Algorithm Theoretical Basis Document (ATBD) for the

SMOS Level 2 Soil Moisture Processor Development Continuation Project

Contract Number: 4000125649/15/I-SBo

Submitted by:

ARGANS Ltd.
1 Davy Road, Plymouth Science Park
PL68BX, Plymouth, United Kingdom

Prepared by:

CESBIO, IPSL-Service d'Aéronomie,
INRA-EPHYSE, Tor Vergata University, WSL/Gamma RS and FMI

Prepared for:

The European Space Agency (ESA)

Algorithm Theoretical Basis Document (ATBD) for the

SMOS Level 2 Soil Moisture Processor Development Continuation Project

Contract Number: 4000125649/15/I-SBo

Author:	_____	09/09/2019
	SM-ESL	Date
Project Manager/ Project Engineer:	_____	09/09/2019
	Manuel Arias	Date
Quality Assurance:	_____	09/09/2020
	Yann Kerr	Date

**Algorithm Theoretical Basis
Document (ATBD)
for the**

**SMOS Level 2 Soil Moisture
Processor Development Continuation
Project**

Contract Number: 4000125649/15/I-SBo

Revision History

ARGANS Version	ESL Internal Version / Rév.	Date	Pages	Changes
0.2.b Pre-draft	0	20/01/2005		General layout
	01	20 Feb 2005	all	Pre-predraft
	02.b	16/03/2005	all	Pre-Draft
	02.d	26/04/2005	all	Draft
0.3	02.e	13/05/2005	most	Draft
	02.f		Tables 9,20	Draft
	02.g	29/06/2005	§3.1.1, §3.1.2.6 – §3.1.2.8, §3.1.4.6, §3.1.5.1, §3.2.2.1, §3.2.2.2, §3.2.2.4, §3.2.2.5, §3.2.3.7, §3.2.4.3, §3.2.4.4, §3.4.4.1, §3.4.4.2.	Draft
	02.h	01/07/2005	§3.2.2.3, §3.2.2.2.4, §3.2.2.2.5.7, §3.2.5	Draft
	02.i	07/07/2005	Tables 4, 25, 26, 30, §3.1.2.4, §3.1.2.5, §3.1.2.7, §3.1.2.8, §3.1.2.9.	draft
	02.j	07/07/2005	All parts	draft
	03.a	09/08/2005	All parts	draft
0.4	03.b	27/09/2005	All parts	draft
	03.c	27/10/2005	All parts	draft
	03.d	25/01/2006	All parts	draft
0.5	05	15/03/2006	All parts	draft
	0.5a	15/03/2006	All parts	draft
0.6	0.5ab	1/8/2006	All Parts	Draft for Pre-QR
1.0	1.0a	31/08/2006	All parts	Issue 1 for QR1
2.0	2.a	15/06/2007	All parts	Draft for Pre-QR2

ARGANS Version	ESL Internal Version / Rév.	Date	Pages	Changes
3.0	3.a	05/12/2007	All parts	Core-V3 FAT (Antenna retrieval updates)
3.1-Draft	3.b	10/24/2008	All parts	Changes resulting from algorithm validation activity and processor debugging. Note that this is a work-in-progress version.
3.2	3.c	04/15/2009	All pages	See ESL change note on 3.c.
3.3-Draft	3.d	03/29/2010	All pages	See ESL change note on 3.d. This version of the ATBD is work-in-progress. The working copy circulated between ESL has a number of comments indicating areas where updates are needed. Those comments are removed by Array under the assumption that they are not beneficial to the outside readers.
3.4-Draft	3.e	12/15/2010	All pages	See ESL comments in 3.e.
3.4	3.e	01/24/2011	All pages	Minor editorial corrections to the draft.
3.5-Draft	3f	06/30/2011	See ESL updates 3.f	See ESL updates for 3.f.
3.5	3f	07/28/2011	See ESL updates 3.f	Note: ATDB and DPM are aligned. Addressed comments from FAT.
3.6-Draft	3g	11/30/2011	See ESL updates	Adding description of Mironov dielectric model.
3.6	3g	12/19/2011	See ESL updates	Final release for L2SM v05.50. Addressed comments from FAT.
3.7	3h	03/01/2013	See ESL updates	Final release for L2SM v06.00. Addressed comments from FAT.
3.8	3i	11/20/2013	See ESL updates	Final release for L2SM v06.10. Addressed comments from FAT.
3.9	3j	October 24, 2014	See ESL updates	Release for L2SM v06.20.
3.10	3k	May 19, 2017	See ESL updates	Release for L2SM v06.50.

ARGANS Version	ESL Internal Version / Rév.	Date	Pages	Changes
3.11	3k	August 31, 2017	See ESL updates	Updated to address FAT comments on v3.10. Document Status Sheet: Updated to list changes including clarification that Simplified IGBP has been implemented. 1.1: Updated to 3.k. 1.2.3: Updated. Please see Document Status Sheet. 1.4.1: Added TGRD as [AD 14]. Updated document issues as needed. 1.2.3, 3.1.2.8, 3.2.4.4.4, 3.2.5.5.3: Made editorial updates. 3.2.4.4.4: Updated to clarify that DQX enhancement is not activated. 3.2.5.2.3 and 3.2.5.5.2.1: Added reference to TGRD. 3.7.2: Updated to add missing reference to IMS for snow.
4.0	4.a	22/08/2018	§3.1.2.2.3	Bircher empirical organic soil dielectric constant model added
		22/08/2018	§3.1.2.2	Weighted average of mineral – organic soil added
		22/08/2018	§3.2.4.3	Added image reconstruction error along profile (oscillation) as fixed variance on top of radiometric uncertainty at cost function level
		09/10/2018	Authors and industry	Change by adding new ESL's (WSL and FMI) and New Industry (ARGANS replacing Array Systems)
		09/09/2019	§3.1.2.2.3	Added missing absolute SM value for Bircher empirical model to comply with Mironov symmetrisation
		09/09/2019	§3.2.4.3	Emphasis that the added image reconstruction error concerns only COV_T not COV_{Prior}
		09/09/2019	§3.2.3.4.2	Text fix; when merged FNO/FFO retrieval occurs (driven by UPF TH_TAU_FN) then the UDP flag FL_DUAL_RETR_FNO_FFO is raised.

Table of Contents

List of Figures

List of Tables

1 REFERENCE INFORMATION

1.1 Identification

This document is the issue 4.0 of the Algorithm Theoretical Basis Document (ATBD) **SO-TN-ESL-SM-GS-0001-4a** for the SMOS Soil Moisture (SM) level 2 processor prototype. Updates are specified with respect to the former version 3k.

1.2 Purpose and structure

1.2.1 Purpose

This ATBD was prepared by the soil moisture ESL (current lead CBSA with University Tor Vergata, Finnish Meteorological Institute and Gamm/WSL) in view of the SMOS Level 2 Soil Moisture Prototype Processor Development (**SMPPD**) by ARGANS Ltd. (ARGANS) under contract with the European Space Agency.

According to ESA guidelines [AD 3], the purpose of ATBD is to "describe the algorithms which will produce higher level SMOS products. The document should focus on the scientific justification for the algorithms selected to derive the product, an outline of the proposed approach and a listing of the assumptions and limitations of the algorithm".

The ATBD is a detailed and extended answer to the initial SMOS Level 2 Processor High Level Requirements as defined in [AD 10]

1.2.2 Structure

The structure of this document follows closely the recommended ESA guidelines for ATBD [AD 3].

We should consider that an ATBD is both a scientific, technical and project answer to both end users and industrial requirements. ESA's point of view is confirmed by the fact that it is asked to comment upon calibration and validation issues. With both end users and industry in mind, this document describes the physical basis and approach to produce L2 products.

Accordingly, this document consists of three sections

- Section 1 (the present one) gathers reference information
- Section 2 provides background information, gives the rationale for selecting the algorithm, and presents its general layout as well as broad indications concerning limitations and output.
- Section 3 first provides the theoretical (physical) basis for SMOS measurements over land surfaces, and then gives a detailed description of the modules of the retrieval processing, a sketch of the error budget, and some practical considerations.

1.2.3 Main updates of present ATBD version

This current V4.0 version is an update of the previous V4.0 version; the main changes are:

- The introduction of the Bircher empirical organic soil dielectric constant, in order to improve Soil Moisture retrievals over organic soil and considering a linear weighting according to the organic soil fraction within the pixel. This has been also done with consideration to the Mironov symmetrisation, in order to add the missing absolute SM value for the Bircher' s model.

- We have also introduced the image reconstruction error along the TBs profile, in the form of a fixed variance on top of the radiometric uncertainty, at cost function level, and in order to improve retrievals by providing a better understanding of the existing errors, propagating them to the convergence scheme.

From previous releases:

- Version 3 with respect to version 2
  - The introduction of the experimental DQX enhancement feature as a general move for a better description of errors and uncertainty in all the algorithms compartments.

- The introduction of rescaled Chi2 using statistics over the mission is an intermediate step to a better description of the observation uncertainty.

- A safeguard is introduced to limit the values of DQX stored in current files to a configurable minimum floor.

- Introduction of the ECOCLIMAP landcover updates.

- Removal of the NPE flood rule which proved to be inadequate.

- The postprocessor update rule when a still valid in time value no longer uses DQX improvement but Chi2 improvement. DQX decrease does not mean that the retrieval solution is better; Chi2 decrease can only be used for this purpose.

- The difference between DQX values and RSTD are explicated to prepare future algorithms updates on the error budget. The RSTD is the radiometric accuracy translated into parameter uncertainty through the model sensitivity. DQX can capture more, RFI impact, model errors, image reconstruction errors ...

- V3.g introduced the symmetrisation of soil dielectric constant, HR(SM, ...) and T_{ge}(SM, ...) needed to follow the same approach. The post-retrieval considers now the absolute retrieved soil moisture to be checked against the retrieved range as the consequence of the symmetrisation makes the modelled MTB(-SM) strictly equal to MTB(SM). A negative SM is thus a mathematical acceptable solution and reported as |SM| in the User Data Product.

- As a first move of the above, C_{RFI} is no more applied to enhance the radiometric accuracy but directly used to enhance the RSTD which was our initial intent, forming the DQX values of the retrieved parameters.

- Fix of an oversight in RFI half 1st Stokes detection. The absolute value was missing in the anomaly |TBS1 - <TBS1>|.

- R_{RFI} is computed on a moving time window using the two last DGG_CURRENT_RFI at 12 days distance.

- The use and update of the "current parameters maps" which is now split per orbit pass, ascending or descending.

- The introduction of a DFFG snow map that can be used instead of ECMWF snow prediction.

- The soil properties are now provided directly on the DFFG grid.

- Major modifications to introduce the retrieval directly at the antenna reference frame. These modifications span many sections.

- Removal of aggregated observed TB X/Y to surface TB H/V, MR2$^{=1}$ (MR4$^{=1}$), COV2s (COV4s) concepts

- Introduced forward modelling up to antenna TBs by the direct use of MR2 (MR4)

- Updates of the required matrices form.

- Introduced a full polarization section on specific aspects affecting the algorithm

- Intermediary modifications on UDP/DAP, Standard mode – ESL mode

- Some minor fixes, typos, formatting, etc …

- New replaced temporary section for revisiting the UDP' modelled TBs use and definition for next updates.

- RFI screening thresholds on the TBs are now dynamic and function of the physical surface temperature, the criteria of TB rejection becomes stricter.

- The Mironov' dielectric constant model is added and the associated symmetrisation around SM=0 to prevent optimal SM retrieval to be found for unphysical too negative values. Dobson model inherited also of this symmetrisation. The choice between model type and model subtype (symmetrised or not) is configurable.

- Version 2 with respect to version 1
    - 1.2.4: A new entry is added to account for the angle bias problem.
    - 3.2.3.2.2: NPE update of DFFG cells is revisited, verified and finalized.
    - 3.2.3.2.3: NPE global winter case rewritten correctly and controlled by DGG_CURRENT_TAU_FO **(new)**

Except for the winter case where a fix is suggested but not yet implemented, the updates performed in this version transcribe the modifications carried out on the algorithm (and DPM) during the verification and validation phases.

For easier reading and referencing, we have kept, as much as possible, the same structure as in the earlier versions. However, the V3.a to V3.g releases bring many changes in the algorithm section 2. Therefore, it was not possible to keep some sections and new ones have been introduced.

The concluding section (3.7) keeps track of the future improvements, modifications or updates which are foreseen.

1.2.4   Open issues

This section is only concerned with open issues that are relevant for **implementing** the initial prototype and must definitely be closed before acceptance.

- Address cases of missing data

Every TBC annotation concerning numerical values of operational constants has been **removed**. While many of them will need tuning, note those values, when presented in the ATBD, are only illustrative: the figures to be considered are provided by the tables in TGRD, together with TBC comments when appropriate.

Several useful developments will not be accomplished in time for defining the prototype, because data are lacking, or further scientific work is needed. While decisions have been made in order to close at best the corresponding issues, it is necessary to keep them in the forefront of the scientific agenda. The most prominent are listed in a concluding section 3.7

1.3   End Users' Requirements

End users requirements are described in the Mission Requirement Document [AD 1] derived from [1]. Taking the example of soil moisture, the requirements are (see also [2-5]):

Table 1: End users' requirements

	Property	User requirement
1	Soil moisture accuracy	0.04 m^3 m^{-3} (i.e. 4% volumetric soil moisture) or better
2a	Spatial Resolution: size	The "average" dimensions of the footprint should not exceed 55 km: < TH_SIZE = 55 km
2b	Spatial Resolution: elongation	The elongation of the footprint should not exceed 1.5 < TH_ELON = 1.5
3	Global Coverage	± 80° latitude or higher
4	Revisit time	3-days max

While requirements 3 & 4 are met thanks to the mission scenario features, requirement 2 is met based on the SMOS interferometer performances [6], bounding the usable Field of view to SMOS pixel sizes and elongations. Note, that the initial 50 x 50 km^2 spatial resolution requirement has been extended to 55 x 55 km^2 to adapt to the true antenna pattern characterization that appears wider that the theoretically used one in [6].

Requirement **1** has been extensively assessed in the framework of ESA's SMOS retrieval study. Results are presented in the error budget section.

Formally and briefly, the algorithm should deliver soil moisture every 3 days max all over the globe with the nominal spatial resolution. Retrieving vegetation opacity, simultaneously to soil moisture (so-called "two-parameter retrievals"), requires a large range of incidence angles which is available only in the central part of the FOV. The estimation of vegetation opacity will thus be made with a lower repetitivity (~ every week). This is not an issue since it is expected that, except for possible effects of rainfall interception by the canopy, opacity varies slowly with time at rather coarse spatial resolution. Should for some reason the algorithm be less reliable, a flag should be issued, and when no soil moisture retrievals are possible the algorithm should nevertheless deliver either information about the equivalent dielectric constant or both dielectric constant and canopy opacity.

1.4   References

1.4.1   Applicable Documents

N°	Reference	Content	Issue
[AD 1]	EEOM-SMOS-MRD, V5	SMOS Mission Requirement Document	1.0 040701
[AD 2]	SO-RS-ESA-SYS-0555	SMOS System requirement document issue	4.1 040928
[AD 3]	SO.RS.ESA.GS.1351	Algorithm Theoretical Basis Document Guidelines	1.0 040701
[AD 4]	SO-TN-IDR-GS-0005	SMOS Level 1 and Auxiliary Data Products Specifications	5.26 130930
[AD 5]	SO-TN-IDR-GS-0006	SMOS Level 2 and Auxiliary Data Products Specifications	7.1 120520
[AD 6]	SO-DS-DME-L1PP-0006	SMOS L1 System Concept	2.9 101029
[AD 7]	SO-DS-DME-L1PP-0007	SMOS L1 Processor L0 to L1a Data Processing Model	2.17 130502

N°	Reference	Content	Issue
[AD 8]	Deleted		
[AD 9]	SO-TN-ESA-GS-1250	High level Product Definition	1.6 050929
[AD 10]	SO-TN-CBSA-GS-0003	SMOS L2 High level requirements	2.b 040720
[AD 11]	SO-TN-CBSA-GS-0015	Level 2 Soil moisture algorithm Validation Plan	1.g 061109
[AD 12]	SO-TN-CBSA-GS-0011	SMOS L2 Processor Discrete Flexible Fine Grid definition	1.c 060123
[AD 13]	SO-TN-DME-L1PP-0024	SMOS L1 Full Pol Processing	1.6 070716

1.4.2 Literature and reference documents

[1] Y. H. Kerr, "The SMOS Mission: MIRAS on RAMSES. A proposal to the call for Earth Explorer Opportunity Mission," CESBIO, Toulouse (F), proposal 30/11/1998 1998.

[2] Y. H. Kerr and P. Thibaut, "MIRAS WP 1110 Scientific Requirements," 1994.

[3] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger, "Soil Moisture Retrieval from Space: The Soil Moisture and Ocean Salinity (SMOS) Mission," *IEEE Trans.Geosci.and Remote Sens.,* vol. 39, no. 8, pp. 1729-1735, 2001.

[4] Y. H. Kerr *et al.*, "Proceedings of the Consultative Meeting on Soil Moisture and Ocean Salinity Measurement Requirements and Radiometer Techniques (SMOS)," in *Consultative Meeting on Soil Moisture and Ocean Salinity Measurement Requirements and Radiometer Techniques (SMOS)*, Noordwijk, 1995, p. 118: ESA-ESTEC.

[5] Y. H. Kerr, A. Chanzy, J. P. Wigneron, T. J. Schmugge, and L. Laguerre, "Requirements for assessing soil moisture from space in arid and semi arid areas," in *Soil Moisture and Ocean Salinity (SMOS) Measurement Requirements and Radiometer Techniques*, vol. ESA WPP-87, N. ESA, Pays-Bas, Ed., 1995, pp. 15-33.

[6] Y. H. Kerr and P. Waldteufel, "Selection of a baseline configuration for SMOS.," CESBIO, Toulouse France, NOTE 9/5/2001 2001.

[7] F. Bayle, "MIRAS demonstrator, Avignon campaign test report," ASTRIUM, Toulouse France446.RP.FB.3659.00, 14/06/2001 2001.

[8] F. Bayle *et al.*, "Two-dimensional synthetic aperture images over a land surface scene," *IEEE Trans. Geosci. Rem. Sens.,* vol. 40, no. 3, pp. 710-714, 2002.

[9] M. Berger *et al.*, "The EuroSTARRS campaign in support of the Soil Moisture and Ocean Salinity Mission.," in *IGARSS02*, Toronto Ca, 2002.

[10] A. Chanzy, Y. H. Kerr, P. J. Van Oevelen, and T. Schmugge, "Surface soil moisture maps using airborne PBMR and Portos measurements in Hapex-Sahel," in *EGS Workshop, Instrumental and Methodical Problems of Land Surface Flux Measurements*, Grenoble, France, 1994, vol. 12, p. C452.

[11] P. de Rosnay, Y. Kerr, J.-C. Calvet, F. Lemaître, M.-J. Escorihuela, and J.-P. Wigneron, "SMOSREX: Surface Monitoring Of the Soil Reservoir EXperiment," in *MICRORAD'04*, Rome (Italy), 2004.

[12] D. M. Le Vine, M. Haken, S. Bidwell, C. T. Swift, and T. J. Jackson, "ESTAR Measurements during SGP-99," in *IGARS'00*, Honolulu, Hawaii (USA), 2000: IEEE.

[13] F. Lemaître *et al.*, "Design and test of the ground based L band Radiometer for Estimating Water in Soils (LEWIS)," *IEEE Geoscience and Remote Sensing,* vol. 42, no. 8, pp. 1666-1676, August 2004 2004.

[14] T. J. Schmugge and Y. H. Kerr, "Microwave Radiometer observations in Hapex Sahel," in *6th International symposium on physical measurements and signatures in remote sensing*, Val d'Isère France, 1994, pp. 549-550.

[15] Y. H. Kerr and E. G. Njoku, "A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space," *IEEE Trans. Geosci. Remote Sens.,* vol. 28, no. 3, pp. 384-393, 1990.

[16] L. Laguerre, J.-C. Calvet, Y. H. Kerr, A. Chanzy, J. P. Wigneron, and S. Raju, "Influence of Surface Roughness on the Microwave Emission from Bare Soils for Surface Soil Moisture Retrieval Algorithms," in *PIERS'94*, Noordwijk, Netherlands, 1994.

