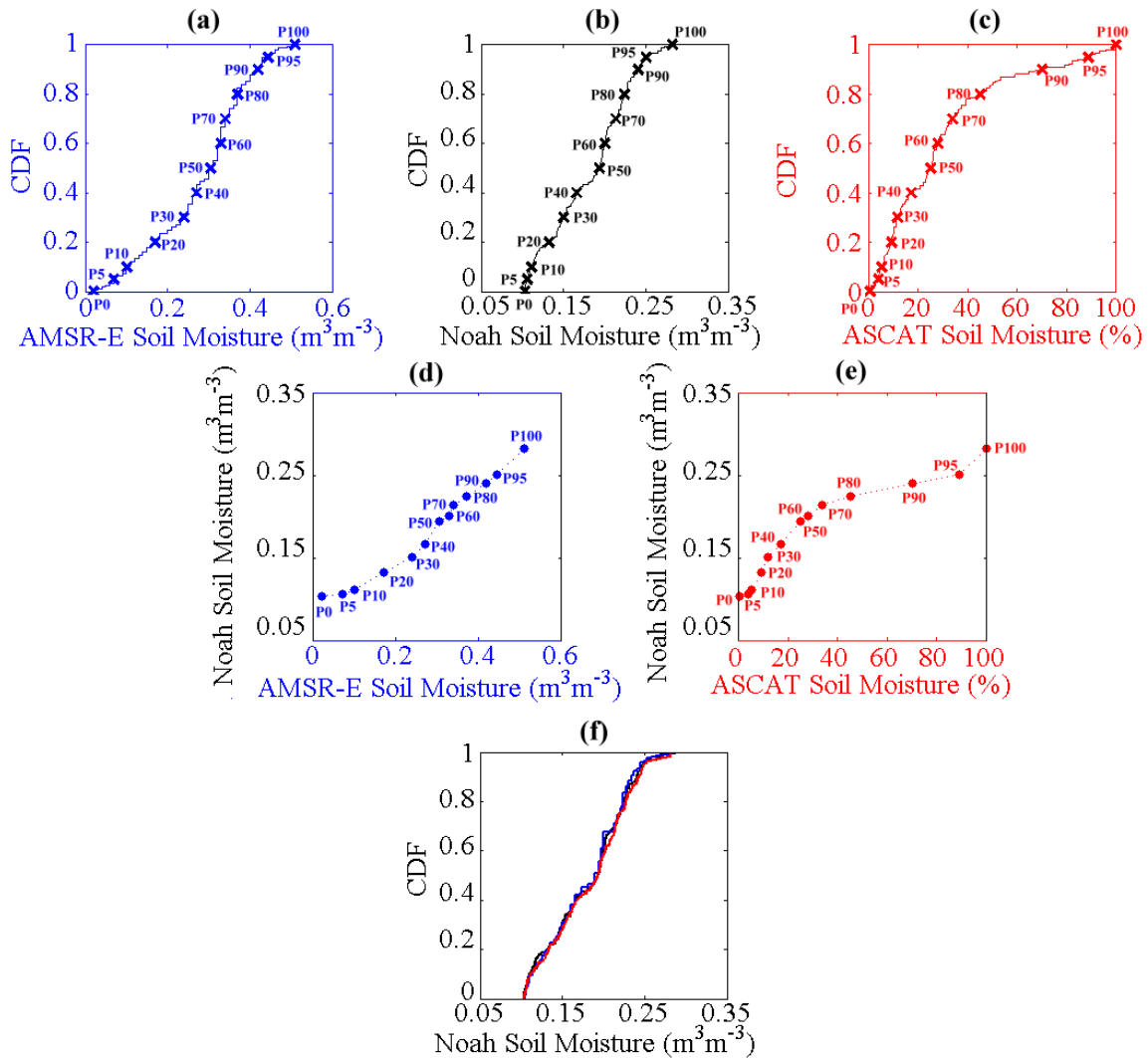
Figure 10: Example illustrating how the cumulative distribution function (CDF) matching approach was implemented to rescale original AMSR-E and ASCAT against Noah soil moisture product in this study. (a, b, c) CDF curves of AMSR-E, GLDAS-noah and ASCAT soil moisture estimates for the grid cell shown in Figure 8 (d) Linear regression lines of AMSR-E against Noah for 12 segments. (e) Same as (d), but for ASCAT and Noah. (f) CDF curves of GLDAS-Noah (black), rescaled AMSR-E (blue) and rescaled ASCAT (red) soil moisture products. (Figure taken from Liu et al. 2011)

8.3.3 Error characterization

Errors in the individual active and passive products are characterized by means of triple collocation analysis. These errors are used both for estimating the merging parameters and for characterizing the errors of the merged product (see section 8.3.4).

8.3.3.1 *Triple collocation analysis*

Triple collocation analysis is a statistical tool that allows estimating the individual random error variances of three data sets without assuming any of them 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:

$$\sigma_{\varepsilon_a}^2 = \sigma_a^2 - \frac{\sigma_{ap}\sigma_{am}}{\sigma_{pm}}$$

$$\sigma_{\varepsilon_p}^2 = \sigma_p^2 - \frac{\sigma_{pa}\sigma_{pm}}{\sigma_{am}}$$

Eqn. 8-1

where σ_ε^2 denotes the error variance; σ^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). Notice 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.