[17] J. P. Wigneron and J. Shi, "Modelling the soil microwave emission," in *Radiative Transfer Models for Microwave Radiometry*, C. Mätzler, Ed. Stevenage, UK, 2005.

[18] J.-P. Wigneron, T. Pellarin, J.-C. Calvet, P. de Rosnay, K. Saleh, and Y. Kerr, "L-MEB: A simple model at L-band for the continental areas - Application to the simulation of a half-degree resolution and global scale data set," in *Radiative Transfer Models for Microwave Radiometry*, C. Mätzler, Ed. Stevenage, UK: Institution of Electrical Engineers, 2005.

[19] Y. H. Kerr and J. P. Wigneron, "Vegetation models and observations : A review," in *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*, B. J. Choudhury, Y. H. Kerr, E. G. Njoku, and P. Pampaloni, Eds.: VSP, The Netherlands, 1995, pp. 317-344.

[20]   J.-C. Calvet, V. Vogt, J.-P. Wigneron, C. Prigent, P. De Rosnay, and Y. Kerr, "Mesure du contenu en eau des sols dans le réseau météorologique : le projet SMOSMANIA," in *Ateliers d'Expérimentation et d'Instrumentation*, Toulouse, 2005.

[21]   Y. H. Kerr, "Microondas passivas," in *Teledeteccion*, S. J.A., Ed. Valencia (Sp): Servicio de Publicacions, Universidad de Valencia, 2001, pp. 199-239.

[22]   Y. H. Kerr and T. J. Jackson, "Working Group report on Soil Moisture," in *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*, B. J. Choudhury, Y. H. Kerr, E. G. Njoku, and P. Pampaloni, Eds.: VSP, 1995, pp. 661-667.

[23]   T. J. Schmugge and T. J. Jackson, "Mapping soil moisture with microwave radiometers," *Meteorol. Atmos. Phys,* vol. 54, pp. 213-223, 1994.

[24]   Y. H. Kerr, "Modélisation de la végétation en micro-ondes passives.," (Cepadues Editions. 1990, p.^pp. Pages.

[25]   L. Simmonds, J.-C. Calvet, J.-P. Wigneron, P. Waldteufel, and Y. Kerr, "Study on soil moisture retrieval by a future space-borne earth observation mission," University of Reading, Reading, UK, final report ESA-ITT 3552, December 2004 2004.

[26]   P. Waldteufel, C. Cot, and F. Petitcolin, "Soil moisture retrieval for the SMOS mission, Retrieval Concept Document," ACRI-ST, Sophia antipolis, France, Technical note SMOS-TN-ACR-SA-001, 25/11/2002 2002.

[27]   P. Waldteufel and J.-L. Vergely, "Soil moisture retrieval for the SMOS mission, Retrieval Algorithm Document," ACRI-ST, Sophia antipolis, Technical note SMOS-TN-ACR-SA-002, 25/03/2003 2003.

[28]   B. Berthelot, "annexe 1 statistiques sur les jeux de données ISBA " NOVELTIS, Toulouse, Final Report NOV-3050-NT-1966, 21/01/2004 2004.

[29]   B. Berthelot, "Inversion de l⁊humidité de surface en utilisant une approche neuronale," Noveltis, Toulouse, Final report NOV-3050-NT-1965, 31/01/2004 2004.

[30]   P. Richaume, Y. H. Kerr , P. Waldteufel, S. Dai, and A. Mahmoodi, "SMOS L2 Processor Discrete Flexible Fine Grid defintion," CBSA, Toulouse, Technical note SO-TN-CBSA-GS-0011-1.a, 23/01/2006 2006.

[31]   Deimos, "SMOS L1 Processor Discrete Global Grids Document," DEIMOS / ESA, Lisboa, Portugal, Technical note SMOS-DMS-TN-5200, 2004 2004.

[32]   V. Masson, J.-L. Champeau, F. Chauvin, C. Meriguet, and R. Lacaze, "A global data base of land surface parameters at 1 km resolution in meteorological and climate models," *Journal of Climate,* vol. 16, no. 9, pp. 1261-1282, 2003.

[33]   -. S.-. Martín-Neira M.: TOS-ETP/2002.106.v2/MMN, 2pp, "Strip adaptive processing for SMOS mission," ESA, Noordwijk, NL, Technical note TOS-ETP/2002.106.v2/MMN, 3-September-2002 2002.

[34]   F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing - Active and Passive*. Norwood, USA: Artech House, 1981.

[35]   C. Mätzler, "Microwave permittivity of dry sand," *IEEE Trans Antennas and Propagation,* vol. 36, no. 1, pp. 317-319, Jan 1998.

[36]   M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. Elrayes, "Microwave Dielectric Behavior of Wet Soil .2. Dielectric Mixing Models," *Ieee Transactions on Geoscience and Remote Sensing,* vol. 23, no. 1, pp. 35-46, 1985.

[37]   M. T. Hallikainen, F. T. Ulaby, M. C. Dobson, M. A. Elrayes, and L. K. Wu, "Microwave Dielectric Behavior of Wet Soil .1. Empirical-Models and Experimental-Observations," *Ieee Transactions on Geoscience and Remote Sensing,* vol. 23, no. 1, pp. 25-34, 1985.

[38]   V. L. Mironov, L. G. Kosolapova, and S. V. Fomin, "Physically and Mineralogically Based Spectroscopic Dielectric Model for Moist Soils (vol 46, pg 2059, 2009)," *Ieee Transactions on Geoscience and Remote Sensing,* vol. 47, no. 7, pp. 2085-2085, Jul 2009.

[39]   V. L. Mironov and S. V. Fomin, "Temperature Dependable Microwave Dielectric Model for Moist Soils," in *Piers 2009 Beijing: Progess in Electromagnetics Research Symposium, Proceedings I and Ii*, J. A. Kong, Ed. (Progress in Electromagnetics Research Symposium, 2009, pp. 831-835.

[40]   V. L. Mironov and S. V. Fomin, "Temperature and Mineralogy Dependable Model for Microwave Dielectric Spectra of Moist Soils," presented at the PIERS, Moscow Russia, August 18–21, 2009 2009.

[41]   A. Mialon *et al.*, "Intercomparison of Dobson and Mironov dielectric constant modelling in the context of SMOS Level 2 retrieval algorithm," Decembre 2011 2011.

[42]   M. Drusch, T. Holmes, P. de Rosnay, and G. Balsamo, "Comparing ERA-40-Based L-Band Brightness Temperatures with Skylab Observations: A Calibration/Validation Study Using the Community Microwave Emission Model," *Journal of Hydrometeorology,* vol. 10, no. 1, pp. 213-226, Feb 2009.

[43]   S. Bircher *et al.*, "Soil moisture sensor calibration for organic soil surface layers," *Geoscientific Instrumentation Methods And Data Systems,* vol. 5, no. 1, pp. 109-125, 2016.

[44] S. Bircher *et al.*, "Towards an improved soil moisture retrieval for organic-rich soils from SMOS passive microwave L-band observations, ," in *EGU Genral assembly*, Wien, Austria, 2017.

[45] N. R. Peplinski, F. T. Ulaby, and M. C. Dobson, "Dielectric Properties of Soils in the 0.3-1.3-GHz Range," *IEEE Trans Geosci. Remote Sens. ,* vol. 33, no. 3, pp. 803-807, May 1995 1995.

[46] N. R. Peplinski, F. T. Ulaby, and M. C. Dobson, "Corrections to "Dielectric Properties of Soils in the 0.3-1.3-GHz Range"," *IEEE Trans Geosci. Remote Sens.,* vol. 33, no. 6, p. 1340, November 1995 1995.

[47] G. P. de Loor, "dielectric properties of heterogeneous mixtures containing water," *J. Microwave Power,* vol. 3-2, pp. 67-73, 1968.

[48] J. K. Mitchell, *Fundamentals of soil behaviour*. New York (USA): Wiley, 1976.

[49] S. Bircher, Y. H. Kerr, and J.-P. Wigneron, "SMOSHiLat – Microwave L-band emissions from organic-rich soils in the northern cold climate zone and their impact on the SMOS soil moisture product.," CESBIO, Toulouse France2015.

[50] M. J. Escorihuela, Y. Kerr, P. de Rosnay, J.-P. Wigneron, J.-C. Calvet, and F. Lemaitre, "A Simple Modeling of a Bare Soil Emission at L- Band and Soil Moisture Retrieval on the SMOSREX site," in *Microrad'06*, San Juan Puerto Rico, 2006.

[51] M.-J. Escorihuela, Y. Kerr, J.-P. Wigneron, J.-C. Calvet, and F. Lemaître, "A Simple Model of the Bare Soil microwave emission at L- Band " *IEEE Geosci. Remote Sens.,* vol. 45, no. 7, pp. 1978-1987, 2007.

[52] T. J. Schmugge, T. J. Jackson, and H. L. McKim, "Survey of Methods for Soil-Moisture Determination," *Water Resources Research,* vol. 16, no. 6, pp. 961-979, 1980.

[53] J.-P. Wigneron, A. Chanzy, L. Laguerre, and Y. Kerr, "A simple parameterization of the L-band microwave emission from rough agricultural soils," *IEEE Trans. Geosci. Rem. Sens.,* vol. 39, no. 8, pp. 1697-1707, 2001.

[54] T. Pellarin, Y. H. Kerr, and J.-P. Wigneron, "Global simulation of brightness temperatures at 6.6 and 10.7 GHz over land based on SMMR data set analysis," *IEEE Trans Geosci.Remote Sens.,* vol. 44, no. 9, pp. 2492-2505, 2006.

[55] P. de Rosnay and J.-P. Wigneron, "Parameterizations of the effective temperature for L-band radiometry. Inter-comparison and long term validation with SMOSREX field experiment,," in *Radiative Transfer Models for Microwave Radiometry*, C. Mätzler, Ed. Stevenage, UK: Institution of Electrical Engineers, 2005.

[56] T. R. H. Holmes *et al.*, "A new parameterization of the Effective Temperature for L-band Radiometry.," *Geophys. Res. Letters,* vol. 33, p. L07405, 2006.

[57] P. de Rosnay, J.-P. Wigneron, T. Holmes, and J.-C. Calvet, "Parameterizations of the effective temperature for L-band radiometry. Inter-comparison and long term validation with SMOSREX field experiment,," in *Thermal Microwave Radiation: Application for Remote sensing*, C. Mätzler, P. W. Rosenkranz, A. Battaglia, and J.-P. Wigneron, Eds. London, UK: IEE Electromagnetic Waves series 52, 2006.

[58] J.-P. Wigneron *et al.*, "Characterizing the dependence of vegetation model parameters on crop structure, view angle and polarization at L-Band," *IEEE Trans. Geosc. Remote Sens.,* vol. 42, no. 2, pp. 416-425, 2004.

[59] T. J. Jackson and T. J. Schmugge, "Vegetation Effects on the Microwave Emission of Soils," *Remote Sensing of Environment,* vol. 36, no. 3, pp. 203-212, Jun 1991.

[60] J.-P. Wigneron, A. Chanzy, J.-C. Calvet, and N. Bruguier, "A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields," *Remote sens. environ.,* vol. 51, pp. 331-341, 1995.

[61] J.-P. Wigneron, A. Chanzy, J.-C. Calvet, A. Olioso, and Y. Kerr, "Modeling approaches to assimilating L-Band passive microwave observations over land surfaces," *J. Geophys. Res.,* vol. 107, no. D14, p. DOI10.1029/2001JD000958, 2002.

[62] K. Saleh *et al.*, "Impact of rain interception by vegetation and mulch on the L-band emission of natural grass," *Rem. Sens. Environ.,* vol. 101, no. 1, pp. 127-139, 15 march 2006 2006.

[63] K. Saleh, J.-P. Wigneron, P. de Rosnay, J.-C. Calvet, and Y. Kerr, "Semi-empirical regressions at L-Band applied to surface soil moisture retrievals over grass," *Remote Sens. Env.,* vol. 101, pp. 415-425, 2006.

[64] J.-P. Wigneron *et al.*, "L-band Microwave Emission of the Biosphere (L-MEB) Model: description andcalibration against experimental data sets over crop fields," *Rem. Sens. Environ.,* vol. 107 pp. 639–655, 2007.

[65] P. de Rosnay *et al.*, "SMOSREX: A Long Term Field Campaign Experiment for Soil Moisture and Land Surface Processes Remote Sensing," *Remote Sens. Env.,* vol. 102, pp. 377-389, 2006.

[66] J.-P. Wigneron, J.-C. Calvet, and Y. Kerr, "Monitoring water interception by crop fields from passive microwave observations," *Agricultural and Forest Meteorology,* vol. 80, pp. 177-194, 1996.

[67] D. M. Le Vine and M. A. Karam, "Dependence of attenuation in a vegetation canopy on frequency and plant water content," *IEEE Trans Antennas and Propagation,* vol. 34, no. 5, pp. 1090-1096, Sep 1996.

[68] J.-P. Wigneron *et al.*, "Soil moisture retrievals from biangular L-band passive microwave observations," *Geoscience and Remote Sensing Letters, IEEE,* vol. 1, no. 4, pp. 277-281, 2004.

[69] K. Saleh *et al.*, "Estimates of surface soil moisture under grass covers using L-band radiometry," *Rem. Sens. Environ.,* vol. 109, pp. 42-53, 2007.

[70]   J. P. Grant *et al.*, "Calibration of the L-MEB model over a coniferous and a deciduous forest," *IEEE Transactions Geoscience and remote Sensing,* submitted feb 07.

[71]   S. A. Komarov, V. N. Kleshchenko, V. V. Shcherbinin, and V. L. Mironov, "Measurement and Simulation of L-Band Emission for a Larch Forest Stand, in Proceedings," in *IGARRS'04,* , Anchorage, USA., 2004.

[72]   T. J. Schmugge, J. R. Wang, and G. Asrar, "Results from the Push Broom Microwave Radiometer Flights over the Konza Prairie in 1985," *IEEE Trans Geosci.Remote Sens.,* vol. 26, no. 5, pp. 590-597, Sep 1988.

[73]   J. R. Wang, J. C. Shiue, T. J. Schmugge, and E. T. Engman, "The L-Band Pbmr Measurements of Surface Soil-Moisture in Fife," *Ieee Transactions on Geoscience and Remote Sensing,* vol. 28, no. 5, pp. 906-914, Sep 1990.

[74]   W. Putuhena and I. Cordery, "Estimation of interception capacity of the forest floor," *Journal of Hydrology,* vol. 182, pp. 283-299, 1996.

[75]   J. P. Wigneron, Y. H. Kerr, A. Chanzy, and Y. Q. Jin, "Inversion of Surface Parameters from Passive Microwave Measurements over a Soybean Field," *Remote Sens. Envir.,* vol. 46, pp. 61-72, 1993.

[76]   J. M. Stiles and K. Sarabandi, "Electromagnetic scattering from grassland. I. A fully phase-coherent scattering model," *IEEE Trans Antennas and Propagation,* vol. 38, no. 1, pp. 339-348, 2000.

[77]   K. Saleh *et al.*, "Impact of rain interception by vegetation and litter on the L-band emission of natural grasslands (SMOSREX Experiment)." *J. Geophys. Res.,* submitted 2005.

[78]   M. Pardé, J.-P. Wigneron, A. Chanzy, P. Waldteufel, Y. Kerr, and S. Huet, "Retrieving surface soil moisture over a wheat field: comparison of different methods," *Remote Sensing of the Environment,* vol. 87, pp. 334-344, 2003.

[79]   R. H. Lang, C. Utku, P. de Matthaeis, N. Chauhan, and D. M. LeVine, "ESTAR and model brightness temperatures over forests:effects of soil moisture," in *IGARSS'01*, Sidney, Australia, 2001.

[80]   C. Matzler, "Microwave Transmissivity of a Forest Canopy - Experiments Made with a Beech," *Remote Sensing of Environment,* vol. 48, no. 2, pp. 172-180, May 1994.

[81]   P. Ferrazzoli and L. Guerriero, "Passive microwave remote sensing of forests: A model investigation," *IEEE Trans Geosci. Remote Sens.,* vol. 34, pp. 433-443, 1996.

[82]   A. Della Vecchia and P. Ferrazzoli, "A large scale approach to estimate L band emission from forest covered surfaces," Tor Vergata Univesrity, Technical note SO-TN-TV-GS-0001-01.a, 21/02/2006 2006.

[83]   P. Ferrazzoli, L. Guerriero, and J. P. Wigneron, "Simulating L-band emission of forests in view of future satellite applications," *IEEE Trans. Geosci. Remote Sens.,* vol. 40, no. 12, pp. 2692-2708, 2002.

[84]   A. Della Vecchia *et al.*, "Modeling the multi-frequency emission of broadleaf forests and their components," *IEEE Geosci. Remote Sens.,* vol. 48, pp. 260-272, 2010.

[85]   R. Rahmoune and P. Ferrazzoli, "Reprocessed forward model of L-Band emission from broadleaf forests," Tor Vergata22/04/2010 2010.

[86]   R. Rahmoune and P. Ferrazzoli, "Suggested actions for the forward model of broadleaf forests," University, Tor Vergata, ESA Technical note 04/06/2010 2010.

[87]   F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing - Active and Passive*. Norwood, USA: Artech House, 1986.

[88]   J. A. Lane and J. A. Saxton, "Dielectric dispersion in pure polar liquids at very high radio-frequencies, measurements on water, methyl and ethyl alcohols," *Proc. Royal Soc.,* vol. A213, pp. 400-408, 1952.

[89]   L. A. Klein and C. T. Swift, "An improved model for the dielectric constant of sea water at microwave frequencies," *IEEE Trans Antennas and Propagation,* vol. AP-25, pp. 104-111., 1977.

[90]   A. Stogryn, "The brightness temperature of a vertically structured medium," *Radio Sci.,* vol. 5, pp. 1397-1406, 1970.

[91]   C. Matzler, "Passive Microwave Signatures of Landscapes in Winter," *Meteorology and Atmospheric Physics,* vol. 54, no. 1-4, pp. 241-260, 1994.

[92]   M. T. Hallikainen, "Retrieval of snow water equivalent from NIMBUS-7 SMMR data - effect of land-cover categories and weather conditions," *IEEE Journal of Oceanic Engineering,* vol. 9, no. 5, pp. 372-376, 1984.

[93]   M. Schwank, M. Stähli, H. Wydler, J. Leuenberger, C. Mätzler, and H. Flühler, "Microwave L-Band emission of freezing soil," *IEEE Trans Geosci.Remote Sens.,* vol. 42, no. 6, pp. 1252-1261, June 2004 2004.

[94]   C. Matzler and U. Wegmuller, "Dielectric-Properties of Fresh-Water Ice at Microwave- Frequencies," *Journal of Physics D-Applied Physics,* vol. 20, no. 12, pp. 1623-1630, Dec 14 1987.

[95]   C. Matzler and A. Standley, "Relief effects for passive microwave remote sensing," *International Journal of Remote Sensing,* vol. 21, no. 12, pp. 2403-2412, Aug 2000.

[96]   Y. Kerr, F. Sécherre, R. Gauthier, and J.-P. Wigneron, " Analysis of topography effects.," CESBIO, Toulouse, Final report ESA ITT 3552 (CNN) December 2004 2004.

[97]   Y. H. Kerr, F. Sécherre, J. Lastenet, and J.-P. Wigneron, "SMOS: Analysis of perturbing effects over land surfaces," in *IGARSS'03*, Toulouse France, 2003: IEEE.

[98]   Y. H. Kerr, F. Sécherre, J. Lastenet, and J.-P. Wigneron, "Influence of Topography on SMOS measurements," in *Microrad'04 specialist meeting*, Rome, Italy, 2004.

[99]   Y. H. Kerr, J.-P. Wigneron, and T. Pellarin, "SMOS: Analysis Topography effects," in *IGARSS'04*, Anchorage Alaska, 2004.

[100]   L. Coret, Y. H. Kerr, P. Richaume, and J.-P. Wigneron, "Flagging the topographic impact on the SMOS signal," in *MicroRad'06*, San Juan Puerto Rico, 2006.

[101]   A. Mialon, L. Coret, Y. H. Kerr, F. Sécherre, and J.-P. Wigneron, "Flagging the topographic impact on the SMOS signal," *IEEE Geosci. Remote Sens.,* vol. 46, no. 3, pp. 689-694, 2008.

[102]   P. Waldteufel, J.-L. Vergely, and C. Cot, "A cardioid model for multi-angular radiometric observations," *IEEE Trans Geosci.Remote Sens.,* vol. 42, pp. 1059-1063, 2004.

[103]   M. Peichl, V. Wittmann, E. Anterrieu, B. Picard, N. Skou, and S. Sobjaerg, "Scientific inputs for the SMOS Level 1 Processor development," DLR, Munich, Final report for ESA contract No. 10508/02/NL/GS 2004.

[104]   H. J. Liebe, G. A. Hufford, and M. G. Cotton, " Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz," in *AGARD 52nd Specialists' Meeting of the Electromagnetic Wave Propagation Panel*, Palma de Mallorca, Spain, 1993, pp. 3-1 - 3-10.

[105]   Y. H. Kerr, "Study of African Arid Regions Using SMMR Data," Jet Propulsion Laboratory, Pasadena, CaliforniaInterim Report N° 2, May 19, 1988 1988.

[106]   R. K. Crane, "Propagation Phenomena Affecting Satellite Communication Systems Operation in the Centimeter and Millimeter Wavelength Bands," *Proceedings of the IEEE,* vol. 59, pp. 173-188, 1971.

[107]   J. W. Waters, "Absorption and Emission of Microwave Radiation by Atmospheric Gases," in *Methods of Experimental Physics*, vol. 12, Part B, Radio Astronomy, section 12, M. L. Meeks, Ed.: Academic Press, 1976.

[108]   A. Benoit, "Signal attenuation Due to Neutral Oxygen and Water Vapor, Rain and Clouds," *Microwave Journal,* vol. 1, no. 11, pp. 73-80, 1968.

[109]   D. M. LeVine and S. Abraham, "Galactic noise and passive microwave remote sensing from space at L-Band," *IEEE Trans Antennas and Propagation,* vol. 42, pp. 119-129, 2004.

[110]   P. Reich and W. Reich, "A radio continuum survey of the northern sky at 1420 MHz," *Astron. Astrophys. Suppl. Ser.,* vol. 63, pp. 205-292,, 1986.

[111]   J.-Y. Delahaye, P. Golé, and P. Waldteufel, "Calibration error of L-band sky-looking ground-based radiometers," *Radio Sci.,* vol. 37, no. 1, pp. 11/1-11/11, 2002.

[112]   D. Hartmann and W. B. Burton, *Atlas of galactic neutral hydrogen*. New York, USA: Cambridge university press, 1997, p. 235.

[113]   E. M. Arnal, E. Bajaja, J. J. Lararte, R. Morras, and W. J. L. Poppel, "A high sensitivity HI survey of the sky at δ<=− 25°," *Astron. Astrophys. Suppl. Series,* vol. 142, pp. 35-40, 2000.

[114]   ACRI-ST, "ACRI, Reqts_L2Draft-2c.doc," ACRI-ST, Sophia Antipolis, Draft Report Draft-2c.doc, 2005 2005.

[115]   DEIMOS, "Earth Explorer Mission CFI Software Mission Conventions Document.," DEIMOS/ESA, Lisboa Portugal 01/11/2008 2008.

[116]   J. P. Claassen and A. K. Fung, "The Recovery of Polarized apparent Temperature Distributions of Flat Scenes from Antenna Temperature Measurements," *IEEE Trans.Antennas Propagat.,* vol. AP-22, pp. 433-442, 1974.

[117]   P. Waldteufel and G. Caudal, "Off-axis Radiometric Measurements : Application to Interferometric Antenna Designs," *IEEE Trans. Geoscience and Remote Sensing,* vol. 40, no. 6, pp. 1435-1439, 2002.

[118]   D. M. Le Vine and S. Abraham, "The Effect of the Ionosphere on Remote Sensing of Sea Surface Salinity From Space: Absorption and Emission at L Band," *IEEE Trans Antennas and Propagation,* vol. 40, no. 4, pp. 771-782, 2002.

[119]   T. Pellarin, J.-P. Wigneron, J.-C. Calvet, and P. Waldteufel, "Global soil moisture retrieval from a synthetic L-band brightness temperature data set.," *J. Geophys. Res.,* vol. 108, no. D12, pp. 4364, doi: 10.1029/2002JD003086, 2003.

[120]   Deimos, "Computation of the Synthetic Antenna Directional Gain as interface to L2 Processor, Synthetic Antenna " DEIMOS/ESA, Lisboa, Portugal, Technical note 26 Apr 2005 2005.

[121]   E. Anterrieu *et al.*, "A strip adaptive processing approach for the SMOS space mission," in *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, 2004, vol. 3, pp. 1922-1925 vol.3.

[122]   J.-L. Vergely, "Soil moisture Retrieval Study," ACRI-ST, Sophia Antipolis, Final Report SMOS-FR-ACR-SA-007, 2005 2005.

[123]   P. Waldteufel and S. Zine, "Approximating the weighting function to be used in the SMOS L2 processor.," CBSA, ToulouseSO-TN-CBSA-GS-0010 version 1.b, 18/06/2005 2005.

[124]   J. R. Wang and B. J. Choudhury, "Remote-Sensing of Soil-Moisture Content over Bare Field at 1.4 Ghz Frequency," *Journal of Geophysical Research-Oceans and Atmospheres,* vol. 86, no. NC6, pp. 5277-5282, 1981.

[125]   D. W. Marquardt, "An algorithm for least-squares estimation of non linear parameters," *J. Soc. Ind. Appl. Math,* vol. 11, pp. 431-441, 1963.

[126]   M. Abramovitz and I. Stegun, *Handbook of Mathematical Functions* (New York, USA). Dover publications, 1964.

[127] R. L. Armstrong and M. J. Brodzik, "Recent Northern Hemisphere snow extent: a comparison of data derived from visible and microwave sensors," *Geophysical Research Letters* vol. 28, no. 19, pp. 3673-3676, 2001.
[128] A. T. C. Chang, J. L. Foster, and D. K. Hall, "Nimbus-7 derived global snow cover parameters," *Annals of Glaciology,* vol. 9, pp. 39-44, 1987.
[129] R. L. Armstrong and M. J. Brodzik, "A twenty year record of global snow cover fluctuations derived from passive microwave remote sensing data," in *5th Conference on Polar Meteorology and Oceanography*, Dallas, TX, 1999, pp. 113-117: American Meteorological Society.

Definitions, acronyms and abbreviations

1.4.3   Definitions

Table 2: Definitions

Term	Definition
Aggregated fraction	Weight to be applied to a collection of surface areas when computing the radiometric signal as a sum of contributions. The aggregation assumes that a single radiative model using average physical properties can represent radiometric signals from every elementary fraction.
Antenna best fit plane	Best fit plane to the phase centres of the LICEF elements
Apodization function (APF)	Function applied to visibilities in order to attenuate the effects of the sharp cut-off at the boundaries of the baseline domain
Area coverage	Surface area enclosed by the SMOS pixel.
Auxiliary data	Those data required by the L2 processor that are not part of SMOS data products. We differentiate two categories: fixed and evolving (or time varying).
Baseline	Physical distance between any 2 elements of the interferometer
Baseline domain	Star shaped domain covered by every baseline provided by the instrument.
Boresight	Antenna axis : angular direction perpendicular to the antenna best fit plane
Correlation products	Raw data provided by the instrument and down-linked
Current	Auxiliary data LUT which should be updated accounting for L2 processing results
Default contribution	Contribution to the radiometric signal computed with physical parameters obtained from auxiliary data only.
DFFG working area	Subset of a map on the DFFG grid, which surrounds a given DGG grid node
Director cosines	Natural reference frame at antenna level. Director cosines are $\xi = \sin(\theta)\cos(\phi)$ and $\eta = \sin(\theta)\sin(\phi),$ where θ and ϕ are here respectively the angle to antenna boresight and the azimuth in the antenna plane
DQX	Retrieval error estimate associated to each parameter product with the same unit.
Dwell line	The (not quite straight) line along the FOV on which are located views of the same area when compounding successive snapshots, for various incidence angles
Evolving, time varying	Those data, which are time dependent by nature, so they are subject to possible availability issues. For example, parameter maps coming from other EOS satellites, forecast models.
External fixed	Data subject to possible external issues (authorization rather than strictly availability). They are not expected to change in time, except for upgrade. For example, it could be land cover information, coastline boundary, IGBP maps.
Fine grid, DFG	Highest resolution grid where auxiliary surface data must be provided
Fixed Parameter reference value	Geophysical quantity obtained through pre-processing auxiliary data files in order to obtain view dependent values or initial guesses for a parameter
Flexible Fine grid, DFFG	Aggregated DFG to a variable coarser resolution where computation must be done
Forward (fwd) model	Radiative model used to compute TB from physical medium properties
Fraction	Weight to be applied to a surface area when computing the radiometric signal

Term	Definition
	as a sum of contributions
Free Parameter reference value	Geophysical quantity obtained through pre-processing auxiliary data files in order to apply decision tree and obtain initial guesses for a parameter
Homogeneous pixel	SMOS pixel which consists of a single fraction when considering the surface characteristics (including evolving features such as snow)
L1c node ; L1c pixel	A given TB data record in the L1c data product. Such a record is defined for an earth location (L1 Auxiliary Earth Grid file DGG) and contains, among others, a collection of L1c views: SMOS pixels TB and their associated geometric, radiometric and identification information.
L1c view	Data subset of the L1c pixel that makes an individual measurement. It consists in brightness temperature, flags including polarization, radiometric uncertainty, incidence angle, azimuth angle, elliptical footprint semi-axes and snapshot ID.
Local DFG grid	Subset of a map on the DFG grid, which surrounds a given DGG grid node
Localization	The determination of the exact area covered by the SMOS pixel.
Mean fraction	Aggregated fraction where the weight is computed using a mean weighting function which does not depend on the incidence angle
MEAN_WEF	Weighting function used for carrying out weighted sums over the DFFG independently of incidence angle such as the mean fractions or the free parameter reference values.
Mixed pixel	Alternate name for a non homogeneous pixel
Normal soil	Soil which has an upper layer which is able to store liquid water
Product Confidence Descriptor	Subset of processor outputs that includes indications about the quality of the product. It contains both confidence value and flags.
Product Process Descriptor	Subset of processor outputs that includes information about process options and status. A small subset is given in User Data Product, and the main part is stored in DAP for ESL analysis after launch.
Product Science Flags	Subset of processor outputs that includes information about geophysical external features
Reconstruction	Computation by the L1 processor of brightness temperatures fields from visibilities
Reference value	Value of a physical parameter used in a radiative model, obtained through averaging physical parameters provided as auxiliary parameters for an elementary area, over an aggregated fraction that aggregates the concerned elementary areas.
SMOS Field of view (FOV)	The extent of the snapshot, bounded by both aliased images and spatial resolution. The FOV may be defined in the antenna frame of reference or in a geographical system at Earth's surface level.
SMOS fixed grid, DGG	Equal surface grid, defined once and for all, on the nodes of which the soil moisture will be retrieved. The average inter-node distance is close to 15 km. For land surfaces only (including large ice covered areas), the grid should include about 6.5×10^5 nodes.
SMOS pixel	This expression refers loosely (through its 3dB contours) to the weighting function which characterizes the spatial resolution of the interferometer.
SMOS snapshot	The image reconstructed from SMOS interferometric data averaged over the elementary period. A snapshot includes one image for a given single polarization (X or Y) in dual polarization mode or two images in full polarization mode, one for a given polarization (X or Y) and one for the cross polarization (XY).
Spurious	Refers to radiometric data being contaminated by point or nearly point radio-sources, either natural (Sun) or manmade (RFI)
Topography (strong /soft)	Topography is said to be strong when the topography index is higher than a given value and soft below this value and above the flat terrain
Topography index	Index derived from digital elevation model characterizing the slope distribution of the terrain

Term	Definition
Uniform pixel	Homogeneous SMOS pixel within which every physical parameter is identical. This is a convenient concept but probably does not exist over land surfaces.
User Data Product	This is the L2 product intended for all end-users. It is organized as a list of fixed-size records that contains the L2 retrieval outputs (parameters, flags …).
User Parameter File (UPF)	This is a file described in TGRD which explicates every parameters values, either parameters in models or algorithms controls, thresholds, switches ... we want to keep user or operator configurable and thus not hardcoded. True mathematical constants, such π, are not included.
Visibilities	Data obtained from correlation products after correcting for system noise
Weighting function (WEF)	Function derived from the apodization function, to be applied to every elementary area inside the SMOS pixel in order to give the proper weight to the corresponding contribution to up-welling radiation.

1.4.4 Acronyms, abbreviations and notations

Table 3: Acronyms and Abbreviations

Acronyms & Abbreviations	Meaning
AF or EAF	Array Factor or Equivalent Array Factor
AMS	American Meteorological Society
AFP	Antenna Footprint (Weighting function bulk properties)
AOCS	Attitude and Orbit Control System
APF	Apodization function
ASL	Above Surface Layer
ASTD	A priori Standard Deviation
ATBD	Algorithm Theoretical Basis Document
BARC	Beltsville Agricultural Research Center
CBSA	CESBIO – IPSL/SA
CESBIO	Centre d'Etudes Spatiales de la Biosphère
CPC	Climate Prediction Centre
DAP	Data Analysis Product
DC	Director Cosines (CHI & ETA)
DEM	Digital Elevation Model
DFFG	Discrete Flexible Fine Grid
DFG	Discrete Fine Grid
DGG	Discrete Global Grid: the SMOS grid
DMSP	Defense Meteorological Satellite Program
DQX	Data Quality indeX
DTED	Digital Terrain Elevation Data
EAF or AF	Equivalent Array Factor or Array Factor
ECMWF	European Centre for Medium-range Weather Forecasting
ESL	Expert Support Laboratory
ESTAR	Electronically Steered Thin Array Radiometer
EUMETSAT	European organization for the exploitation of meteorological satellites
FOV	SMOS alias-free Field Of View
FWF	Fringe-Washing Factor
GHRSST	GODAE High Resolution Sea Surface Temperature
GTOPO30	Global 30 Arc Second Elevation Data
IGBP	International Geosphere-Biosphere Programme

Acronyms & Abbreviations	Meaning
I-HKTM	Instrument Housekeeping Telemetry
INRA	Institut National de la Recherche Agronomique
IPSL/SA	Institut Pierre Simon Laplace / Service d'Aéronomie
L0	SMOS Level 0 Data Products
L1a	SMOS Level 1a Data Products
L1b	SMOS Level 1b Data Products
L1c	SMOS Level 1c processor or Data Products
L2	SMOS Level 2 processor or Data Products
LAI	Leaf Area Index
LAI_max	Maximum value of the LAI over one year for a forest stand
LAT, LON	Latitude, Longitude
LICEF	Lightweight Cost Effective Front-end; the SMOS antenna-receiver element. The SMOS instrument consists of 69 LICEFs
L-M	Levenberg-Marquardt minimization algorithm
LSM	Land Sea Mask
LUT	Look-Up Table
MD; MD0	Dielectric index radiative model; default version
MD2; MD3; MD4	Dielectric index model retrieval options
MDa	Additional Dielectric index retrieval
MDd	Dielectric index model applied to inhomogeneous scenes
METOP	METeorological Operational satellite
MIRAS	Microwave Interferometric Radiometer with Aperture Synthesis
MN; MN0	Vegetated soil radiative model; default version
MN2; MN3; MN4	Vegetated soil radiative model retrieval options
MODIS	Moderate Resolution Imaging Spectroradiometer
MS	Snow cover radiative model
MW; MW0, MWS, MWF	Open water radiative model; default version, saline and fresh sub-models
MW2; MW3; MW4	Open water radiative model retrieval options
NCEP	National Centre for Environmental Prediction
NIR	Noise Injection Radiometer
NN	Neural Network
NOAA	National Oceanic and Atmospheric Administration
NPE	Non-Permanent (meteorological conditions)
NSIDC	National Snow and Ice Data Centre
PALS	Passive and Active L- and S-band (PALS) airborne microwave sensor
PBMR	Push Broom Microwave Radiometer
PCD	Product Confidence Descriptor
PPD	Product Process Descriptor
PSD	Process Science Descriptor
RFI	Radio Frequency Interference
RMSE	Root Mean-Square Error
RTE, RT	Radiative Transfer Equation, Radiative Transfer
SC-HKTM	Satellite Housekeeping Telemetry
SM	Soil volumetric Moisture content
SMOSREX	Surface Monitoring Of the Soil Reservoir EXperiment
SMPPD	Soil Moisture Prototype Processor Development
SRTM	Shuttle Radar Topography Mission
SSM/I	Special Sensor Microwave/Imager
TB	Short notation for brightness temperatures
TBC	To Be Confirmed
TBD	To Be Decided

Acronyms & Abbreviations	Meaning
TBH	Brightness Temperature for Horizontal polarization at the surface of the Earth
TBV	Brightness Temperature for Vertical polarization at the surface of the Earth
TB_X	Brightness Temperature for X axis polarization at antenna frame
TB_Y	Brightness Temperature for Y axis polarization at antenna frame
TEC	Total Electron Content
TGRD	Table Generation Requirements Document
TOA	Top Of Atmosphere
TOV	Tor Vergata University - DISP
UDP	User Data Product
UPF	User Parameter File
USDA	United States Department of Agriculture
UTC	Coordinated Universal Time
VWC	Vegetation (volumetric) Water Content
WA_{DFFG}	Matrix of DFFG cells making the Working Area
WEF	SMOS pixel WEighting Function
WMO	World Meteorology Organization

Plain notation (no fancy font, subscript …) has been systematically added even when not used in the document, allowing organizing the table in a strict alphabetic order.

Attempts to be systematic are made: C for coefficient, TH for threshold, FL for flags. For some short notations, "local use" means they are used with a definite purpose in a single subsection.

CODE values: i (L1 input), d (aux data input), u (user input), v (variable), t (theoretical description), o (output).