8.3.4 *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 in case of insignificant Pearson correlation (p-value < 0.05) between any of the data sets. In these areas, error estimates are derived from the mean VOD (derived from AMSR-E in the entire mission period) at that particular location:

$$SNR_x = \sum_{\{i=0\}}^{N} a_i VOD_x^i$$

Eqn. 8-2

Where the subscript denotes the spatial location; and the parameters a_i are derived from a global polynomial regression between VOD and TCA based error estimates at locations where they are considered reliable (i.e., all data sets are significantly correlated). 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.

8.3.5 Merging

The merging procedure consists of (1) merging the original passive microwave product into the PASSIVE product, (2) merging the original active microwave products into the ACTIVE product, and (3) merging the original active and passive microwave products into the COMBINED product. 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. Such weighted average is calculated as

$$\Theta_m = \sum_{i=1}^{N} w_i \cdot \Theta_i \qquad \qquad Eqn.\ 8\text{-}2$$

where Θ_m denotes the merged soil moisture product; Θ_i are the soil moisture products that are being merged, and w_i are the merging weights.

8.3.5.1 Weight estimation

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

$$w_i = \frac{\sigma_{\varepsilon_i}^{-2}}{\sum_{j=1}^{N} \sigma_{\varepsilon_j}^{-2}} \qquad \qquad Eqn.\ 8\text{-}3$$

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. 8-3. 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 harmonization 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 8.3.2. Therefore, the calculation of the weights using Eqn. 8-3 suffices, keeping in mind that they represent rescaled error variances of rescaled data sets.

8.3.5.2 *Merging passive microwave products*

Differences in sensor specifications, particularly in microwave frequency and spatial resolution, result in different absolute soil moisture values from SMMR, SSM/I, TMI and AMSR-E. Even though SMMR and AMSR-E have a similar frequency (i.e., C-band), their absolute values are different. Therefore, a Spearman and Pearson correlation analysis was performed between the different soil moisture products to identify differences and correspondences between the data sets (Liu et al. 2012). Based on this analysis, the AMSR-E soil moisture retrievals were identified as more accurate than the other passive products due to the relatively low microwave frequency and high temporal and spatial resolution of the sensor. Thus, soil moisture retrievals from AMSR-E are selected as the reference to which soil moisture retrievals from SMMR, SSM/I, TMI, WindSat, SMOS and AMSR2 are rescaled and merged on a pixel basis according to the following steps. SMAP is later CDF-matched to the rescaled AMSR2 data.

Merging SSM/I and TMI with AMSR-E

1. Rescale original TMI against the AMSR-E reference using the piece-wise linear cumulative distribution function (CDF) matching technique (Section 8.3.2) based on their overlapping period (Figure 11a),
2. Decompose SSM/I and AMSR-E time series into their own seasonality and anomalies (Figure 11b). This is done for their overlapping period from July 2002 through December 2007. The seasonality for each sensor was calculated by taking the average of the same day of the year for their overlapping period. The seasonality ($\overline{SM}$) is one time series of 366 values, one value for each day of the year (*DOY*):

$$\overline{SM}_{DOY} = \left(\sum_{YR=2002}^{2007} SM_{DOY}^{YR} \right) \Big/ N$$

Eqn. 8-4

where *YR* represents the year 2002 through 2007; *N* represents the number of valid soil moisture retrievals. The value of $\overline{SM}_{366}$ is only taken from the year 2004 as that is the only leap year (i.e., 366 days) between 2002 and 2007. The anomalies (*ANO*) over their individual entire periods were obtained by removing the sensor's seasonality $\overline{SM}$ from the original (*ORI*) time series:

$$ANO_{DOY}^{YR} = ORI_{DOY}^{YR} - \overline{SM}_{DOY}$$

Eqn. 8-5

where *YR* represents the year 1987 through 2007 for SSM/I and 2002 through October 2011 for AMSR-E.

3. Rescale "anomalies of SSM/I" against "anomalies of AMSR-E" using the piece-wise linear CDF matching technique (Figure 11c).
4. Add the AMSR-E seasonality to the "rescaled SSM/I anomalies" (from Step 3) and obtain reconstructed SSM/I (Figure 11d).
5. Merge the reconstructed SSM/I, rescaled TMI, and original AMSR-E to obtain the merged SSM/I-TMI-AMSR-E dataset (Figure 11e). The lower the measurement frequency, the more accurate soil moisture retrievals can be expected. Therefore AMSR-E is used for July 2002 – December 2008 and the rescaled TMI is used for January 1998 – June 2002 between N40º and S40º. Otherwise the reconstructed SSM/I is used.

Merging SMMR with SSM/I-TMI-AMSR-E

The overlapping period between SMMR and other sensors is too short to perform the rescaling as conducted on retrievals from other sensors. In order to incorporate SMMR (1979 – 1987) soil moisture retrievals into the merged product, we assumed that the dynamic range of SMMR retrievals is the same as the range of merged SSM/I-TMI-AMSR-E dataset. Following this assumption, we produced the rescaled SMMR (Nov 1978 to July 1987) by matching the CDF curve of SMMR against that of the merged SSM/I–TMI–AMSR-E dataset for each grid point. The CDF curve is calculated based on all observation of both data sets. Together with the merged SSM/I-TMI-AMSR-E dataset, we obtained the merged SMMR-SSM/I-TMI-AMSR-E soil moisture product covering the period Nov 1978 – Sep 2007 (Figure 11). It should be emphasized that the CDF matching process changes the absolute values of SMMR, SSM/I and TMI products, but does not change the relative dynamics of the original retrievals, which is demonstrated in Liu et al. (2011).