Table 4: Notations

CODE	Plain Notation	Notation	meaning
v	a	a	Argument of rotation matrix
v	A_card		Dielectric constant index
u	a_L , b_L		Coefficients for computing Mg_L from estimated SM
v	a_ST	a_{ST}	Intermediate function in water dielectric constant
v	A_t	A_t	Weighting coefficient for T_gc
t	A1 to A4		Brightness temperatures at antenna level (local use)
i	AF_FOV, EAF_FOV, BORDER_FOV; SUN_FOV, SUN_POINT, SUN_TAILS, SUN_GLINT_FOV, SUN_GLINT_AREA, RFI		L1c flags
v	ALPHA	α	Exponent in EPS_s (local use)
t	ALPHA_sct	α_{sct}	Coefficient for effective albedo
i	ASCENDING_FLAG		L1c SPH flag
i,o	AVG_TIME		Mean (median time) between first and last view of a node
u	B'_F , B"_F		Coefficients for computing TAU_0F
t	B_p		Coefficient for computing TAU_p (unused)
u	B_S , B"_S		Coefficients for computing TAU_S_nad
u	B_t	B_t	Coefficient for computing A_t
t	BD		Bandwidth
u	BD_S	BD_S	SMOS receiver bandwidth
u	BERE_1 to 3, BEIM_1 to 3	$\beta_{e'}, \beta_{e"}$	Components for exponent BETA
v	BETA	β	Exponent complex function in EPS_s
u	BETA_i_1 to 3		3 coefficients for computing imag(BETA)
u	BETA_r_1 to 3		3 coefficients for computing real(BETA)
v	BS_L		Dry litter biomass
u	c		Light velocity
u	C_BORDER,C_ EAF, C_SUN_TAILS, C_SUN_GLINT_AREA		Enhancing DTBa factors triggered by L1c flags
o	C_FM0, C_FM, C_FV		DAP numerical information on fractions
t	C_pol	C_{pol}	Parameter for computing TAU_V (unused)
u	C_RFI, C1_RFI, C2_RFI		Enhancing uncertainty factor and coefficients
v	C_T	C_t	Function for computing T_g
u	C_WEF_1 to C_WEF_4		WEF coefficients
u	CA_TBS1, CB_TBS1		Coefficients for Stokes 1 RFI L2 test
u	CCX, CCXi		Sensitivity function and coefficients
v	CHI	ξ	Director cosine
v	CHI', ETA'	ξ', η'	Differential director cosines
v	CHI_2	χ^2	Retrieval quality index
u	CL_P		Coefficient for litter optical thickness
v	COST		Cost function

14

CODE	Plain Notation	Notation	meaning
v	COV_Post		Retrieval parameter posterior covariance matrix
v	COV_Prior		Cost function parameter prior covariance matrix
v	COV_T		Antenna radiometric uncertainty covariance matrices
v	CP		Convective precipitation
u	CPA_1 to 3		Coefficients for EPS_pa
u	CVAL_2, CVAL_4	C_{val_2} C_{val_4}	Coefficients for computing MVAL_0
u	d, f0		Element spacing, central frequency (local use)
t	D_cp. WI-MAX . ALPHA_I		Factors to compute TAU_Ip (unused)
t	D_tau_G, ZM, D_Z		Atmospheric coefficients (local use)
u	D_TSURF; D_A_CARD, D_SM, D_TAU, D_TTH, D_(TTV/TTH), D_OMH, D_(OMV-OMH), D_HR, D_SNPAR	$\sigma0_{Tsurf...}$	ASTD (a priori standard deviations) values
v	DELTA, ALPHA	δ, α	Declination, right ascension (local use)
u	DFFG_STEP		STEP of the DFFG
u	DFG_STEP		STEP of the DFG
v	DIFF		(Data-model) difference matrices
v	DIFF_OM	$\omega_V - \omega_H$	Vegetation albedo polarization difference
U	DLCC	DLCC	Uncertainty in reference values (cover classes)
u	DP_SM, DP_A_card, DP_TAU_nad, DP_T_SURF, DP_TTH, DP_RTT, DP_OMH, DP_DIFF_OM	ΔSM etc.	Increments for derivatives
u	DPD		Increment vector for derivatives
v	DRV		Derivative matrix
v	DT_G, DT_O2, DT_H2O		Equivalent atmospheric layer temperature to surface differences
u	DTB_F		Scaling coefficient for computing MVAL_0
i	DTBa		Antenna radiometric uncertainties
t	EH, EV		Electric fields
v	EL, H		Elevation, sidereal angle (local use)
u, v	EPS', EPS"	$\varepsilon', \varepsilon''$	Dielectric constant real and imaginary part (generic notation)
u	EPS_0	ε_0	Permittivity of free space (=8.854 10^{-12} Fm^{-1})
o	EPS_D	ε_D	Dielectric constant inferred from additional MDa retrieval
u	EPS_dry-sand ,EPS_sand	$\varepsilon_{dry\text{-}sand}$, ε_{sand}	Dielectric constant for dry sand
u	EPS_frz, EPS_ice	ε_{frz}, ε_{ice}	Dielectric constants
v	EPS_fw	ε_{fw}	Dielectric constant of free water
u	EPS_pa	ε_{pa}	Dielectric content of solid particles
u	EPS_rock	ε_{rock}	Dielectric constant for barren areas
v	EPS_s, EPS_b	ε_s, ε_b	Complex dielectric constant for whole surface, smooth bare medium (additional subscript for polarization)
u	EPS_urban	ε_{urban}	Dielectric constant still missing
v	EPS_W	ε_W	Free water dielectric constant (real part)
v	EPS_wo	ε_{w0}	Static dielectric constant of water

CODE	Plain Notation	Notation	meaning
u	EPS_woo	$\varepsilon_{w\infty}$	High frequency limit of the static dielectric constant of water EPS_w0
v	ETA	η	Director cosine
u	ETA_FS	η_{FS}	Free space impedance (= 377 Ohms)
u	f		Mean SMOS frequency Hz
v	f, f0		Frequency, absorption line frequency (local use)
t	F_VOL	F_{VOL}	Vegetation volumetric fraction
u	FCV1, FCV2		Retrieval algorithm convergence criterions
v	FDE		Sum of non nominal fractions (excl water) in FM class
v	FEB, FEI, FUH, FUL, FFO; FNO, FRZ, FSN, FWL, FWP, FWS,FSI, FTI		Aggregated fractions in the FM list
v	FEB, FEI, FUL, FUH, FFO, FNO, FRZ, FSN, FTS, FTM, FWL, FWO,FS, FTI		Aggregated fractions in the FM0 list
o	FL_CE		Flag for computational exceptions (place holder)
o	FL_CURRENT_RFI, FL_CURRENT_TAU, FL_CURRENT_HR, FL_CURRENT_FLOOD		Flags driving the request for updating the RFI, TAU, HR and FLOOD maps, after processing
o	FL_DATA_MISS		Place holder
o	FL_DEW, FL_LITTER, FL_FLOOD		Scene flags
o	FL_DQX		Retrieved parameter DQX flag
o	FL_MD_A		Flag for failure of additional MDa retrieval
o	FL_MVAL0, FL_MVAL		Flags for invalid pixels
o	FL_OW, FL_OPAQ_SNOW, FL_FROST, FL_FOREST, FL_TAU_FO, FL_WETLANDS, FL_BARREN, FL_ICE, FL_URBAN,FL_SEA_ICE, FL_COAST, FL_INTERCEPT		Scene flags
o	FL_PR		Polarization index flag
o	FL_QVAL		Fit quality flag
o	FL_R2, R3, R4		Flags reporting failed retrievals
o	FL_RAIN		Rain occurrence flag
o	FL_RANGE		Retrieved parameter range flag
o	FL_RFI_PRONE		Flag for potential RFI coming from aux RFI map
o	FL_SCENE_T		Aggregated scene flag
o	FL_TB_RANGE		Flag for deleted views (out of TB range)
o	FL_TOPO		Topography flags
v	FLA		Land aggregated fraction
v	FM0_n		Mean fractions to drive the decision tree
v	FM_n		Mean fractions to compute the reference values for the free parameters of the retrieval model(s)
v	FV_n_p		Angle dependent fractions to weight the models contributions and to compute the references value for their fixed parameters
v	FRE		Fraction selected for retrieval
t	GAMMA, GAMMA_1	γ_1	Atmospheric line width parameters
v	GAMMA, GAMMA_p	γ , γ_p	Vegetation attenuation factor
t	GAMMA_ST	γ_{ST}	Electric field product to TB coefficient
v	GQX		Global quality figure

CODE	Plain Notation	Notation	meaning
i	GRID_POINT_ID, GNID		DGG identifier
i	HH, MM, SS		Hour, minutes and second (time – local use)
u,v	HR; alternate Hsoil, H_ROUGHNESS		Roughness soil dependent exponent factor
i	JD,Y, M,		Julian day, year, month (local use)
t	K_BC	k_{BC}	Boltzmann constant
t	K_ext, K_sca	$\kappa_{ext}, \kappa_{sct}$	Extinction & scattering coefficients
u	k0_DT_H2O, k1_DT_H2O, k2_DT_H2O		Coefficients for H2O layer temperature difference
u	k0_DT_O2, kT0_ DT_O2, kP0_ DT_O2, kT02_ DT_O2, kP02_ DT_O2, kT0P0_ DT_O2		Coefficients for O2 layer temperature difference
u	k0_tau_H2O, k1_tau_H2O, k2_tau_H2O		Coefficients for H2O optical thickness
u	k0_tau_O2, kT0_tau_O2, kP0_tau_O2, kT02_tau_O2, kP02_tau_O2, kT0P0_tau_O2		Coefficients for O2 optical thickness
t	KAPPA, KAPPA_OX KAPPA_r, KAPPA_22, KAPPA_H2O, KAPPA_G	κ	Atmospheric lineic absorption coefficients
u	KDIA, FDIA, KDIA_MAX		Retrieval algorithm control parameters
d	LAI_max		Climatological maximum annual LAI
u	LAMBDA	λ	SMOS mean operating wavelength
i,o	LAT, LON(G)		Latitude, longitude
v	LH		Layer height (local use)
V	LSCP		Large scale precipitation
v	LWC		Litter water content
v	M_AVA0, M_AVA		Initial and validated number of L1c views
v	M_card		Cardioid model intermediate function
u	MEAN_WEF		Mean weighting function
v	Mg_L		Litter moisture content for estimating LWC
i	MODE		Operating mode
v	MR4, MR2		Rotation matrices
u	MU_s	μ_s	Soil magnetic permeability
u	MU_w	μ_w	Water magnetic permeability
v	MVAL		Fraction weighted validation index
v	MVAL_0		Initial validation index
o	N_CLEANED		Counter for outliers removed
o	N_RET		Number of times the retrieval has been ran
o	N_RFI		Number of cases with detected RFI
o	N_SKY		View counter for strong galactic source
v	N_SNAP		Total current number of snapshots
o	N_WILD		Counter for persisting outliers
u	NB_TH_DEC		Number of decision tree stage 1 thresholds
v	NF		Number of fractions in SMOS pixel
v	NFD		Number of degrees of freedom
v	NIT		Number of iterations needed for convergence
u	NITM, FCOND		Retrieval algorithm limit parameters
u	NR_p, NR_H, NR_V	$NR_p, \dots$	Roughness theta exponent factor
v	NT, NP		Number of valid data & free parameters
t	NU, NU_0	ν, ν_0	Sky radiation theory: line frequencies
u	OM_F	ω_F	Forest albedo
u	OMEGA, OM_H, OM_V	ω	Albedos
u	OMEGA_E,	Ω_E	Earth rotation rate
u	OW_01 to OW_32		Coefficients for static water dielectric constant EPS_W0
t	P, P_int		Sky radiation theory (Power) (local use)

CODE	Plain Notation	Notation	meaning
t	P, T		Atmospheric pressure & temperature (local use)
t	P_i	p_i	Free parameter
t	P_i_0, SIGMA_i_0	p_{i0}, σ_{i0}	Free parameter first guess & ASTD
d	P0	hPa	Surface pressure
t	PHI	Φ	Astronomical azimuth
v	Pint		Integrated power
v	Plobe		Normalized antenna power pattern
o	PR, PR_INDEX		Polarization ratio, polarization ratio index
u	PR_INCI		Angle for computing PR_INDEX
t	PSI, OM_Fa	ψ, ω_{Fa}	Claassen angle, Faraday angle
u	QR		Roughness polarization coupling coefficient
t	r		Fringe-wash factor in AF equation (local use)
t	R_bH, R_bV, R_bp	r_{bp}	Smooth bare soil reflectivities
t	R_E, H_rad		Earth radius, spacecraft altitude (local use)
t	R_gp, E_gp	r_{gp}, e_{gp}	Rough soil reflectivity, emissivity
v	R_RFI		RFI statistics
t	R_sp, E_sp	r_{sp}, e_{sp}	Reflectivity, emissivity including above surface layer
v	R_TAU		Initial TAU_nad value
t	R1, R2		Theta & polarization dependent terms in RTE equation
	RATIO_AVA		Percentage of valid L1c views
u	RO b, RO s	ρ_b, ρ_s	Soil dry bulk and soil particles densities
v	RO_DC	ρ_{DC}	Distance in director cosine frame
t	RO_V	ρ_V	Atmospheric water vapour density
v	RSTD		A posteriori (retrieval) standard deviations
v	RTAU_W	$r\tau_W$	Relaxation time of water
v	RTT	Rtt	$R_{tt} = tt_V / tt_H$
i	S, C		Sand & clay fractional soil content
o	S_TREE_1, S_TREE_2		Status descriptor elements for retrieval
u	SAL		Salinity of water in soil
u	SGEF_1 to 4		4 coefficients for computing SIG_eff
t	SIG_0	σ_0	Bistatic reflection coefficient
v	SIG_eff	σ_{eff}	Function in EPS_fw
d, v, o	SM		Soil Moisture
u	SM_min, SM_max; same suffixes for A_card, T_SURF, HR, TT_H, RTT, OMH, DIFF_OM		Acceptable ranges for retrieved parameters
v	SST, SSS		Sea surface temperature & salinity (Local short: T, S)
t	ST1, ST2, ST3, ST4		Stokes parameters
v	T_c	T_c	Physical vegetation (canopy) temperature
v	T_g	T_g	Soil effective surface-deep physical temperature
v	T_gc	T_{gc}	Effective soil-vegetation composite temperature
v	T_SURF		Surface temperature for retrieval
v	T_s	T_s	Effective composite temperature
v	T_SNOW		Physical snow temperature
d	T_soil_depth , T_soil_surf		Soil physical temperatures
d	T0		Surface air temperature

CODE	Plain Notation	Notation	meaning
v	TAU		Short for TAU_nad
v	TAU _O2, TAU_H2O, TAU_G	$\tau_{O2,H2O,G}$	Atmospheric nadir optical thicknesses
t	TAU_atm	τ_{atm}	Mean atmospheric nadir opacity
t	TAU_atu, TAU_atd	τ_{atu}, τ_{atd}	Atmospheric nadir opacity up and down
t	TAU_c	τ_v	Vegetation (canopy) opacity
v	TAU_FNAD	τ_{FNAD}	Forest nadir optical depth
v	TAU_Ip	τ_{Ip}	Interception TAU_p component
v	TAU_Lp	τ_L	Litter vegetation TAU_p component
v	TAU_nad	τ_{NAD}	Nadir optical depth
v	TAU_p	τ_p	Modified nadir optical depth
v	TAU_S_nad	τ_{S_NAD}	Standing vegetation nadir optical depth
v	TAU_Sp	τ_{Sp}	Standing vegetation TAU_p component
v	TB		Brightness temperature
o	TB_ASL_THETA	$TB_{\theta°}^{ASL}$	Corrected surface (Above Surface Level) simulated TB field at incidence angle theta
t	TB_atm	TB_{atm}	Mean atmospheric radiation
t	TB_atu, TB_atd	TB_{atu}, TB_{atd}	Atmospheric radiation up and down
v	TB_O2, TB_H2O, TB_G	TB_{O2}, TB_{H2O}, TB_G	Atmospheric radiative contributions
v	TB_sk, TB_sky	TB_{sk} TB_{sky}	Sky radiation
o	TB_TOA_THETA	$TB_{\theta°}^{ASL}$	Top of Atmosphere simulated TB field at incidence angle theta
t	TBH, TBV, TB3, TB4		Surface Brightness temperatures
v	TBM, TBF		Measured and simulated TB values
v	TBS1; <TBS1>		Halved ST1 view parameter; mean value
i	TB_X, TB_Y, TB_{XY}		Antenna Brightness temperatures
u	TB_X_MIN, TB_X_MAX, TB_Y_MIN, TB_Y_MAX, $TB_{XX}\_RE\_MIN$, $TB_{XX}\_RE\_MAX$, $TB_{XX}\_IM\_MIN$, $TB_{XX}\_IM\_MAX$, $TB_{YY}\_RE\_MIN$, $TB_{YY}\_RE\_MAX$, $TB_{YY}\_IM\_MIN$, $TB_{YY}\_IM\_MAX$, $TB_{XY}\_RE\_MIN$, $TB_{XY}\_RE\_MAX$, $TB_{XY}\_IM\_MIN$, $TB_{XY}\_IM\_MAX$,		Ranges for antenna TB
i	TEC	TEC_n	Vertical total electron content
u	TH_23, TH_34		Thresholds on a priori TAU for decision tree stage 2
u	TH_AVA_Min		Low threshold on minimum number of paired views for half 1st Stokes L2 RFI filtering
u	TH_CHI_2		Thresholds for setting retrieval quality flag
u	TH_DQX_SM; same for A_card, TSURF, HR, TT_H, RTT, OMH, DIFF_OM		Thresholds for acceptable DQX on retrieved parameters
u	TH_EB, TH_EI, TH_UL, TH_UL, TH_F2, TH_NO, TH_R1, TH_R2, TH_S1M, TH_S1W, TH_S2M, TH_S2W, TH_TM, TH_TS, TH_W1, TH_W2, TH_WL		Decision tree stage 1 thresholds
u	TH_EB_D, TH_EI_D, TH_UL_D, TH_UL,_D, TH_F2_D, TH_NO_D, TH_R1_D, TH_R2_D, TH_S1M_D, TH_S1W_D, TH_S2M_D, TH_S2W_D, TH_TM_D, TH_TS_D,		Key for ratio denominator when applying decision tree stage 1 thresholds

CODE	Plain Notation	Notation	meaning
	TH_W1_D, TH_W2_D, TH_WL_D		
u	TH_EB_N, TH_EI_N, TH_UL_N, TH_UL_N, TH_F2_N, TH_NO_N, TH_R1_N, TH_R2_N, TH_S1M_N, TH_S1W_N, TH_S2M_N, TH_S2W_N, TH_TM_N, TH_TS_N, TH_W1_N, TH_W2_N, TH_WL_N		Aggregation key relevant for decision tree stage 1 thresholds
u	TH_EB_R, TH_EI_R, TH_UL_R, TH_UL_R, TH_F2_R, TH_NO_R, TH_R1_R, TH_R2_R, TH_S1M_R, TH_S1W_D, TH_S2M_R, TH_S2W_R, TH_TM_R, TH_TS_R, TH_W1_R, TH_W2_R, TH_WL_R		Rank of decision tree stage 1 thresholds (defines order of branches)
v	TH_FF		Decision tree stage 1 threshold computed from TH_F1
u	TH_FIT		Threshold coefficient for repeating retrieval
U	TH_FLOOD		Threshold at which rain amounts raise the flood flag
u	TH_INDS, TH_INDM		Thresholds for topography index
u	TH_CUR_HR_VAL_PERIOD, TH_CUR_TAU_NAD_LV_VAL_PERIOD		Maximum delays for using current maps
u	TH_MMIN0		Low threshold for L1c pixel initial validity index
u	TH_MMIN1, TH_MMIN2, TH_MMIN3		Thresholds for selecting retrieval conditions
u	TH_MVAL0		Threshold of minimum MVAL0 to grant current update flags FL_CURRENT_TLV, FL_CURRENT_TFO and FL_CURRENT_HR to be possibly raised.
u	TH_PR, TH_INTERCEP		Polarization index and interception flags threshold
u	TH_RFI_ST4		Threshold for RFI using stokes 4
u	TH_SAND,TH_SEA-ICE		SAND and sea ice thresholds
u	TH_SIZE, TH_ELON		Spatial requirement thresholds
u	TH_TAU_F1		Threshold for validating TH_FF from TH_F1
	TH_TAU_FN		Threshold on TAU_FNAD for SM retrieval
u	TH_TDRY, TH_TWET		Thresholds on T_g for categorizing snow
i	THETA	θ	Incidence angle
t	THETA, PHI_I	θ, ϕ_i	Incidence, azimuth angles (local use)
t	THETA_a, PHI_a,	θ_a, ϕ_a	Polar angles (antenna frames)
u	THETA_B	θ_b	Incidence angle for computing $TB_{\theta°}^{ASL}$
t	THETA_g, PHI_g	θ_g, ϕ_g	Polar angles (geographical frames)
v	THETA_L, THETA_G0	Θ_L, Θ_{GO}	Sidereal time at point and at Greenwich (local use)
u	TILT		SMOS antenna plane tilting angle (unused)
t	Tsky_refl, Tsky_refl_lobe		Elementary, integrated received sky contribution
u	TT_V , TT_H	tt_V, tt_H	Coefficients for computing the TAU_Sp
t	u, v		Baseline coordinates in frequency domain
u	U_card, B_card		Cardioid model coefficients
v	U_T	U_T	Universal time
o	VRES		DAP numerical information on residuals
t	VWC		Parameter for computing TAU_p (unused)
t	W(u,v)		Apodization function

CODE	Plain Notation	Notation	meaning
u	W_0, b_w0	w0, b$_{W0}$	Coefficients for computing C_T
t	WEF		Weighting function
t	WEF_A	WEF$_A$	WEF approximation
u	WEF_SIZE		Size (km) of the DFFG working area (WA$_{DFFG}$)
u	WVC		Water vapour atmospheric content
v,o	X_SWATH		Dwell line abscissa

2 ALGORITHM OVERVIEW

2.1 Background information

Passive microwave radiometry has been used for some years at ground level, airborne and spaceborne experiments. At high frequencies, it has now reached a significant level of maturity, especially for atmospheric retrievals.

At low frequency and especially at L-band, most of the background lies with ground experiments (BARC, PORTOS 91, 93, PAMIR, MIRAS 99, Avignon 01, SMOSREX, Bordeaux 04,…) with a few spaceborne campaigns, mainly in the US (using PBMR, ESTAR, PALS) such as SGPnn, HAPEX SAHEL, SMEXmm, Eurostarrs[7-14]. From these experiments, models representing emission from soil and vegetation were elaborated and somewhat validated [4, 15-19]. There is thus now a consensus on the models and limitations, although a certain level of empiricism in the different approaches is still present [20-24].

The step to SMOS data is, however, still significant. The challenge will be mainly with large pixels including a variety of targets (water, crops, fallow layer urban/roads etc mixed) with potential caveats, not always well understood and /or modelled (RFI, topography…). Finally, it should also be said that, in many field experiments, the targets were rather pure, which hardly happens over land surfaces in real life. For instance, under natural vegetation, a layer of litter (dead matter) may develop, giving way to very specific signals as a function of the litter moisture content.