Table 4 Used passive sensors in the PASSIVE product

Time Period	Passive Sensors (mode: ascending (a) or descending (d))
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 – 40N], [90S – 40S], TMI (a/d) [40N – 40S]
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 - 2019-12-31	SMOS (a), AMSR2 (d), SMAP (d)

Merging SMOS, WindSat, SMAP, and AMSR2 with SMMR, SSM/I, TMI, AMSR-E

WindSat data (1 October 2007 to 31 June 2012) bridge the operational time gap between AMSR-E, which failed to deliver data from 4 October 2011 onwards, and AMSR2, for which data are available from 02 July 2012 onward. SMOS data in ascending satellite mode are available from 1 July 2010 onward. The CDFs between WindSat and AMSR-E, and SMOS and AMSR-E are calculated based on their respective overlapping time periods with AMSR-E. AMSR2 and SMAP do have no temporal overlap with AMSR-E and can therefore not be rescaled directly to it. Instead, the first 3 years of AMSR2 are scaled to the last 3 years of AMSRE. SMAP is the scaled to the rescaled AMSR2.

Within the time period from 1 October 2007 to May 2015 there are various combinations of data overlap Figure 8, Table 4, and Figure 14b illustrate these overlaps. The data periods AMSR-E & WindSat (1 October 2007 to 30 June 2010), AMSR-E & WindSat & SMOS (1 July 2010 to 3 October 2011), WindSat & SMOS (4 October 2011 to 30 June 2012), are then extended with AMSR2 & SMOS (1 July 2012 to 31 December 2018). The resulting product hereafter is referred to as the PASSIVE product. The following paragraph describes the in more detail the process of merging these datasets, when more than one sensor is used.
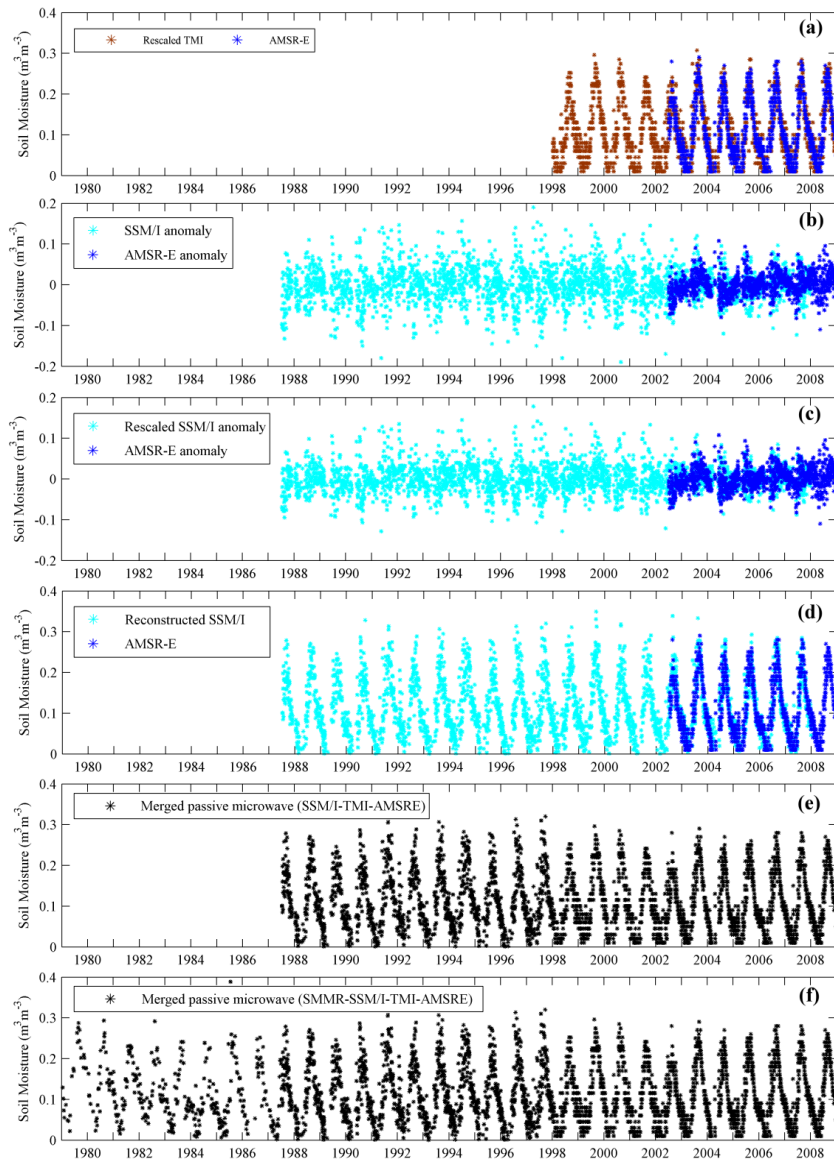
Figure 11: Example illustrating how (a) the TMI was rescaled against AMSR-E, (b-e) the SSM/I anomalies were rescaled against AMSRE-E anomalies, reconstructed and merged with rescaled TMI and AMSR-E, and (e) the SMMR was rescaled and merged with the others. The grid cell is centred at 13.875°N, 5.875°W (Image courtesy Liu et al. 2012).

Merging in periods where more than one sensor is used

As it can be seen from Figure 14 there are periods where more than one passive dataset is available, i.e., AMSR-E & WindSat. In these periods, a weighted average of the respective sensors is used to construct the merged PASSIVE product (see Eqn. 8-5). Error estimates are obtained from triple collocation analysis (see section 8.3.3) using ASCAT/ERS and GLDAS-Noah data to complement the respective triplets.

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.

8.3.5.3 *Merging active microwave products*

Different sensor specifications between ERS1/2 and ERS2 (e.g. spatial resolution) need to be compensated by using the same rescaling techniques performed on the radiometer data sets. The CDF curves for ERS2 are calculated based on the overlap with ERS1/2. Rescaling ERS2 against ERS1/2 and then merging them generates the AMI-WS active data set, which is subsequently scaled and merged to the Metop-A ASCAT data (Figure 8).

Table 5 and Figure 14a show the sensors used in the ACTIVE product for the individual time periods.

Table 5 Used active sensors in the ACTIVE product

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 – 2019-12-31	Metop-A ASCAT, Metop-B ASCAT

An example of a soil moisture time series from AMI-WS ERS1/2 and Metop-A ASCAT for the grid point centred at 13.875°N, 5.875°W (Niger River basin in southern Mali) is shown in Figure 12, where the AMI-WS ERS1/2 is labelled as SCAT to denote its predecessor role to ASCAT. The AMI-WS ERS1/2 and Metop-A ASCAT soil moisture variations are scaled between the lowest (0%) and highest (100%) values over their individual operational period. The limited overlap

	Algorithm Theoretical Baseline Document (ATBD)	Product Version v05.2 Doc Issue 1.0 Date 29-05-2020

soil moisture cci

in time (i.e., a few months) and space (i.e. only Europe, Northern America and Northern Africa) rules out the global adjustment method based on the information of their overlapping period, such as applied between TMI and AMSR-E. Figure 12 also shows the evident AMI-WS ERS1/2 data gap from 2001 to 2003.

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 use for each grid point the CDF curves of both datasets to rescale AMI-WS to Metop-A ASCAT before merging them (Figure 12b). Metop-A ASCAT data from 1 January 2007 to 5 November 2012 are joined with AMI-WS data from 5 August 1991 to 31 December 2006. In the time period from 6 November 2012 to 2019-12-31 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. Joining AMI-WS & Metop-A ASCAT from 5 August 1991 to 5 November 2012 with Metop-A ASCAT & Metop-B ASCAT from 6 November 2012 to 2019-12-31 generates the ACTIVE product (Figure 8 and Figure 14a).
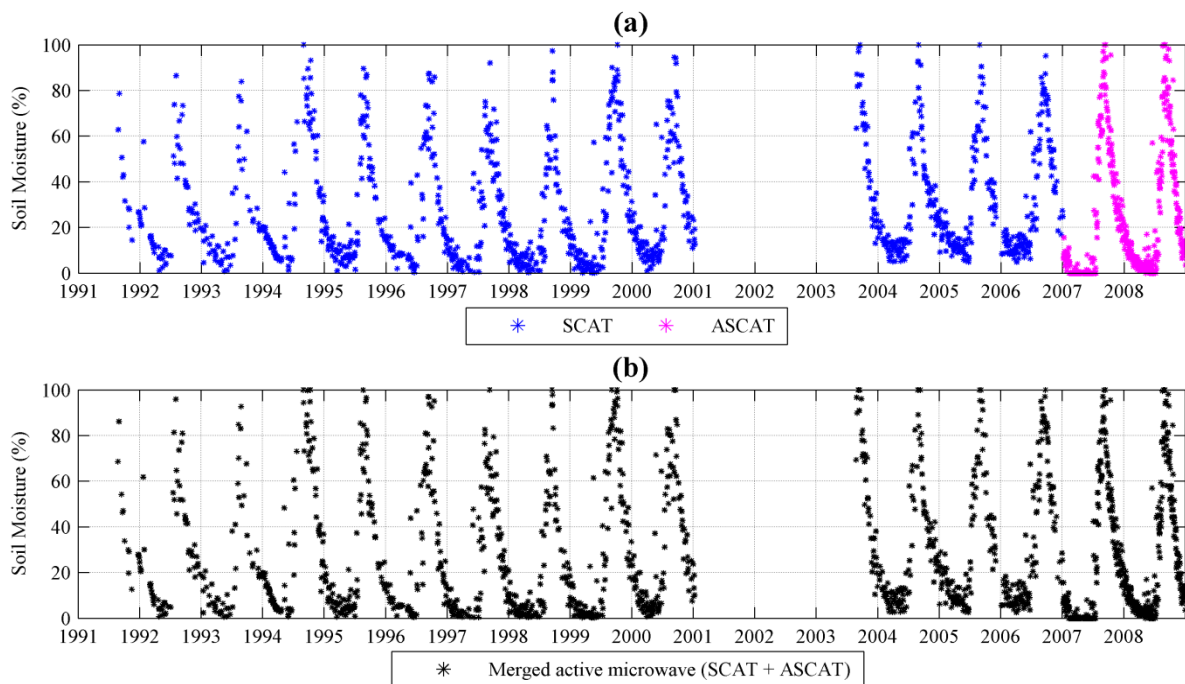

Figure 12: Example illustrating fusion of ERS1/2 (SCAT) with ASCAT. Note the data gap from 2001 – 2003, which will be filled by ERS2 data. The grid point is centred at 13.875°N, 5.875°W (Image courtesy Liu et al. 2012)