Such factors imply a good part of humility as to the validity range of existing algorithms as they were very often developed and tested in specific conditions.

Currently known facts are as follows:

- Retrieval of soil moisture over bare soil with low vegetation should be easy **but,**

- Snow is a very tricky target, as snow conditions may evolve very quickly with drastic changes in the signal

- Bare dry soil has behaviour that is not well understood/modelled

- Frozen soil behaves as dry soil

- Forest emission and attenuation are mostly correlated with "branch" water content (not the "leaves" water content). Under dense forests, practically no SM can be retrieved or with a so large uncertainty that the value would be useless.

- Urban areas are yet to be modelled

- Water bodies will have to be taken into account, including seasonal effects and fractional coverage

- Topography will reduce signal quality until no retrieval is possible

- Litter, when substantial, can appreciably modify soil emission

- Surface roughness at SMOS scale appears to be relatively small and a function of soil moisture

- Sun-glint might not be negligible

- …

It may be noted that the list consists mainly of limitations.

2.2 Selected approach

The basis for the approach taken here lies with the results of an ESA Study on soil moisture retrieval for SMOS [25-27]. The principle is to find the best-suited set of soil moisture (SM) and vegetation characteristics by minimizing the differences between modelled direct and measured brightness temperature (TB) data. Other potential methods could have been:

- Direct retrieval. However, direct retrieval is not feasible because the relationship between SM and TB is not unique. Moreover, direct retrieval would not allow accounting for the heterogeneous characteristics of the pixels.

- Empirical / statistical approaches (see [25-27]) where a regression is built between SM and TBs.

- Neural network approaches (see [25, 28, 29]).

- The main issue with statistical and neural network approaches is that in the SMOS case it will require measurements and can only be implemented sometime after launch. A simple inter-comparison table is presented below (Table 5).

Table 5: Statistical modelling vs. Physical modelling

Method	Advantages	Disadvantages
Empirical statistical	Quickness Robustness Simplicity	Opaque Need a learning data base every time it is upgraded Requires real data (hence after launch in our case) Clumsy for variable range of incidence angles (i.e. SMOS conditions) Limited validity range/area depending on training areas and conditions
Iterative using forward physical models	Close to the physics Easy to upgrade Provide theoretical uncertainty	Heavy Strong demand on auxiliary data Limited by the availability of reliable direct models!

We understand that ESA might want to have all the placeholders defined so that sometime after launch (at least 3 months after the end of the commissioning phase), a statistical / NN approach might be implemented.

It is however clear that the efficiency of the statistical approach will depend on available reliable data, which is per se a challenge. The baseline is thus an **iterative approach**.

Dual Pol vs. Full Pol

At the onset of this study and even this project, so as to have a pragmatic approach, the baseline was to rely solely on the dual polarization mode (i.e. H and V polarizations at the Earth surface). And this for two main reasons: the first one being that all models are rather well-defined dual pol but the behaviour for Stokes 3 and 4 is not so well known. The second reason was that full polarization mode, by reducing the integration time, decreases the sensitivity. Incidentally, the full polarization also "generates" more data.

However, in dual polarization the transfer from antenna to surface and vice versa causes ambiguities degrading significantly the number of useful views away from the track and hence retrieval quality/efficiency. With the full polarization mode, the gain obtained here could very well counterbalance other drawbacks.

Not having any conclusive elements enabling us to make a decision between dual and full pol, it will have to wait for the commissioning phase before any choice can be made based on "real" data. In this ATBD however, both options are tentatively addressed, clearly indicating the advantages of any option when it is not self-evident.

2.3 General Overview

2.3.1 Algorithm overview – a tentative layman description

In the iterative approach, one essentially aims at minimizing a **cost function** through minimizing the sum of squared weighted differences between measured and modelled brightness temperature (TB) data, for a variety of incidence angles. This is achieved by finding the best-suited set of the parameters, which drive the direct TB model, e.g. soil moisture (SM), and vegetation characteristics.

Despite the simplicity of this principle, the main reason for the complexity of the algorithm is that SMOS "pixels" which contribute to the radiometric signal are rather large areas, and therefore strongly heterogeneous. Moreover, the exact description of pixels is given by a **weighting function** which expresses the directional pattern of the SMOS interferometric radiometer and depends on incidence angle.

The goal is to retrieve soil moisture over fairly large and thus heterogeneous pixels. The retrieval is carried out at the **nodes** of a fixed Earth surface grid.

The first step will be to assess the input data quality (at each node) and filter out all unwanted data (outside the spatial mask requirement, L1c data quality flags etc).

Auxiliary data including time varying data and data having an impact on the SMOS products (meteorological data, vegetation opacity) are then ingested.

Afterwards, the retrieval process per se can be initiated. This cannot be done blindly as the direct model will be dependent upon surface characteristics (snow is different from vegetated soil and water for instance). It is thus necessary to first assess what the dominant[1] land use of a node is. For this an average weighting function (MEAN_WEF) which takes into account the "antenna" pattern is run over the high resolution land use map to assess the dominant cover type. This is used to drive the **decision tree**, which step by step, selects the type of model to be used as per surface conditions.

Obviously, over any pixel, the variety of land use type will be rather large, and it is not realistic to hope to retrieve everything. Some assumptions have to be made. It is thus considered that the node is divided generally in two areas, one where the retrieval will take place and one where the contributions to the overall node signal will be estimated. This latter part is then considered as fixed (**default contributions**) and the retrieval is made on the remaining – dominant – area. For instance, if there is an area of low vegetation with a dense forest and a lake, we will estimate the contribution of the lake and that of the forest using either external data or predetermined values of the surface characteristics: the **reference values**. This default contribution will be assumed constant in the modelled signal, and the retrieval adjustment performed on the remaining part.

For the main part of the node, as it is not possible to infer all the model parameters, reference values are also used, either for setting the models parameters, which are not retrieved, or as first guess values for the retrieved parameters.

On a node, as said above, a very large variety of surfaces may happen to be present; for instance, wheat, maize and sorghum, deciduous and coniferous forest. These land use classes can be grouped as elements having the same overall characteristics and behaviour at L-band. To have manageable items, the classes are thus **aggregated** into a small number (about 10) of generic classes having the same modelling characteristics and similar parameters. It is over this aggregated (and variable) area that the parameters are averaged to produce the reference parameters/values.

At this level, two options are possible: for each generic class of interest for a node, we can

- Either compute for each element its brightness temperature and produce the k class contribution (i.e. take for the forest class all the different types of forest available on the node and for the view and for each compute TB_k using the fine classification reference value)
- Or derive an average set of reference values for each generic class of interest in the node and for the view (i.e. estimate the "average "reference value for – say- all the forest types available using the characteristics of each class of the fine classification

The first approach is the only valid from the radiometric point of view. It is however almost intractable in our case and we have to use the second approach. This is not as bad as it seems however knowing that: i) the elementary reference values within a

---

[1] Dominant for the well-behaved node (i.e., with normal land use). When the majority of the surface is occupied by a target of no direct interest for soil moisture (e.g., water), "dominant" applies to the complementary part of the node.

generic class and a given region are sufficiently close to make errors negligible in most cases; ii) the assumptions and values obtained for the fine classes are very often coarse (if not arbitrarily the same).

In consequence, the second approach is considered thereafter; there will be a single "forest" fraction, a single "low vegetation" fraction, a single "bareland" fraction, etc ... in the SMOS pixel for a given view.

For any given node, there will be a varying number of views and each view will have different spatial extent (geometrical effects) meaning that the weighing function will not always cover the same points and thus have different land use characteristics: for instance a forest on the border might appear in some views and not in others (i.e. the "forest" fraction will depend on incidence angle). The issue of fractions in the retrieval algorithm is detailed further below in 2.3.2.

All this being said and done, the retrieval procedure starts if all the conditions are satisfied, ideally to retrieve 3 free parameters or more over the dominant class (the so-called rich retrieval). If the algorithm does not converge satisfactorily, a new trial is made with less free parameters ("poorer retrieval") until either result are satisfactory or the algorithms are considered to fail.

In all the above it was assumed that the dominant class was a nice area (i.e., gently rolling hills of green pasture) which is not always the case. The soil can be frozen, or covered with snow or rocks, there might be water only (node within a lake) we might be over a large urban area, or in the Himalayas, the apparently green surface might be a rice field or marshes … etc.

For each of these strange cases different approaches are proposed. The first distinction is related to the fact that the exotic characteristic is complementary to the rest (i.e., water body) or superimposed (i.e., topography) as both cases cannot be approached in the same manner. The second point is linked to the availability of a direct model, which is not always the case.

Finally, it was decided that, when it is not possible to retrieve soil moisture, we could at least retrieve a dielectric constant parameter (using the so-called cardioid approach). In addition, once the retrieval has converged, the brightness temperature could be computed at a given preset angle (e.g. 42.5° to compare with L1c browse products) using the selected forward models applied to the set of parameters obtained at the end of the retrieval process.

Therefore, the output product of the level 2-soil moisture algorithm should include node position, soil moisture, dielectric constants, computed brightness temperature at 42.5°, flags and quality indices.

2.3.2   More about fractions

The signal collected by the SMOS radiometer is generated by the area" illuminated" by the antenna directional gain pattern, which is characterized by a **weighting function** at surface level. The signal is thus essentially the sum of fractional signals, i.e. weighted sum of radiations upwelling from each **elementary** area.

We consider here as an elementary area (about $4{\times}4$ km^2) the pixel of the **DFFG** (Discrete Flexible Fine Grid, see next section for a quick overview). Even with 16 km^2 pixels, this may induce strong heterogeneity at the scale of the SMOS pixel, around 40 km diameter on average (1260 km^2).

This pixel contains a list of aggregated classes that result from the aggregation of a higher resolution (referred to as Discrete Fine Grid DFG) land cover classification. While this DFG land cover classification features more than 200 classes (to which are to be added non-permanent frost or snow conditions and presence of strong topography), it is possible to **aggregate** them together in about 10 aggregated classes, in such a way that non-homogeneities within each aggregated class are considerably reduced, from the viewpoint of the radiated signal.

Each aggregated class, stored in the DFFG pixel, is associated with sub pixel features: its **geometric surface fraction** (with respect to the whole pixel surface) and the most representative high-resolution land cover class among all the aggregated ones.

Therefore, in the SMOS L2 SM processor, surface areas are represented as **aggregated** (over DFFG) **fractions** for **aggregated** (over land cover) **classes.**

For each aggregated fraction, it is possible to apply **specific radiative models**, in such a way that the radiometric contribution depends on physical parameters that are characteristic of the aggregated fraction. We want to retrieve some of these parameters, and specifically of course soil moisture. However, for some aggregated fractions, there is no soil moisture to retrieve (for example open water or completely barren soil). For this and other reasons, in the general case, the retrieval will not concern **every** aggregated fraction. Some contributions to the signal will be assumed to be given by auxiliary data. They are called **default contributions**.

In order to compute default contributions, it is necessary to compute (through a weighted average) the aggregated physical parameters relevant for each concerned aggregated fraction. These aggregated parameters are called **reference values**. Even for the fractions (usually a single one) where a physical parameter is retrieved, reference values are needed, both for the parameters that are not retrieved and for those that are retrieved, since the algorithm then needs initial guess values.

Weighting functions, which represent the SMOS synthetic antenna directional gain, are used to estimate the presence and magnitude of aggregated fractions. Based on the population of "true" weighting functions, which depend on incidence angle, a **mean weighting function** (called **MEAN_WEF**) is thus built for these purposes. The resulting aggregated fractions belong to a list of **FM** values.

During the retrieval however, it has been found that the incidence angle dependence could not be neglected. Therefore, the radiative contributions are computed using the true weighting functions **WEF** (see the flow chart, iterative retrieval loop). The corresponding aggregated fractions **FV** depend on incidence angles.

Either the MEAN_WEF, or the incidence angle dependent WEF whenever necessary, are used to compute the contribution fractions of the different aggregated classes of surface and their associated reference values.

It will be seen that, as a final complication, the aggregated fractions needed to select which fraction(s) is (are) considered for retrieval, using which direct model and for which parameters (i.e. drive the decision tree), are not exactly the same as those discussed so far. This is because, for example, while such decisions depend on the amount of mountainous terrain, there is no direct model for mountainous terrain so far. Therefore, it is necessary to define a list of non-incidence angle dependent **FM$_0$** values, which are obtained using the MEAN_WEF, but differ slightly from the FM values.

2.3.3 Introducing the SMOS L2 SM grids

The SMOS L2 SM [30] processor has to manipulate several gridded data on different formats at different scales. However, the approach and algorithms presented in this document are generic and they require in reality only two gridding systems to be defined.

2.3.3.1 The Discrete Global Grid

The first grid is associated with the measured brightness temperatures given at each node and comes out to naturally be the DGG (Discrete Global Grid) **used by L1 processor** and where the L1c products are defined. This gridding system is the ISEA-4-9 that paves the Earth surface with quasi equal-area cells and minimal distortion all around the globe; the inter-node distance is practically the same everywhere on the globe and is about 15 km. Consequently, this grid is irregular for both longitude and latitude. For more details see [31].

The full Earth is covered with $\approx 2.6 \times 10^6$ DGG nodes; only $\approx 7 \times 10^5$ are relevant for SM. An L1c product contains the measurements for a SMOS half orbit swath that represents, at most, ≈ 150000 DGG nodes. Again, for SM, it will be much less depending on the swath position on Earth (even only a very few above oceans e.g. Pacific)

This DGG grid defines the L2 processing grid; the ATBD's algorithms are meant to be applied at each DGG node for producing the L2 products.

2.3.3.2 The Discrete Flexible Fine Grid

The second grid system is called the DFFG (Discrete Flexible Fine Grid) [AD 12]. This DFFG defines an almost equal-area grid system close to the reduced Gaussian ECMWF standard. The description of the grid property is very similar to the way **reduced** Gaussian grids are described in the World Meteorological Organization (WMO) GRIB specification, though lightened to satisfy only our need and, in our case, with a uniform sampling of latitude instead of a Gaussian one.

The DFFG samples meridians and parallels with an almost equal distance increment equal to DFFG_STEP_KM. Consequently, this grid is regular in latitude since the number of latitude samples along any meridian is always the same but is irregular in longitude since the number of longitudes samples decreases with the latitude. For more details see [30]

This DFFG is called "flexible" for the two following reasons:

DFFG_STEP_KM can be set to any resolution. This information is included within the format file and thus can be handled by the L2 processor without modifying the code. The purpose of this flexibility is to operate a trade-off between accuracy and computation time in order to obtain a workable solution on current computer; more powerful the computer, higher the resolution and accuracy.

Preliminary assessments indicate that the computational power of currently available computers requires DFFG_STEP_KM = 4*DFFG_STEP_KM_MIN, where DFFG_STEP_KM_MIN ≈ 0.927 km and corresponds to a 30" equatorial arc length and is linked with the resolution of [32] ECOCLIMAP 2004 which was our reference landcover. With the resolution increase of new landcovers (such as MODIS IGBP, MERIS GLOBECOVER or future SENTINEL derived landcovers) or other ADFs this might be reassessed. Note that, choosing DFFG_STEP_KM as an integer multiple of DFFG_STEP_KM_MIN is just convenient, but absolutely not mandatory.

Indeed, L2 algorithms require many computations to be done for every DFFG cells contained in a square DFFG **working area** (WA_{DFFG}) that surrounds each DGG node times the number of nodes to process. The dimension of WA_{DFFG} is 123 km × 123 km. At the full DFFG resolution (≈1km) WA_{DFFG} contains $123^2 \approx 15000$ cells, which is not much compatible with the power of current computers, especially in labs. At ≈4 km of resolution, WA_{DFFG} contains "only" $32x32 \approx 1000$ cells, which appears to be tractable. In the future, we will benefit from the flexibility of the DFFG and increase its resolution as computers power increases.

The DFFG acts as a fixed interface between the processor and the huge diversity of auxiliary data. Thus, the processor will benefit from any improvements, or changes, on those data, better coverage, better resolution … with no modifications of the processor code.

Since the beginning of the algorithm development, we used the auxiliary data we found to be the best suited for the SMOS L2 processor. For example, we chose ECOCLIMAP as our reference landcover both for its fine resolution and for the richness of its code-set (218 ecosystems). to specify our algorithms. In the most recent implementation of the algorithms, the ECOCLIMAP dataset has been replaced with a simplified version of IGBP to build the DGG_INFO ADFs (see the TGRD for details) with no change of the algorithms.

Indeed, one important characteristic of the algorithms is that they do not rely or weakly rely on specific datasets which are described in TGRD with the best set we can use at a particular moment. The future may offer new opportunities, with better auxiliary data that can be ingested within the DFFG without any changes concerning the algorithms and thus the processor.

The DFFG defines the true L2 working grid; it provides the support to hold the high-resolution Earth surface properties that algorithms need.

2.3.4 Simplified flow chart

The following table summarizes a general view of the flow for the algorithm. This presentation relies heavily on the discussion conducted with the contractor concerning the **data processing chain**. In order to improve the readability of ATBD, relevant sections are indicated.

Table 6: Simplified flow chart

ATBD reference	over… Half orbit	L1c pixel	L1c view	Action	inputs	outputs
2.3.4				**General layout (this chart)**		
	$			ingest L1c	L1c	
	$			Ingest static aux data		
3.2.2.2.1	$			Pre-process static aux data	TGRD	
3.2.2.1.3	$		$	Obtain and pre-process angles for DGG	L1c angles	
3.2.2.1.4	$		$	apply spatial resolution requirements	L1c FOOTPRINT	
3.2.2.1.5	$		$	Filter L1c views	L1c FLAGS	update RFI map
3.2.2.1.6	$		$	Enhance radiometric uncertainties	L1c FLAGS, L2 current RFI map	
3.2.2.1.7	$	$		Filter L1c pixel		
3.2.2.2.2	$	$		ingest time dependent aux data	ECMWF, tau, HR, RFI, LAI	
		$		select DFFG area for given DGG node		
3.2.2.3		$		pre-process ECMWF data to DFFG		flags NPE snow, frost
		$		pre-process ECMWF atmospheric data		rain flag
		$		Any pre-processing other auxiliary data		
3.2.2.5		$		select MEAN WEF	account for NPE snow and frost, flag sand	
3.2.3.2		$		compute aggregated fractions using MEAN WEF		
3.2.3.5		$		compute reference values using WEF		
3.2.3.3&4		$		apply decision tree first part		
		$		select retrieval fraction group and model		
3.2.3.6		$		apply decision tree second part		
3.2.4		$		iterative retrieval (it.ret.) begins		
3.1.2.2, 3.1.3, 3.1.4			$	compute dielectric constants when necessary		
0			$	apply roughness correction		
3.1.2.6& 7			$	apply vegetation layer		
3.1.5			$	compute atmospheric and sky contributions		
3.2.2.4.3			$	compute (incidence) fractions using WEF		
3.1.1.3			$	compute TOA TB H/V composite forward model		
3.2.2.1.8			$	compute TOA TB X/Y	Faraday angles and Geometric angles	
3.2.4.2		$		compute cost function		
3.2.4.4		$		compute derivative matrices, increments		
		$		apply convergence test		
		$		apply L-M descent		
		$		iterative retrieval (it.ret.) ends		
3.2.5.1		$		Retrieval analysis		
3.2.5.1		$		Optional repeated attempts (it. retr. loop)		
3.2.5.4 & 5		$		Diagnostics		
3.2.5.2.1 to 4		$		Update current tables		(TAU, HR, RFI)
3.2.5.3		$		Compute surface TB (single angle)		
3.2.5.5 & 6		$		Generate L2 flags and indexes		
3.4.4.1	$	$		Build L2 output User data product		UDP file
3.4.4.2	$	$		Build L2 output Data Analysis Product		DAP file

The following Figure 1 is a much aggregated graphical representation of Table 6.
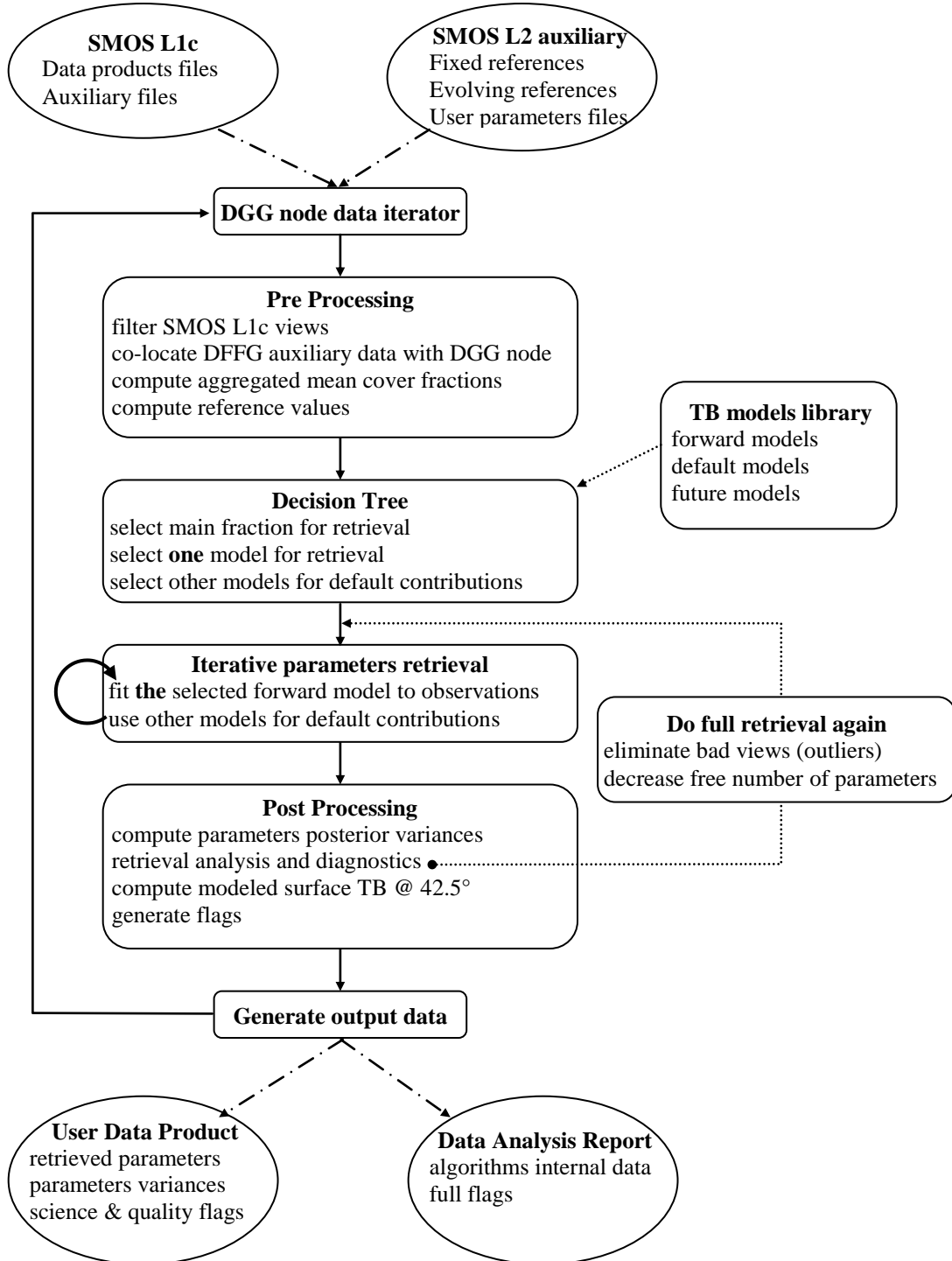
Figure 1 : General Layout

2.4 Known limitations

1. A first category of limitations is due to the **direct models and their parameterizations**:

   - Whenever available, they include simplifications with respect to the detailed physics.

   - For some scenes, validated direct models are not available.

   - Some of their parameters are fed from auxiliary data that may not be always available and may include errors.

2. A second category is due to direct model scale validity which is essentially local. Then, due to the large size of the SMOS pixel:

   - Average values have to be retrieved over heterogeneous targets where in reality a spatial distribution lies.

   - Errors may be due to the presence of nonlinear behaviour and saturation effects either natural (water bodies and large dense forests) or manmade (RFI).

3. Concerning the **retrieval algorithm** (formerly in **section 3.4.4.3)**

   - In the retrieval, use is made of fractions (as defined above) where the radiometric contribution is assumed to be known (depending on land cover types). This may involve strong assumptions. Hopefully, the SMOS data themselves will help to improve these estimates.

   - Failure of the retrieval cannot be excluded and must be provided for. However, the forward models are well behaved; no occurrence of divergence has been found so far, provided the input for auxiliary data and initial values are realistic.

   - Some inaccuracy on the **estimated posterior uncertainties** will result if the input uncertainties are not Gaussian distributed. This will happen if the uncertainties are large, as the parameters are non-negative.

2.5 Expected outputs

Depending on the data available and the nature of the SMOS pixel, the L2 processing will result in the following basic categories:

1. No valid retrieval whatsoever can be attempted.
2. SM retrieval is attempted and succeeds. Values for SM as well as for other parameters, typically vegetation optical thickness, are obtained.
3. Retrieval is carried out for geophysical quantities which do not include SM and succeeds.
4. Retrieval is attempted and fails.

Along with the retrieved products, the output should include information and flags whenever necessary concerning the quality and reliability of radiometric data and retrieved estimates, and information about the status of retrieval options.

The content of outputs is developed in section 3.4.4

2.6 Statistical/NN retrieval option

From ESA's comments: "the interface for a neural network should be defined in order to populate this net at a later date if needed". The reality of this option and the need to include this interface now can be discussed, but at least the architecture shall allow for a switch between methods, and preferably the interface definition for the NN".

The implementation of NN retrieval is expected to be vastly different. Before defining an interface, it is necessary to try to list the main differences. These are in Table 7 below:

Table 7: NN retrieval features

Input and Pre-processing		
	Auxiliary data	Certainly, non-necessary beyond those needed by the physical method.
	Data pre-processing	Incidence angle coverage of TB must be interpolated/averaged to categories to be defined; NN will be developed for each category (probably a few tens).
	Weighting functions and fractions	It is not clear that NN can handle incidence angle dependent fractions in retrieval. Possibly the strip adaptive apodization [33] should be preferred. In any case, a fraction of computation (probably simpler) remains needed.
Offline data and tools		
		A major part of the algorithm, offline, consists of building the learning base and tuning the NN on it. Learning bases must be built and NN tuned for every incidence angle coverage and retrieval option.
Decision tree		
	Overall structure	There should still be one including thresholds for some cover fractions. Probably simpler than when using the physical method. However, the branching according to incidence angles must be added.
	Default models	It is not foreseen that NN will accommodate default contributions.
	Retrieval options	Some will be kept provided specific learning bases and NN algorithms are incorporated.
Post-processing		
	Quality control	Require a specific NN architecture to provide *posterior* conditional variance estimates or require a NN architecture that provides directly an estimate of the posterior conditional probability distribution of retrieved parameters.

This table, although built from a very preliminary analysis, shows that the implementation is bound to be **vastly different**, with probably two main areas of difference:

- Whereas in the physical method there is a "kernel" (the cost function minimization algorithm and the channels to feed it) around which the whole architecture is built, no such kernel probably exists in the NN option. While in the physical method the decision tree mainly selects the physical modules to be fed to the optimization routine, in NN the tree selects the full specific retrieval algorithm.

- Whereas in the physical method no major off-line component is identified (leaving aside external calibration), in the NN a major component is necessary with the data necessary to build the **learning basis** and the software necessary to train the networks. It is true that there is some similarity, since indeed a learning base can only be obtained from calibration data, but the calibration is not basically constitutive of the physical method.

The simplest solution consists in keeping both algorithms completely separated. If this architecture is not wished, then there might probably be **two** interfaces, one somewhere in the pre-processing stage, and the other one when building the output product.

3 ALGORITHM THEORY DESCRIPTION

3.1 Physics of the problem

3.1.1 Overview of the radiative contributions

3.1.1.1 Thermal radiation

The SMOS system is a microwave imaging radiometer with aperture synthesis. It collects TOA directional polarized (TBH & TBV or Stokes parameters) radiances coming from the scene viewed by SMOS antennas through their power patterns. At SMOS frequency (1.4 GHz), the Plank's law Rayleigh-Jeans approximation holds very well, meaning that brightness temperature and radiances are directly proportional.

Instantaneous up-welling radiation is described by electric fields E_H and E_V for horizontal and vertical polarizations, as shown in Figure 2. The horizontal field component is perpendicular to the plane defined by the nadir SO and line of sight SP lines, while the vertical component lies in this plane.

The polarization state of partially polarized thermal radiation may be fully characterized by the set of four Stokes parameters (in the geographical reference frame) ST_1, ST_2, ST_3, ST_4. Because most conventional radiometers for Earth remote sensing perform TB_H and TB_V measurements, an alternate representation of the Stokes vector may be given in terms of brightness temperatures:
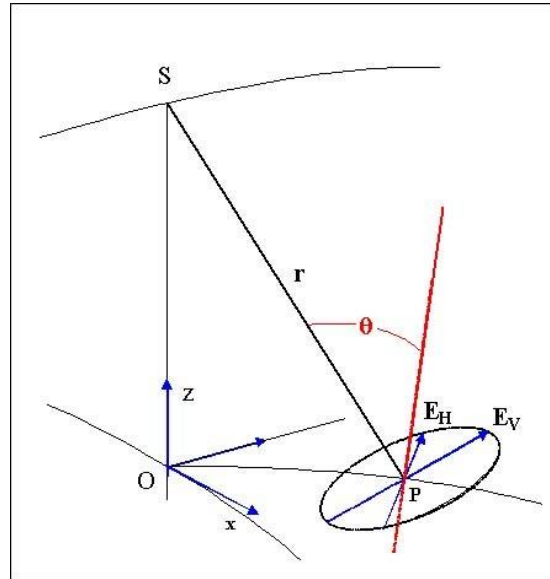
Figure 2: Surface electric field components

$$TB_H = \gamma_{ST} \langle E_H E_H^* \rangle; \quad TB_V = \gamma_{ST} \langle E_V E_V^* \rangle;$$
$$TB_3 = 2\gamma_{ST} \, Re\langle E_V E_H^* \rangle; \quad TB_4 = 2\gamma_{ST} \, Im\langle E_V E_H^* \rangle$$

Eq 1

where γ_{ST} □is a multiplying factor relating each brightness temperature TB to the electric power density: $\gamma_{ST} = \lambda^2 / (k_{BC} \, \eta_{FS} \, BD)$, where λ is the operating wavelength, k_{BC} is the Boltzmann constant, η_{FS} is the intrinsic impedance of free space, and BD is the bandwidth. The "$\langle \rangle$" symbol designates ensemble average.

TB_H (= $(ST_1-ST_2)/2$) and TB_V (= $(ST_1+ST_2)/2$) are the horizontal and vertical up-welling TB components.

Figure 3 depicts the signal measured at satellite level. It is a brightness temperature consisting of four main contributions: i) the up-welling atmospheric emission TB_{atu}; ii) the Earth's surface emission TB_{sp}, attenuated by the atmosphere iii) the atmospheric down-welling atmospheric emission TB_{atd} reflected (scattered) at the surface and attenuated along the upward path by the atmosphere; and finally the cosmic background emission TB_{sk} attenuated by the atmosphere, reflected /scattered (r_p) at the surface and attenuated again along the upward path by the atmosphere.

3.1.1.2 Radiative transfer equation

Combining these 4 components gives the general **radiative transfer equation** (RTE)[15]:

$$TB_p = TB_{atu} + TB_{sp} \exp(-\tau_{atu}) + (TB_{atd} + TB_{sk} \exp(-\tau_{atd})) \, r_{sp} \exp(-\tau_{atu})$$

Eq 2

All the terms of the above equation are functions of frequency and incidence angle θ (see Figure 1) between the line of sight and the local normal to Earth surface; the "p" subscript indicates the **polarization**. The "s" subscript refers here to combined (surface + near surface) layers.

The upward and downward path atmospheric opacities τ_{atu} and τ_{atd} depend on the gaseous and liquid droplet attenuating constituents (primarily oxygen, water vapour and clouds). Considering that we are operating at L-band, we can safely assume that τ_{atu} and τ_{atd} are almost equal, as differences are linked to differences in atmospheric temperatures and constituent profiles between the two paths. They will be both assigned as τ_{atm}.

The **surface reflectivity** r_{sp} is the integral of the surface scattering coefficient over all scattering directions. This element is the key to what we need to retrieve. Its main influence over the overall brightness lies in its indirect influence on the surface TB_{sp}, since r_p is the complement to 1 of emissivity (see below). The atmospheric radiation components TB_{atd} and TB_{atu} are dependent upon the vertical profiles of temperature, gaseous constituents and liquid droplets in the atmosphere. Their computation takes into account absorption and scattering. At L-band, atmospheric effects are small and TB_{atd} and TB_{atu} can be considered as equal to TB_{atm}.

At L-band the so-called **Faraday rotation**, linked to the columnar electron content (TEC) of the ionosphere over the path, causes the polarization to be rotated by on average up to a few degrees. This factor has to be taken into account when the TEC (hence the effect) is high (afternoon pass, high solar activity / bursts).

Finally, TB_{sk} is the **sky background**. At L-band, several sources are present; the galactic plane contains a number of significant sources that might have to be accounted for (see 3.1.5.2). One should not forget the **Sun**, which at L-band is a very significant source (100 000 to 300 000 K) and will have to be considered.

Surface variables such as temperature, roughness vegetation, snow, etc… enter the general radiative transfer equation through their effects on surface reflectivity r_{sp} and surface brightness temperature TB_{sp}:

$$TB_{sp} = e_{sp} T_s \qquad\qquad\qquad \textbf{Eq 3}$$

where e_{sp} is the **surface emissivity** ($e_{sp} \cong 1 - r_{sp}$) and T_s is the **effective** (physical) surface temperature.

For **bare soil** surfaces, T_s reduces to a weighted sum T_g of soil temperatures at subsurface levels accounting for the penetration depth.

In the presence of vegetation (or snow, etc), the interpretation of r_{sp} and T_s must be developed further. The general case is indicated in Figure 3 by a uniform vegetation layer above a rough soil surface. For the sake of clarity, we will focus here on the low vegetation layer approximated by a zero[th] order model. More details and other types of layers are described in sections 3.1.2 - 3.1.6.

The vegetation layer scatters and absorbs radiation incident from the atmosphere above and from the soil below in addition to contributing its own emission. Following the classical approaches:

- The above surface (vegetation or canopy) layer is characterized by a **canopy** physical temperature T_c, an opacity τ_c, and an isotropic single scattering albedo ω.
- The underlying soil (ground) surface is described by its effective temperature T_g (see 3.1.2.4) and emissivity e_{gp} ($e_{gp} = 1 - r_{gp}$).

In summary and still considering the soil/atmosphere interface as a simple layer, an expression of the SMOS observed brightness temperature TB_p viewed at an incidence angle θ can be derived by summing the components in Figure 3. It will be assumed that there is negligible reflection at the atmosphere vegetation interface.

The radiation components are, assuming we have one layer above the ground with a temperature T_c

- Atmospheric upward emission TB_{atu}
- Soil-surface emission attenuated through canopy and the atmosphere: $e_{gp} T_g \exp(-\tau_c) \exp(- \tau_{atu})$
- Downward atmospheric and cosmic background (and eventually solar) radiation attenuated through the layer (canopy) and atmosphere: $r_{gp} (TB_{atd} + TB_{sk} \exp(-\tau_{atd})) \exp (-2 \tau_c) \exp (-\tau_{atu})$
- Upward layer (canopy) emission attenuated through the atmosphere: $T_c (1-\omega) (1-\exp(-\tau_c)) \exp (-\tau_{atu})$
- Downward layer (canopy) emission scattered at the surface and attenuated through the layer and the atmosphere $T_c (1-\omega) (1-\exp(-\tau_c)) \exp (-\tau_c) \exp(-\tau_{atu}) r_{gp}$

One obtains the composite equation:

$$TB_p = TB_{atu} + \exp(-\tau_{atu})(TB_{atd} + TB_{sk}\exp(-\tau_{atd}))\, r_{gp}\exp(-2\,\tau_c)$$
$$+ \exp(-\tau_{atu})[\, e_{gp}T_g\exp(-\tau_c) + T_c(1-\omega)(1-\exp(-\tau_c))(1+r_{gp}\exp(-\tau_c))\,]$$

Eq 4

Where τ_c and ω are defined by layer extinction and scattering coefficients κ_{ext} and κ_{sct}:

$$\tau_c = \int_0^{LH\,\sec(\theta)} \kappa_{ext}\, dx$$

Eq 5

where LH is the layer height and θ the incidence angle:

$$\omega = \kappa_{sct}/\kappa_{ext}$$

Eq 6

It has however been shown that, for vegetation (characterized by cylindrical features), scattering is non-isotropic and dominant in the forward direction. For such cases, the normalized phase function for the canopy can be expressed as the sum of a Dirac function and a modified phase function. To account for this, in the non-isotropic case τ_c and ω should be replaced by an **effective** canopy opacity τ_c^* and an effective single scattering albedo ω^* where

$$\tau_c^* = (1-\alpha_{sct}\,\omega)\tau_c$$

Eq 7

$$\omega^* = (1-\alpha_{sct})\omega/(1-\alpha_{sct}\,\omega)$$

Eq 8

and where α_{sct} is a coefficient related to the canopy structure, which characterizes the proportion of radiation scattered in the forward direction. A general formulation is then derived according to the layer type (low vegetation snow etc) as described in sections 3.1.2.7 & 3.1.2.8, where for the sake of simplicity **the superscript * is dropped** .

Finally, it must be said that the layer above the surface can be vegetation but also snow, ice layer, and that in many cases the layer will actually be a set of layers. Just as a complex example we might have above the soil's surface a layer of litter, then snow, then vegetation (understory), then trees and finally snow /ice on the trees themselves.

3.1.1.3 Aggregated radiative transfer equation

At the SMOS scale (25-60 km), pixels are not uniform and we may have a variety of surface types, for instance a rural area with towns and roads, bare fields, fallow land and some crops, thickets or woodland, the occasional river or pond, and again, in the worst case, snow here and there with frozen grounds in some places!

In such cases, the total brightness temperature comes from several classes of emitters. This composite brightness temperature is obtained through an **aggregated** forward model that combines each class of emitting sources weighted by their intra-pixel cover **fractions**.

To show clearly how this aggregation is done, for a given polarization and incidence angle and a **homogeneous** L1c scene, we first rewrite equation Eq 4 assuming that downwards and upward atmospheric contributions are equal:

$$TB_p = TB_{atm} + \exp(-\tau_{atm})[TB_{atm} + TB_{sk}\exp(-\tau_{atm})]\, r_{gp}\exp(-2\,\tau_c)$$
$$+ \exp(-\tau_{atm})[\, e_{gp}T_g\exp(-\tau_c) + T_c(1-\omega)(1-\exp(-\tau_c))(1+r_{gp}\exp(-\tau_c))\,]$$

Eq 9

The reflectivities and emissivities r_{gp} and e_{gp} include both smooth surface effects from the dielectric constant and roughness effects. The method to build a single physical temperature parameter from T_g and T_c is discussed in subsections 3.1.2.4 & 3.1.2.6.

In the description of atmospheric contributions, we shall refer below to an equivalent <u>physical</u> layer temperature, linked very simply to TB_{atm} and τ_{atm}.