8.3.5.4 *Merging passive and active microwave products*
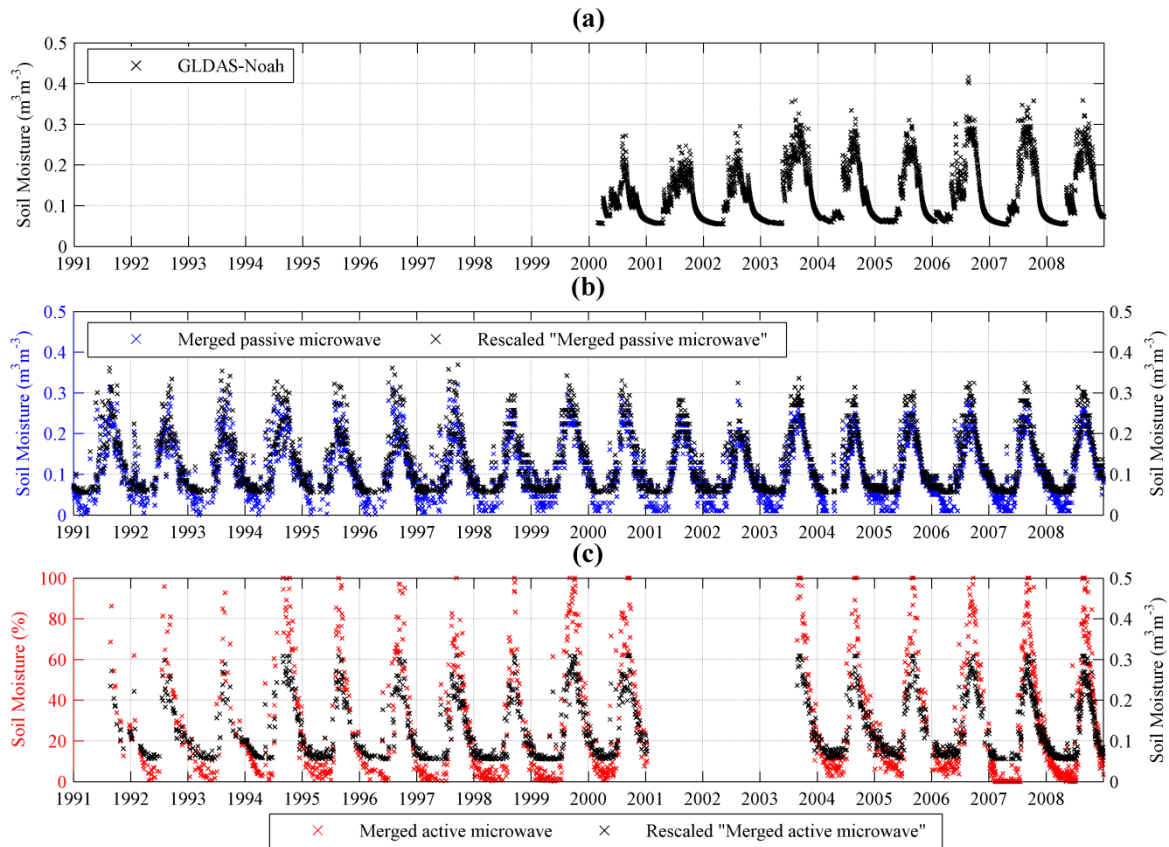
Figure 13: Rescaling the merged passive and active microwave product against the GLDAS-1-Noah simulation. (a) GLDAS-1-Noah soil moisture; (b) merged passive microwave product and one rescaled against GLDAS-1-Noah; (c) same as (b) but for active microwave product. The grid cell is centred at 13,875°N, 5.875°W (Image courtesy Liu et al. 2012). Since CCI SM v04.4 released in November 2018 GLDAS 2.1 is used for rescaling all products.

For generating the combined product, climatologies of all passive and active level 2 data sets are first harmonized by rescaling against GLDAS-2.1 (see Sec. 6.2). Considering the covering period of each microwave instrument we divided the entire time period (1978 – 2019-12-31) into eleven segments. Table 6 list these time periods, and Figure 14c illustrates also the spatial sensor usage at global scale.

Table 6 Used sensors in individual time periods. Note that Metop-B ASCAT data are available from 06 November 2012 onwards.

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-40N], [90S-40S], TMI [40N-40S]
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 - 2019-12-31	Metop-A ASCAT, Metop-B ASCAT	AMSR2, SMOS, SMAP
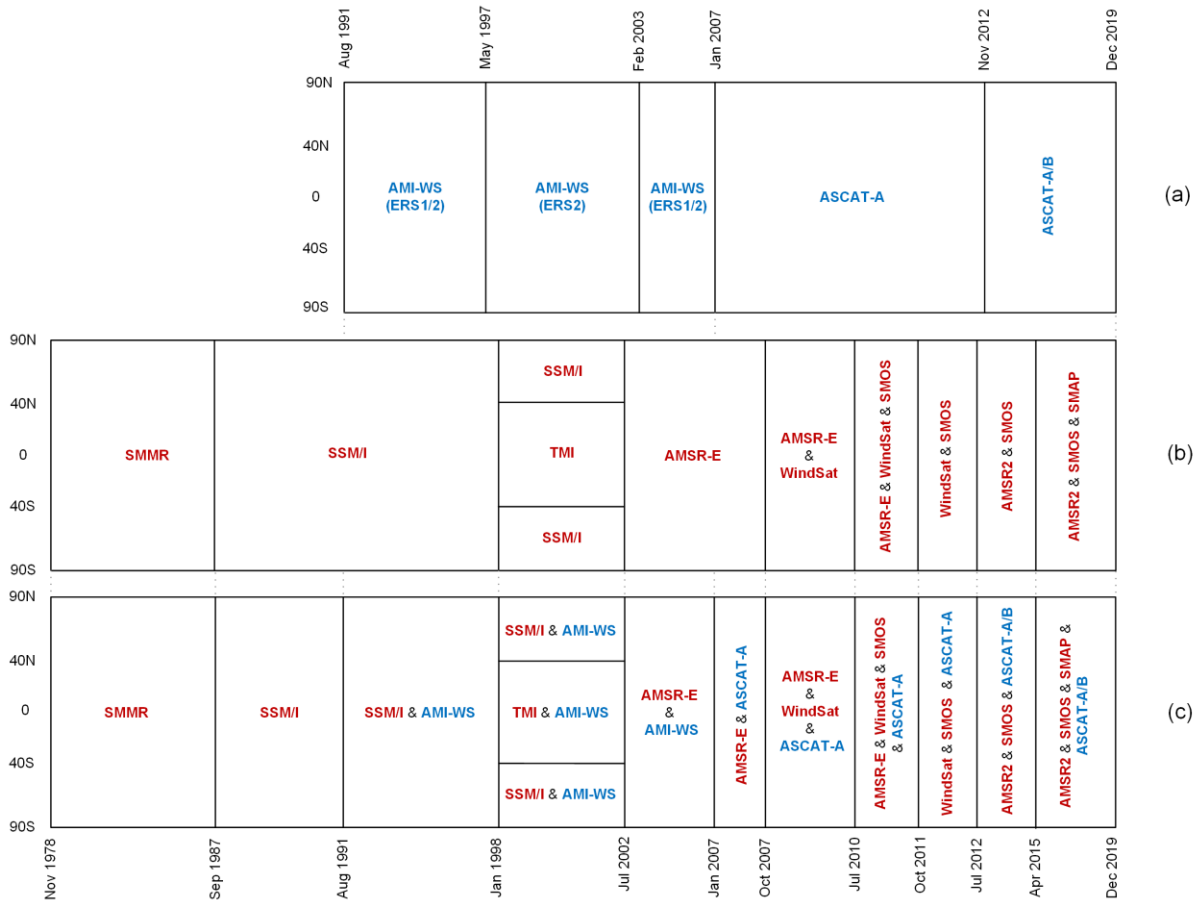
Figure 14: Spatial and temporal coverage of soil moisture products from different sensors in the CCI SM vv05.2 COMBINED product. Figure adapted from (Dorigo et al. 2017).

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.

8.4 Known Limitations

8.4.1 Passive merged CCI product

8.4.1.1 Using night-time observations only

For the current version of the merged passive product only descending overpasses, corresponding to night-time / early morning observations, were considered. 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, recent 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. If day-time observations could be introduced to the blended product, this would significantly increase the observation density.

8.4.2 Active Product

8.4.2.1 Intercalibration of ERS and ASCAT

The generation of the ERS and ASCAT products is still based on their individual time series. The merged ERS + ASCAT could significantly profit from an appropriate Level 1 intercalibration. Besides improving the quality of the individual measurements this would improve the robustness of the calculation of the dry and wet references.

8.4.3.2 Data gaps

Similar as for the passive products, merging ERS and ASCAT into a merged dataset is based on a strict separation in time. Gaps in ASCAT time series can be potentially filled with ERS observations, although the spatial and temporal overlap between both sensors is limited.

8.5 Scientific Advances under Investigation

8.5.1 All products

8.5.1.1 Separate blending of 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 spectral components may be subject to different error magnitudes (Su et al. 2015, Draper et al. 2015). Therefore, we will investigate the feasibility of blending the climatologies and the anomalies of the data sets separately.

8.5.1.2 *Improved sensor inter-calibration*

Currently, inter-calibration between active and passive data sets is done using CDF-matching against a long-term consistent land surface model. However, in order to achieve a full model-independence of the CCI SM products, we will investigate alternative inter-calibration approaches, for instance using lagged-variable based approaches or homogeneity tests (Su et al. 2015, 2016).

8.5.1.3 *Data density and availability*

In the current versions, gaps are only filled if the weight of the available product is above a relatively crudely defined empirical threshold. This threshold will be refined to find a best compromise between data density and product accuracy.

8.5.2 ACTIVE product only

8.5.2.1 *Metop ASCAT wetting trend correction*

Measurements of sensors operating in RFI sensitive frequency bands (C, L) may be disturbed by external sources and show behaviour that is not representative of the actual soil moisture conditions in some areas. These areas should either be flagged as unreliable (passive sensors) or measurements have to be corrected. Within HSAF, soil moisture from ASCAT was found to show RFI caused positive trends in the measured backscatter signals that result in erroneous wetting trends in ASCAT SM in some areas. The impacts of an experimental version of the H-SAF-produced Metop ASCAT dataset where the wetting trend has been corrected are presently being evaluated.

8.5.3 PASSIVE product only

8.5.3.1 *Development of a solely satellite based soil moisture data record*

Within the climate community there is a strong preference for climate records that are solely satellite based. Any additional dataset that is used in a soil moisture retrieval algorithm could potentially lead to a dependency between a model and an observation. This is also why research was set up to investigate the possibility to develop an independent ancillary free soil moisture data set.