Many terms and factors in this expression depend on polarization and incidence angle. This is detailed in forward models below.
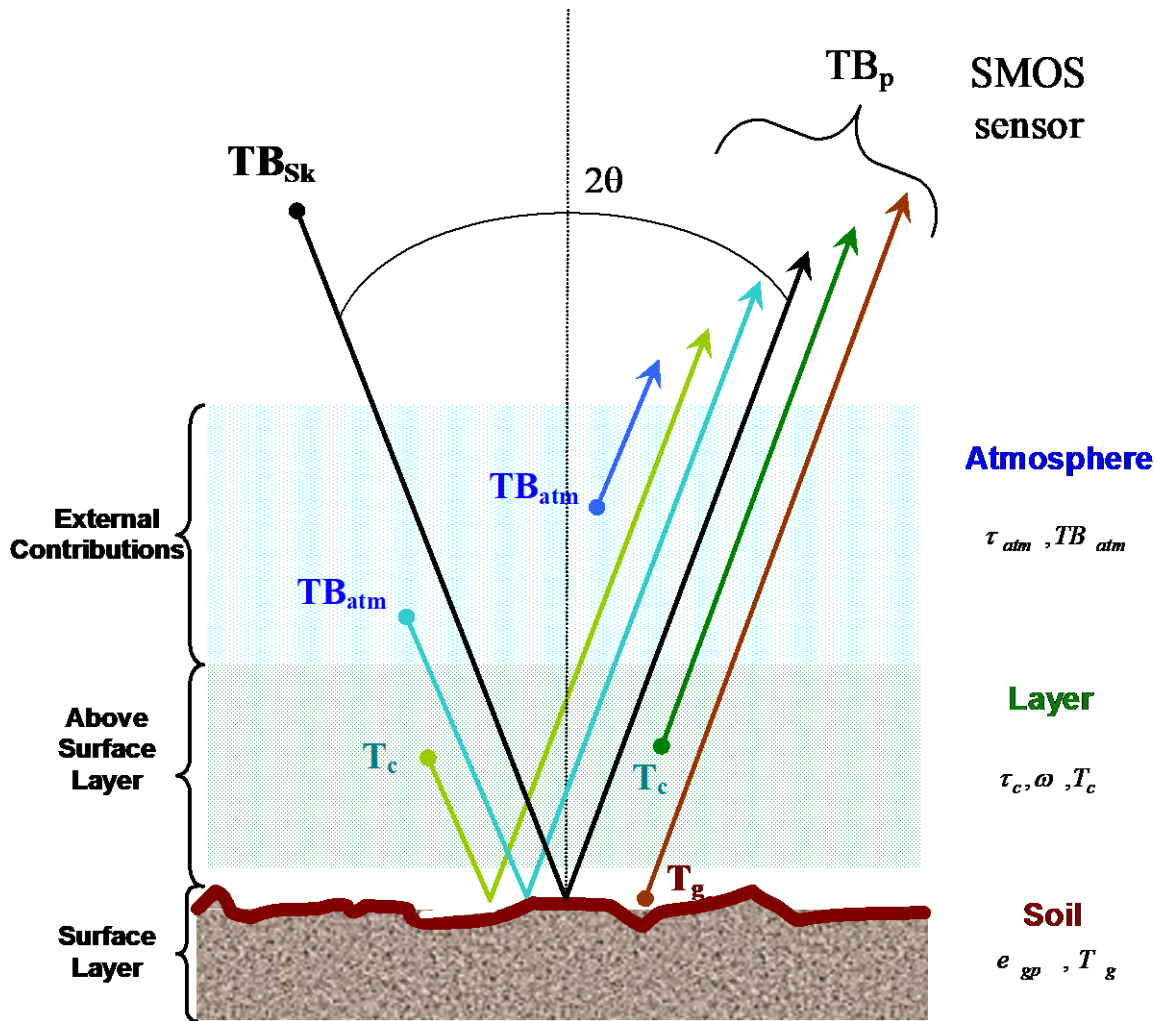
Figure 3: Contributions to TOA brightness temperature

Consider now a **mixed** L1c scene with n = 1 to NF **mean fractions** (over incidence angle) FM_n. Of course, NF is actually a small number. For each L1c view, **incidence angle dependent** values FV_n for fractions are to be computed.

For ease of writing, we rewrite Eq 9 as follows:

$$ = TB_{atm} + \exp(-\tau_{atm}) [TB_{atm} + TB_{sk} \exp(-\tau_{atm})] R1 + \exp(-\tau_{atm}) R2 \qquad \textbf{Eq 9a}$$

Where only the expressions R1 (dimensionless) and R2 (in Kelvin) depend on the fraction n. Then the **aggregated** forward model, for each view, is derived from Eq 9 where:

R1 becomes: $SUM_{n=1:NF} \{ FV_n . R1_n \}$;   R2 becomes: $SUM_{n=1:NF} \{ FV_n . R2_n \}$

The contributions $R1_n$ and $R2_n$ are computed with the help of forward models described in following subsections of section 3.1. Fractions FM_n and FV_n are presented in section 3.2.2.5 and in the decision tree section 3.2.3.

3.1.1.4   Towards elementary radiative models

In the following, elementary radiative models are described whenever available. If no model exists (i.e. urban) it is proposed to put a placeholder with a proxy model (in this case some sort of bare soil). Then:

- We are mainly interested in scenes devoid of strong topographic features, possibly covered by **low vegetation**, for which volume surface moisture can be defined. This will be called the **nominal** SMOS target (in short NO for nominal, or LV for low vegetation). Forward models are available.

- It may happen that, although soil moisture is in principle relevant, forward models are poorly known or not validated. This is e.g. the case for strong topography, snow cover.

- In some cases, soil moisture is no longer relevant. Examples are open water, ice.

We will now address the details of nominal models as well as other cases.

The nominal case develops the way to model surface roughness as well as the vegetation layer. Note that

- surface roughness is also present for other cases excepting all water surfaces;
- vegetation layer is also present for other cases, excepting free water surfaces but including wetlands

3.1.2 Nominal case (vegetated soil)

The modelling approach used here relies on an extensive review of current knowledge and previous studies. It accounts for, as much as possible, emission from various land covers, from bare soil to full vegetation-covered surfaces, snow-covered surfaces, open water, and atmospheric effects.

The **nominal case** is the case where we believe soil moisture retrievals will be feasible. It consists of a mixture of mineral and organic soil covered by low vegetation, eventually a manageable amount of free water. The "manageability" is expressed by **thresholds** for which values are suggested in the decision tree section but will often require confirmation.

3.1.2.1 Bare Soil

Bare soils are quasi-opaque at 1.4 GHz, so the radiative budget is mainly ruled by their emissivity e and reflectivity r, for each polarization p, with:

$$e_{gp} + r_{gp} = 1 \qquad \textbf{Eq 10}$$

The emission of microwave energy is governed by the product of the soil effective temperature, T_g, and soil emissivity, e_{gp}. At L-band, the emissivity e_{gp} is in its turn a function of the soil's characteristics, i.e. moisture, texture, roughness and eventually salinity.

3.1.2.2 Smooth Bare Soil Dielectric Properties

The theory behind the microwave remote sensing of soil moisture is based on the large variation of emissivity with soil water content. This is because the real part of the dielectric constant value of "ordinary" soil varies between that of dry soil (< 4) and that of liquid water (~ 80) depending on its actual water content. Consequently, as soil moisture increases, the emissivity (all other things remaining constant) decreases, and this change is detectable by microwave sensors.

This qualitative description is formalized as follows. The reflectivity r_{bp} of a perfectly smooth surface is given by the Fresnel law that defines the partition of electromagnetic energy at a flat dielectric boundary [34]. The Fresnel reflection coefficients r_{bH} and r_{bV} at H and V polarizations, respectively (a rigorous notation would be r_{bgH}, with the "b" subscript standing for **smooth** and **bare (bald)** soil so not covered by any vegetation layer) are given by:

$$r_{bH}(\theta) = \left| \frac{\mu_s \cos(\theta) - \sqrt{\mu_s \varepsilon_b - \sin^2(\theta)}}{\mu_s \cos(\theta) + \sqrt{\mu_s \varepsilon_b - \sin^2(\theta)}} \right|^2 \qquad r_{bV}(\theta) = \left| \frac{\varepsilon_b \cos(\theta) - \sqrt{\mu_s \varepsilon_b - \sin^2(\theta)}}{\varepsilon_b \cos(\theta) + \sqrt{\mu_s \varepsilon_b - \sin^2(\theta)}} \right|^2 \qquad \textbf{Eq 11}$$

where μ_s is the soil magnetic permeability, **assumed to be unity**, ε_b, is the complex, smooth, bare soil dielectric constant (**medium dependent**), and θ is the incidence angle.

Then, for smooth bare soil, the upwelling soil brightness temperature may be written as a function of the soil effective temperature T_g and soil reflectivity r_{bp} computed from the Fresnel equation:

$$TB_p = (1 - r_{bp}) \, T_g \qquad \textbf{Eq 12}$$

We consider bare soils as a mixture of mineral soils and organic soils defined by the fraction of organic soil RSOM; when RSOM=1 the bare soil is fully organic, when RSOM=0 the bare soil is fully mineral.

The dielectric properties of this mixture are represented by the weighted mean of the dielectric constants of mineral soil, ε_m, and organic soil, ε_o.

$$\varepsilon_b = (1 - RSOM) \times \varepsilon_{ms} + RSOM \times \varepsilon_{os} \qquad \textbf{Eq 13}$$

While the dielectric constant of wet mineral soil, ε_{ms}, has been formulated in different ways by several authors [35], we have selected two approaches The universally used Dobson semi-empirical model [36, 37] which has been originally used in previous releases of this document, and the more recent Mironov semi-physical model [38] [39, 40] introduced in the ATBD version 3.f, which offers a more physical approach and better numerical stability.

Although both model formulation leads to similar soil moisture retrievals in SMOS context [41] the Mironov model has some advantages:

- its formulation, based on refractive index, is more physical and lead to a simpler and more robust mathematical formulation close to SM=0 where Dobson formulation involves an instable hyperbolic uses of SM thus more successful retrievals are obtained for very dry conditions.

- empirical relations which depend on soil texture are valid for a wide range of soil texture as they are obtained from spectroscopic validated model. In comparison Dobson model use five soil texture types with a weak representativeness of very sandy soils.

- Mironov model emissivity is also warmer for dry and sandy situations where Dobson model is known [42] to have some limitations.

The two above dielectric constant models are designed for mineral soils made of a mixture of sand, clay and silt. They are not well suited to model organic soils dielectric constant with correct volumetric soil moisture[43]. In this ATBD version 4.a, the Bircher empirical dielectric constant model [44] is considered and used for organic surfaces.

These three models detailed in the following sections. Both are considered as selectable options for the retrieval, with Mironov model being the default choice for mineral soil and Bircher for organic soils.

Note: although these dielectric constant models are perfectly defined as forward models, they need to be slightly modified when they are used for retrieval in order to insure a good convergence to valid retrieved soil moisture values. Please read the section 3.5.1 about this aspect.

3.1.2.2.1 Dobson mineral soil dielectric constant model

According to the Dobson model [36, 37] and using the Peplinki's formulation [45, 46], the dielectric constant of mineral wet soil, ε_m, can be calculated as:

$$\varepsilon_{ms} = \left(1 + \frac{\rho_b}{\rho_s}\left(\varepsilon_{pa}^{\alpha} - 1\right) + SM^{\beta'}\,\varepsilon_{sfw}'^{\alpha} - SM\right)^{\frac{1}{\alpha}} - j\left(SM^{\beta''}\varepsilon_{sfw}''^{\alpha}\right)^{\frac{1}{\alpha}} \qquad \textbf{Eq 14}$$

where

ε_{msb} = soil dielectric constant for mineral bare soil

- ρ_b = soil bulk density, function of soil texture. Default value is 1.3 gcm^{-3}.

- ρ_s = soil particle density, usually considered as constant. Default value is 2.664 gcm^{-3}

- ε_{pa} = dielectric constant of solid particles. For "normal" soils:

$$\varepsilon_{pa} = (CPA_1 + CPA_2 * \rho_s)^2 + CPA_3 \text{ ; (CPA coefficients in TGRD UPF), yielding } \varepsilon_{pa} \approx 4.7$$

- $\alpha = 0.65$

- $\beta = \beta' - j\,\beta''$ is an empirically-derived complex function of soil texture parameter usually calculated as in [36, 37]

- SM = soil moisture, volumetric water content of soil (m^{3}m^{-3})

- $\varepsilon_{sfw} = \varepsilon'_{sfw} - j\,\varepsilon''_{sfw}$ dielectric constant of free water included in the soil

With

$$\varepsilon'_{sfw} = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w\infty}}{1 + \left(2\pi\, f\ r\tau_w\right)^2}$$

$$\varepsilon''_{sfw} = \frac{2\pi\, f\ r\tau_w\left(\varepsilon_{w0} - \varepsilon_{w\infty}\right)}{1 + \left(2\pi\, f\ r\tau_w\right)^2} + \frac{\sigma_{eff}}{2\pi\, f\ \varepsilon_0}\frac{\rho_s - \rho_b}{\rho_s\ SM} \qquad \textbf{Eq 15}$$

This expression is as described in equations Eq 51a to Eq 52e but adapted to the specific case of soil by approximating the effective conductivity of water using the Stern-Gouy double layer theory (see [36, 47, 48]) and where:

- ε_{w0} is the static dielectric constant of water ε_{w0}

- $\varepsilon_{w\infty}$ is the high frequency limit of the dielectric constant of water ε_{fw}

- f is the mean SMOS frequency (Hz),

- rτ_w is the relaxation time of water,

- ε_0 is the permittivity of free space (8.854 10^{-12} Fm^{-1})

$$\sigma_{eff} = SGEF_1 + SGEF_2\,\rho_b + SGEF_3\,S + SGEF_4\,C$$
$$\beta' = BERE_1 + BERE_2\,S + BERE_3\,C$$
$$\beta'' = BEIM_1 + BEIM_2\,S + BEIM_3\,C$$

Eq 16

Coefficients SGEF, BERE, BEIM in Eq 16 are provided in the TGRD, while S is the sand fractional content of the soil and C the clay fractional content of soil.

Other expressions are linked to water and are given in Eq 48 to Eq 52e.

Note: the notion of soil salinity is not currently used. It is kept for future use. The current algorithm uses only the dielectric constant of pure water given in Eq 51a-b.

3.1.2.2.2 Mironov mineral soil dielectric constant model

According to Mironov model [40], the wet mineral soil dielectric constant, $\varepsilon_{ms} = \varepsilon_s' - j\varepsilon_s''$, is expressed using the complex index of refraction, $n_s^* = \sqrt{\varepsilon_s}$, as the refractive index linear mixing model of the complex refractive indexes of completely dry soil, n_d^*, soil bound water, n_b^*, and unbound water in the soil, n_u^*:

$$n_s^* = n_d^* + \left(n_b^* - 1\right)\left(SM - \Delta_{SM}H\left(\Delta_{SM}\right)\right) + \left(n_u^* - 1\right)\left(SM + \Delta_{SM}H\left(\Delta_{SM}\right)\right)$$

Eq 17

where $\Delta_{SM} = SM - XMVT$ is the difference between soil moisture, SM, and the wilting point, $XMVT$, $H(x)$ is the Heaviside step function, $H(x)=1$ for $x>0$, $H(x)=0$ for $x\le 0$.

The complex index of refraction, n_s^*, of wet soil is more conveniently separated into its real part, the refractive index $n_s = \mathrm{Re}\left(n_s^*\right)$, and the normalized attenuation $k_s = \mathrm{Im}\left(n_s^*\right)$ in Eq 17. From the knowledge of n_s and k_s the dielectric constant of wet soil can then be computed:

$$\varepsilon_s' = n_s^2 - k_s^2$$
$$\varepsilon_s'' = 2n_s k_s$$

Eq 18

The rest of this section is dedicated to the computation of n_p and k_p as a function of the texture through the clay fraction, C, the soil temperature in degree centigrade, TC, the frequency, f, using the following equations, for the three soil states of Eq 17: dry soil, $p=d$, bound water, $p=b$, and unbound water, $p=u$.

Completely dry soil refractive index and normalized attenuation ($p=d$) are given as an empirical function of the clay fraction, C:

$$n_d = ND0 + ND1*C + ND2*C^2$$
$$k_d = KD0 + KD1*C$$

Eq 19

While for bound and unbound water refractive index and normalized attenuation ($p=b,u$) is given by:

$$n_p = \frac{\sqrt{\|\varepsilon_{wp}\| + \varepsilon'_{wp}}}{\sqrt{2}}$$

$$k_p = \frac{\sqrt{\|\varepsilon_{wp}\| - \varepsilon'_{wp}}}{\sqrt{2}}$$

Eq 20

Where $\varepsilon_{wp} = \varepsilon'_{wp} - j\varepsilon''_{wp}$ is the dielectric constant for bound, $p=b$, and unbound, $p=u$, water fractions and is computed as described in the following equations.

$$\varepsilon'_{wp} = \frac{\varepsilon_{w\infty} + (\varepsilon_{0p} - \varepsilon_{w\infty})}{(1 + (2\pi f \tau_p)^2)}$$

$$\varepsilon''_{wp} = \frac{2\pi f \tau_p (\varepsilon_{0p} - \varepsilon_{w\infty})}{(1 + (2\pi f \tau_p)^2)} + \frac{\sigma_p}{(2\pi * \varepsilon_0)}$$

Eq 21

where

- $\varepsilon_{w\infty}$ is the high frequency limit of the dielectric constant of water ε_{fw}

- f is the mean SMOS frequency (Hz),

- ε_0 is the permittivity of free space (8.854 10^{-12} Fm^{-1})

The water relaxation time, τ_p is computed by:

$$\tau_p = \frac{TAU0_p\, e^{(\Delta H_p / TK - \Delta S_p)}}{TK}$$

Eq 22

Where

- $TAU0_b = TAUB0$

- $TAU0_U = TAUU0$

- $TK = TC + 273.15$

- $\Delta H_b = DHBR0 + DHBR1*C + DHBR2*C^2$

- $\Delta S_b = DSRB0 + DSRB1*C + DSRB2*C^2$

- $\Delta H_u = DHUR0 + DHUR1*C$

- $\Delta S_u = DSUR0 + DSUR1*C$

The ohmic conductivity, σ_p, is computed by :

$$\sigma_b = SBT0 + SBT1*C + \beta_{\sigma b}(TC - TF0)$$
$$\sigma_u = SUT0 + SUT1*C + \beta_{\sigma u}(TC - TF0)$$

Eq 23

Where

- $\beta_{\sigma b} = BSGB0 + BSGB1*C + BSGB2*C^2 + BSGB3*C^3 + BSGB4*C^4$

- $\beta_{\sigma u} = BSGU0 + BSGU1*C$

The computation of the static dielectric constant of water, ε_{0p}, is given by:

$$\varepsilon_{0p} = \frac{\left(1 + 2e^{(F_p - \beta_p(TC - TF_0))}\right)}{\left(1 - e^{(F_p - \beta_p(TC - TF_0))}\right)}$$

Eq 24

Where

- $TF_0 = 20°C$

- $F_b = \ln\left(\dfrac{\varepsilon_{b0} - 1}{\varepsilon_{b0} + 2}\right)$

- $F_u = \ln\left(\dfrac{\varepsilon_{u0} - 1}{\varepsilon_{u0} + 2}\right) - \beta_u(TC - TF_0)$

- $\beta_b = BVB0 + BVB1*C + BVB2*C^2 + BVB3*C^3 + BVB4*C^4$

- $\beta_u = BVU0 + BVU1*C$

- $\varepsilon_{b0} = E0PB0 + E0PB1*C + E0PB2*C^2$

- $\varepsilon_{u0} = E0PU$

Finally, the wilting point, *XMVT*, used in Eq 17 is given by:

$$XMVT = XMVT0 + XMVT1*C \qquad \textbf{Eq 25}$$

The values of all the regression coefficients in the above equation are given in the TGRD UPF described in TGRD. They are based on refined values[2] taken from [40].

3.1.2.2.3 Bircher's organic soil dielectric constant model

The HiLat ESA STSE recent project [49] showed the importance of organic soils in carbon storage in particular at high latitudes. SMOS has the capability to observe these areas but so far, the Mironov or Dobson mineral dielectric constant model are not appropriate to model the emissivity of such surfaces. One of the outcomes of the HiLat project was that an simple empirical organic soil permittivity model could be obtained from samples and laboratory permittivity measurements [43]. As a continuation the CCN extension to the HiLat project [44] more validation sites were added to provide better insights toward an improved SMOS SM retrievals for organic soils. It pointed out the necessity of global maps of Soil Organic Matter (SOM) to decide where the organic soil permittivity model should be used instead of the model applicable to mineral soils. As a first start, the SoilGrids ORCDRC map [Hengl jet al., 2014] with adequate thresholds was shown to be suitable for such purpose though there are margins of improvements by fine-tuning and/or adaptation with other dataset.

The Bircher organic soil dielectric constant model is given by the following 3rd order polynomia in SM and coefficients [49] in a symmetrized form where the absolute value |SM| is used. As for the Mironov symmetrized dielectric constant, we have for the organic soil $\varepsilon_{os}(SM) = \varepsilon_{os}(-SM)$.

$$\varepsilon_{os} = \varepsilon'_{os} - i\varepsilon''_{os}$$
$$\varepsilon'_{os} = SOM'_3.|SM|^3 + SOM'_2.|SM|^2 + SOM'_1.|SM| + SOM'_0 \qquad \textbf{Eq 26}$$
$$\varepsilon''_{os} = SOM''_3.|SM|^3 + SOM''_2.|SM|^2 + SOM''_1.|SM| + SOM''_0$$

The eight $SOM'_{\{0,...,3\}}$ and $SOM''_{\{0,...,3\}}$ coefficients are provided in the TGRD UPF.