In addition, ancillary data could also have a strong impact on the spatial distribution of soil moisture as shown in **Error! Reference source not found.**. Here the artificial squared patterns o f the 1 degree FAO soil property map are still visible in the original LPRM soil moisture product. However, these patterns disappear when only the dielectric constant is used. A study is set up to derive soil moisture from the dielectric constant records without making use of any ancillary

datasets, with such an approach you will create an independent dataset that could be used as a benchmark for different modelled soil moisture datasets.
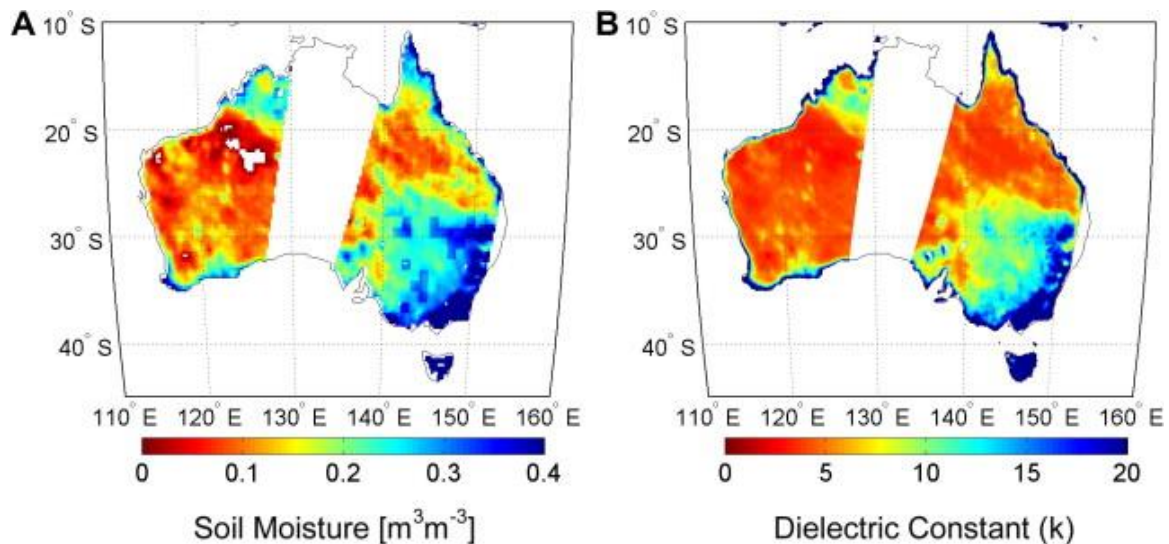

Figure 15: (A) Original descending LPRM Soil Moisture of May 18, 2007 of Australia and (B) the LPRM ancillary data free dielectric constant dataset from the same brightness temperatures. Note the disappearance of the artificial squared patterns in south-eastern Australia.

8.5.3.2 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 optimization procedure for soil moisture retrievals through a quasi-global precipitation-based verification technique, the so-called R_{value} metric. In this optimization, 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% (**Error! Reference source not found.**) 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.
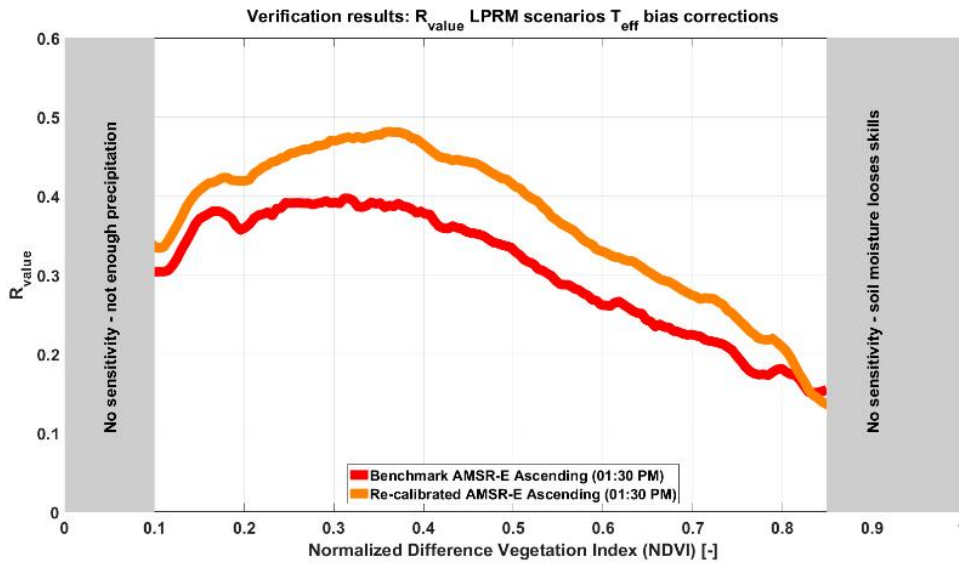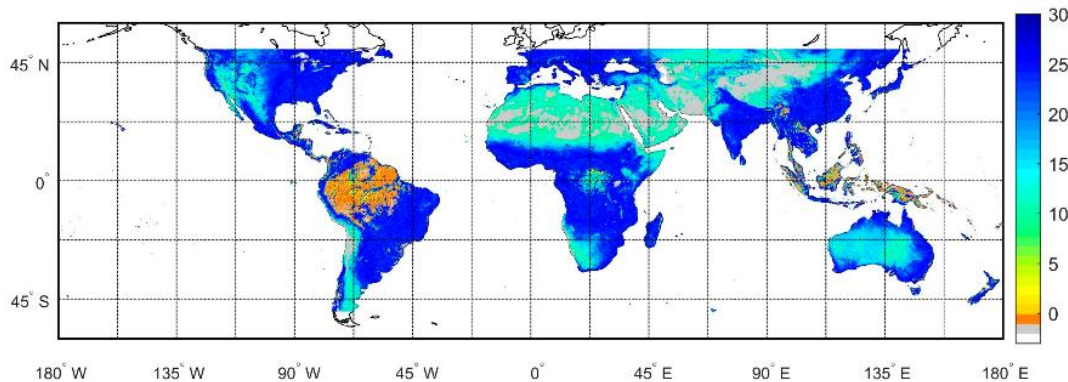
Figure 16: (up) comparison of R$_{value}$ with the old and new daytime land surface temperature binned over NDVI, (down) the difference in R$_{value}$ compared to the old temperature parameterization in [%].

Further work here will focus on similarly updating the temperature for night-time observations. Also, in order to remove model dependency for the L-band soil moisture retrievals, we will look into combining the L-band observations with Ka-band observations from other satellites with similar overpass times. For this, the Ka-band linear regressions need to be optimized specifically to match the L-band sensing depth and overpass times and will follow an approach similar to the one used by Parinussa et al. (2016).

8.5.3.3 *Update error characterization*

As a solid knowledge of the uncertainties and errors of the soil moisture datasets is important for many applications. The results in Section 7.1.2 on the soil moisture uncertainties need to be updated to include results based on L-band observations and the latest parameterization update for the C- and X-band.

8.5.3.4 *Using night-time observations only*

Based on extensive product validation and triple collocation we will try to address the uncertainty of both modes. Based on these results we will decide how both observations modes can be considered in the generation of a single merged passive product, potentially leading to improved observation frequency with respect to the single descending mode used in the CCI SM product. An important step in this step was made by Parinussa et al. (2016).

8.5.4   COMBINED product

8.5.4.1 *L-Band Reference climatology*

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.

8.5.4.2 *SMOS L2 product*

Soil Moisture from passive sensor observations (including SMOS) in the ESA CCI SM is derived using the Land Parameter Retrieval Model. SMOS IC (Fernandez-Moran et al. 2017) is an alternative SM product derived from SMOS brightness temperature measurements, that is as independent as possible from any ancillary data. Replacing the current SMOS LPRMv6 SM with SMOS-IC could improve the passive and combined product quality in the according sub-periods.

8.5.4.3 *Break detection and correction*

When merging active and passive sensors into the combined product, inhomogeneities (or structural breaks) in then mean and variance of observations within adjacent sensor periods (see Figure 14) are potentially introduced. Su et al. (2016) used statistical tests to detect

breaks in the data set. Preimesberger et al. (in review) describe methods to reduce the number of breaks in the dataset that are detected this way and explore their impact on the data. A method based on relative, empirical distribution matching is found which reduces both, inhomogeneities in mean and variance with respect to a reference reanalysis dataset.

8.5.4.4 *Gap-filled product*

Due to temporally varying data availability, the current products contain data gaps. Gaussian process regression models are under investigation to fill those. The development is mostly driven by Vegetation Optical Depth data but a transition to soil moisture is planned in the near future.

For details of recently undertaken work including quality assessment of the combined product, please refer to (Dorigo et al. 2017)

9   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

Albergel, C., Zakharova, E., Calvet, J.C., Zribi, M., Pardé, M., Wigneron, J.P., Novello, N., Kerr, Y., Mialon, A., & Fritz, N.E.D. (2011). A first assessment of the SMOS data in southwestern France using in situ and airborne soil moisture estimates: The CAROLS airborne campaign. *Remote Sensing of Environment, 115*, 2718-2728

Bartalis, Z., Wagner, W., Dorigo, W., & Naeimi, V. (2010). Accuracy and stability requirements of ERS and METOP scatterometer soil moisture for climate change assessment. In, *European Space Agency Living Planet Symposium* (p. 7 p.). Bergen, Norway: European Space Ageny

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

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

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.

Crapolicchio, R., Lecomte, P., & Neyt, X. (2004). The Advanced Scatterometer Processing System for ERS Data: Design, Products and Performances. In, *ENVISAT & ERS Symposium*. Salzburg, Austria, 6-10 September 2004

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 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

Fernandez-Moran, R., Al-Yaari, A., Mialon, A., Mahmoodi, A., Al Bitar, A., De Lannoy, G., Rodriguez-Fernandez, N., Lopez-Baeza, E., Kerr, Y., Wigneron, J. P. (2017) SMOS-IC: An Alternative SMOS Soil Moisture and Vegetation Optical Depth Product. *Remote Sens. 2017, 9(5), 457;* https://doi.org/10.3390/rs9050457

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

H SAF (in prep.) Algorithm Theoretical Baseline Document (ATBD), Metop ASCAT Soil Moisture Data Records v0.8

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

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 polarization 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

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 polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing, 39(8), 1643-1654*

	Algorithm Theoretical Baseline Document (ATBD)	Product Version v05.2 Doc Issue 1.0 Date 29-05-2020

soil moisture cci

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., Holmes, T.R.H., Wanders, N., Dorigo, W.A., & de Jeu, R.A.M. (2015). A Preliminary Study toward Consistent Soil Moisture from AMSR2. *Journal Of Hydrometeorology, 16*, 932-947

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

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

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

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., 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

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

Wilson, J.J.W., Anderson, C., Baker, M.A., Bonekamp, H., Saldana, J.F., Dyer, R.G., Lerch, J.A., Kayal, G., Gelsthorpe, R.V., Brown, M.A., Schied, E., Schutz-Munz, S., Rostan, F., Pritchard, E.W., Wright, N.G., King, D., & Onel, U. (2010). Radiometric Calibration of the Advanced Wind Scatterometer Radar ASCAT Carried Onboard the METOP-A Satellite. *IEEE Transactions on Geoscience and Remote Sensing, 48*, 3236-3255