3.1.2.3 Surface roughness

When the surface is not flat and assuming that only surface scattering occurs, it is possible to estimate, from the flat bare surface reflectivity r_{bp}, an expression for the rough surface reflectivity r_{gp}. The most accepted formulation is an empirical relationship.

As surface roughness increases, the angular signature of TB is affected, requiring correcting the Fresnel law with the following empirical phenomenological expression:

$$r_{gp}(\theta) = \left((1 - QR)r_{bp} + QR\,r_{bq}\right)\exp\left(- HR(SM)\cos^{NR_P}(\theta)\right) \qquad \textbf{Eq 27}$$

where

---

[2] Reference to the publication to add once these new results published by Mironov & al.

- QR is a **polarization coupling factor**, related to the fact that roughness tends to induce polarization mixing.

- HR is an effective surface roughness dimensionless parameter: HR $=(2\ k\ \sigma)^2$ where k is the wave number, σ is the surface RMS height).

- NR_p is an integer used to parameterize the dependence of the roughness effects on incidence angle.

- r_{bq} designates the smooth surface reflectivity for alternate polarization.

Even though empirical, this formula has been tested in various occasions, and found to work well, provided several precautions are taken. At L-band the main issues are related to the fact that soil roughness should rather be seen as a 1.4 GHz effective soil roughness i.e. probably more related to the distribution of water in the top soil rather than a pure geometric soil surface roughness as the latter can only occur when the soil is very wet. Recent work [50, 51] indicates that HR is better modelled using a moisture dependent function. This point will be taken into account by using a soil water contribution in HR. The principle is to have HR as a function of soil moisture with a simple law (see Figure 4). Below a transition moisture point, XMVT(C,S), the roughness is constant as well as above the **field capacity**, FC(C,S), where it takes the classical expression (HR_MIN(LC) $=(2\ k\ \sigma)^2$).

The HR value for dry soil could be set a priori and/or adjusted from the data. The piecewise relationship will also have to be smoothed at SM=XMVT and SM=FC by an adequate function having continuous first derivatives with respect to SM. This property is required by the optimization procedure which uses gradients[3].

It must be understood that these expressions are not fully validated but as suggested by the following figure, the roughness dependence to soil moisture can be cancelled, if necessary, by setting HR_MAX(LC)=HR_MIN(LC).
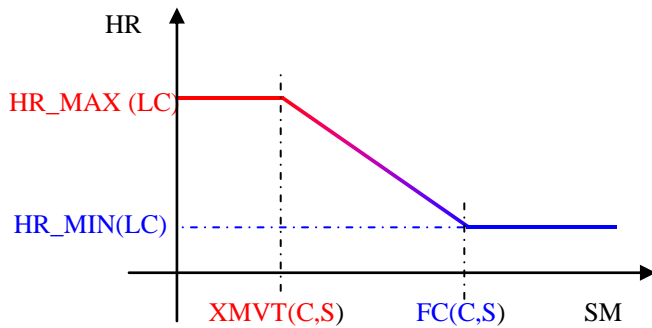
Figure 4: HR(SM): roughness as a piecewise function of SM

The two parameters XMVT and FC are function of the sand, S, and the clay, C, fractions.

From S and C, the transition moisture XMVT can be computed. First, we define the wilting point by:

WP(C,S) = CWP1 +CWP2*S + CWP3*C   **Eq 28a**

CWP1=0.06774, CWP2=- 0.00064 and CWP3=0.00478 are stored in the TGRD UPF.

and the final transition moisture by:

XMVT(C,S) = CXMVT1*WP(C,S) + CXMVT2   **Eq 28b**

CXMVT1=0.49 and CXMVT2=0.165 are stored in the TGRD UPF.

Field capacity is defined accordingly to [52] by:

FC(C,S) = 0.3 – 0.0023*S + 0.005*C
with C and S in percentages.   **Eq 28c**

However both XMVT and FC values will be given in the auxiliary table SOIL_PROPERTIES at the DFFG scale.

---

[3] Note the value of XMVT(C,S) can be 0. To be accounted for in the formulation of the smoothing function.

The HR_MIN and HR_MAX values are function of the land cover type, LC. Their reference values will similarly be given in the **LAND_COVER_CLASSES** auxiliary table. When the HR_MAX parameter is retrieved[4], then its reference value computed as indicated here will be used as a prior value in the retrieval scheme.

The value of NR_p is found to be between 0 and 2 from experimental data [53] (originally, it was considered that $NR_p=2$). Recent results indicated that the NR_p exponent is also polarization dependent. The main issue is the extension of this local scale approach to SMOS resolution. The first analysis of global data sets (such as SMMR or AMSR) is promising [54].

Polarization coupling effects are generally found to be rather weak at low frequencies. Therefore, it is often considered that QR=0 at L-band and this value increases slightly with increasing frequency [53].

The soil moisture dependence introduced in the HR(SM) function is only meaningful for surfaces where soil moisture is defined which is the case only for the nominal and forest surfaces. For the other surfaces we keep the same general formulation but with an adapted parameterization described further in section 3.1.4.8, all the non-nominal surfaces being introduced first.

3.1.2.4 Effective soil temperature

The effective soil temperature T_g depends on the soil properties and moisture content profile within the soil volume. A simple formulation developed originally by [53] and then validated and revised [55] will be used in the algorithm. This formulation introduces two soil temperatures T_{soil_surf} and T_{soil_depth}, to be selected from the 4 values supplied as auxiliary data (ECMWF fields).

The effective temperature is usually computed using a surface temperature and the temperature at a depth where it is almost constant. The actual profile and depth are dependent upon the soil type actual profile and the level at which the deep soil temperature is obtained. Obviously, for a global operational processor such as the one in consideration in this ATBD, such pieces of information are not necessarily available nor really affecting the result. We consider that the first layer and either the deepest or next to deepest layer given by ECMWF will give a good estimate of the surface and deep temperature. The errors induced are no more significant than those derived from using a crude interpolation scheme and have only impact in the case of very dry soils.

The effective soil temperature is written as a function of the soil temperature at depth (T_{soil_depth}, approximately at 0.5 to 1m depth) and surface soil temperature (T_{soil_surf}, approximately between 1 and 5 cm) as follows:

$$T_g = T_{soil\_depth} + C_t (T_{soil\_surf} - T_{soil\_depth})$$ **Eq 29**

where C_t is a parameter depending mainly on frequency and soil moisture. If the soil is very dry, soil layers at depth (deeper than one meter for dry sand) contribute significantly to the soil emission, and the value of C_t is lower than 0.5. Conversely, if the soil is very wet, the soil emission originates mainly from layers at the soil surface and $C_t \approx 1$.

[53] computed C_t as a function of surface soil moisture,

$$C_t = min\{(SM / w_0)^{bw0},1\}$$ **Eq 30**

Where the soil moisture estimate SM is **taken from auxiliary data**; w_0 and b_{w0} are parameters that depend mainly on the soil texture and structure. To simplify we will consider that
w_0, b_{w0} = function (soil type)

Note that [56] developed another parameterization of the C_t parameter based on the dielectric constant. This later parameterization, which has been validated only over one experimental site (SMOSREX) to date, will not be used in the algorithm.

In this study, we will actually select for C_t the former equation from [57], which was tested over both the Avignon and SMOSREX sites. It will be considered, as done in the above-cited references, that the "deep" soil temperature (T_{soil_depth}) is measured at 0.5 m depth while surface soil temperature (T_{soil_surf}) is measured at 5 cm. For these conditions, the value of w_0 was close to 0.3 m^3/m^3 over the bare soil sites of the SMOSREX and Avignon experiments. The value of b_{w0} was close to 0.3 m^3/m^3 over the Avignon site and close to 0.65 m^3/m^3 over the SMOSREX site.

The values $w_0 = 0.3$ m^3/m^3 and $b_{w0} = 0.3$ will be used as **default** values for the w_0 and b_{w0} soil parameters.

As neither the Avignon nor the SMOSREX test sites are really representative, it is suggested during the future SMOS validation to assess the best depth to be used for the deepest temperature for the effective temperature (i.e., after launch). In the meantime, the 21 to 72 cm layer is to be used for T_{soil_depth} i.e. ECMWF soil level 3.

---

[4] Note that the HR Max is what is reported in the SMOS product

3.1.2.5 Summary of bare soil parameters

See Table 8

Table 8 : Bare soil parameters

Surface TB of bare soil	Input/Parameter Name		Range	Units
Dobson or Mironov model to compute soil dielectric constant: ε_b **Note:** Mironov model does not require S, ρ_b, ρ_s	S	Sand fraction	0-100	%
	C	Clay fraction	0-100	%
	ρ_b	Dry bulk soil density	0.5-2.5	[g/cm³]
	ρ_s	Soil particle density	2 – 3	[g/cm³]
	SM[5]	Soil moisture	0-0.5	[m³/m³]
	Sal	Soil salinity	0-12[6]	[ppt]
	F	Frequency	1.4	[Ghz]
	T_g	Effective surface-deep soil temperature	250-350	[K]
Fresnel equations to compute the specular reflectivity H&V for smooth air-soil boundary r_{bp}	ε_b	Bare soil dielectric constant		[F/m]
	θ	incidence angle	0-55	[deg]
Introduce soil roughness to compute bare soil scattering / reflectivity: r_{gp}	r_{bp}	Specular smooth soil reflectivity	0-0.6	[-]
	θ	Incidence angle	0-1.25	[rad]
	QR	H/V polarization coupling factor	0.0-0.5	[-]
	HR	Surface roughness parameter	0-5	[-]
	NR_P	Power law of cos(θ)	0-5	[-]
Computing effective soil temperature	T_{soil_depth}	Soil temperature at depth (~at 46 cm)	250-350	[K]
	T_{soil_surf}	Soil temperature at surface (~ at 3.5 cm)	250-350	[K]
	w_0	Texture parameters used to compute the coupling factor Ct for effective soil temperature T_g computation.	0.05-2	[m³/m³]
	b_{w0}		0 – 2	

3.1.2.6 General considerations about vegetation

The above-surface vegetation layer is a location of multiple interactions and fluxes processing. Its impact on brightness temperature is two-fold:

- It may absorb or scatter the direct bare soil radiation and attenuate or reflect above surface radiation directly and indirectly, through bare soil reflectivity.

- It may provide its own upward and downward radiation; the latter leads to an indirect contribution through soil reflectivity and self -attenuation.

We will consider several classes in the general approach with two main parts:

- Low vegetation: grassland, crops.

- Forest vegetation: coniferous, evergreen and deciduous.

3.1.2.7 Low vegetation (grassland, crop)

When a vegetation layer is present over the soil surface, it attenuates soil emission and adds its own contribution to the emitted radiation. At low frequencies, these effects can be well approximated by a simple model based on the Radiative Transfer Equations (RTE), hereafter referred to as the **τ - ω model**. This model is based on two parameters, the **optical depth τ** and the **single scattering albedo** ω, that are used to parameterize, respectively, the vegetation attenuation properties and the scattering effects within the canopy layer. The reflection at the top of the canopy (at the vegetation–atmosphere interface) is neglected, contrary to the case of snow covers.

---

[5] The yellow is just to remind SM is our main retrieval parameter
[6] For ordinary soils, whereas for salinity affected areas this value may go much higher

No study could demonstrate the interest of using more complex radiative transfer models over rather low vegetation covers, where phase coherent effects (neglected by the RTE) may be significant [20]. Using the $\tau - \omega$ model, global emission from the two-layer medium (soil & vegetation) is for each polarization p the sum of three terms: (1) the direct vegetation emission, (2) the vegetation emission reflected by the soil and attenuated by the canopy layer and (3) soil emission attenuated by the canopy.

$$TB_P = (1-\omega_p)(1-\gamma_p)(1 + \gamma_p\, r_{gp})\, T_c + (1-r_{gp})\, \gamma_p\, T_g \qquad \textbf{Eq 31}$$

where T_g and T_c are the effective soil and vegetation temperatures, r_{gp} is the soil reflectivity, ω_p the single scattering albedo, γ_p the vegetation attenuation factor (where the c subscript has been dropped).

This last term can be computed from the optical depth τ_P as:

$$\gamma_p = \exp(-\tau_p / \cos\theta) \qquad \textbf{Eq 32}$$

The above equation is a way to define a modified nadir optical depth.

While refining these equations and defining the contributions to τ_p is a complicated research issue (see below), it is always possible to write:

$$\tau_p = \tau_{NAD} \times function(\theta, p)$$

where the **nadir estimates of overall optical depth τ_{NAD}** is independent of both incidence angle and polarization.

- **Surface temperature**: two options are considered:

- The first option is the strict application of the tau-omega model described in Eq 4 where two separate temperatures are used: Tg is set to the effective soil temperature described in the previous section 3.1.2.4 and Tc is set to the low vegetation temperature.

- Alternatively, in most studies (forward modelling and retrievals), it is assumed that effective soil (T_g) and vegetation (T_c) temperatures are approximately equal to a single value $T_{gc} \approx T_c \approx T_g$[7]. In particular, the effects of temperature gradients within the vegetation canopy should not be accounted for. With an overpass around dawn, the differences should be minimised, and T_c can be expected to be close to the air temperature, while T_g can be estimated.

An estimate of an **effective** composite temperature T_{gc} (including **both** soil and vegetation media) could be roughly evaluated from the following equation:

$$= A_t\, T_c + (1 - A_t)\, T_g \qquad \textbf{Eq 33a}$$

with
$$= B_t\, (1 - \exp(-\tau_{NAD}))$$
$$A_t \leq 1 \qquad \textbf{Eq 33b}$$

Note: when computing A_t, values exceeding unity are set to unity; values lower than zero are set to zero.

The rationale of this equation is that as the vegetation biomass increases, both (i) attenuation of soil emission and (ii) vegetation emission increase, making the effective temperature closer to the vegetation effective temperature. Conversely, for bare soil conditions (i.e. for LAI=0), T_{gc} is equal to T_g. When θ increases, T_{gc} becomes closer to the vegetation temperature as attenuation by the vegetation increases due to the $1/\cos(\theta)$ dependence. However, in equations (22a-b), this dependence was not considered; simulations showed that this simplified equation remains accurate for most applications.

The above approximate equation is derived from the radiative transfer equation of the τ-ω model. In this equation, T_{gc} is assumed to be a linear function of T_c and T_g and the weighting parameter A_t is assumed to depend on τ_{NAD}. The coefficient B_t used to compute A_t is assumed to depend on the canopy type. Simulations made with the τ-ω model for a large range of values of optical depth, soil and vegetation temperatures and incidence angles, provided an estimate of the default value of B_t: $B_t = 1.7$. As the temperature difference ($T_{gc} - T_g$) is small over low vegetation covers, we can use approximate τ_{NAD} values estimated from default LAI values.

- B_t = function (canopy type)

- default value of B_t is $B_t = 1.7$ (derived from the τ-ω model and for rather general conditions)

- τ_{NAD} computed as a function of canopy type and default LAI value.

For the time being, **it has been decided to select option 1 for the prototype and the operational processor**. Later, the introduction of a switch might be considered to test the Tgc option 2.

---

[7] It is probably not so valid for the afternoon orbit. Impact TBD from real data

- **Scattering effects**: at L-band, the value of the single scattering albedo ω is found to be rather low. For specific crop types (such as corn), ω can reach a value close to 0.1, but for most of low vegetation types, ω is lower than 0.05 and is neglected in most studies [58]. As the dependence of ω on θ could not be clearly demonstrated to date in the literature, it will be neglected in the algorithm. The value of ω will be given in the algorithm as a function of the vegetation type.

  - ω_P = function (canopy type)
  - The default value of ω, which was found to be valid over most types of crops will be $\omega_V = \omega_H = 0$ [58]. It is likely that the dependence of ω_P on polarization is rather low for most of low vegetation canopies.

- **Optical depth**: several studies found that τ_P could be linearly related to the total vegetation water content VWC (kg/m^2) using the so-called b_P parameter according to $\tau_p = b_p$ VWC [59].

  At 1.4 GHz a value of $b_P = 0.12$ +/- 0.03 was found to be representative of most agricultural crops. Recent studies (presented below) found good correlation between τ_p and vegetation indices (such as NDVI) or Leaf Area Index (LAI). In dry conditions (without interception effects), the ratio τ_p / LAI computed over a fallow and several crops from both SMOSREX and INRA experimental data sets ranged from 0.06 to 0.08. A detailed description of the computation of optical depth τ accounting for green vegetation, litter and intercepted water is given in the following.

  - **Modelling the effect of vegetation structure on optical depth**: it was found that τ_p depends on polarization and incidence angle, especially for vegetation canopies with a dominant vertical structure (stem dominated canopy such as cereal crops). Wigneron et al. [60] proposed a simple formulation using a polarization correction factor C_{pol} to parameterize this effect and compute the optical depth for cereal crops:

    $$\tau_H(\theta) = \tau_{NAD} \qquad \textbf{Eq 34}$$

    $$\tau_V(\theta) = \tau_{NAD} [\cos^2\theta + C_{pol} \sin^2\theta] \qquad \textbf{Eq 35}$$

  Within a large-scale SMOS scene, it is likely that the effects due to the vegetation structure for a variety of vegetation types are averaged, so that the dependence of τ_p (and ω_p) on polarization and incidence angle can be neglected over most pixels. However, the possibility of accounting for this dependence should be kept in the algorithm to be used possibly over pixels with rather homogeneous vegetation cover. Thus, **a generalization of these equations valid only for crops with a vertical structure has been developed** and will be given in the following.

  - **Interception**: recent results have shown that the effects of the interception of water by the vegetation canopy may be very significant (optical depth τ may increase by a factor of two or three during and after rainfalls over a fallow for instance). Accurate modelling of these effects is not known. Indexes flagging these events, during which soil moisture cannot be retrieved, are developed currently and could be possibly used in the algorithm.

  - **Litter**: even though it is not well known, it is likely the effect of litter is very significant (see 3.1.2.10.1). For instance, this effect was probably the implicit reason for using very high b_P values ($b_P \approx 0.4$) over natural vegetation cover such as prairies.

- Detailed description of the modelling of the vegetation optical thickness:

  To model the optical depth τ_P we propose accounting for the effects of the standing vegetation cover, litter and water intercepted by the vegetation cover after rainfall or dew events:

  $$\tau_P = \tau_{SP} + \tau_L + \tau_{IP} \qquad \textbf{Eq 36}$$

  where τ_{SP} is the optical depth of the standing vegetation cover, τ_L is the optical depth of all the vegetation materials laying at the bottom of the canopy (including litter mainly), τ_{IP} is used to parameterize the increase in optical depth due to intercepted water by the standing vegetation canopy (water intercepted by litter is included in the term τ_L). The computation of these three terms is given in the following. Note that in the following equations, all parameters are not currently well defined over a variety of vegetation covers. The experimental L-band microwave studies made currently over a variety of vegetation covers should provide new information to provide realistic values of the vegetation parameters for natural herbaceous covers, agricultural fields, matorrals and forests.

  For the retrieval, we shall consider the nadir value τ_{NAD} (i.e. for the incidence angle $\theta = 0$), including the contributions of the standing vegetation cover, litter and water intercepted by the vegetation cover.

  1) τ_{SP} **is the optical depth of the standing vegetation cover** and includes both green and senescent vegetation materials. τ_{SP} is usually found to be correlated to VWC, but it is very difficult to provide estimates of this latter variable at global scale. Instead, we propose to parameterize τ_{SP} as a function of the Leaf Area Index (LAI).

  There are two main reasons for this: (i) it is much easier to build global maps of LAI from spaceborne remote sensing observations in the optical domain or from SVAT modelling with interactive vegetation [61] than maps of VWC; (ii)

several recent studies have also found good correlation between τ_{SP} and LAI ([62, 63] over a fallow; and over several crops [64].

Note also that P. de Rosnay (2006)[65] obtained good correlation over a natural fallow cover between LAI and green vegetation water content ($R^2 = 0.95$) and also between LAI and the total vegetation water content ($R^2 = 0.86$) (including both green and senescent vegetation materials and litter). It is likely the parameterization of τ_{SP} from LAI, rather than from VWC, will be rather efficient as long as the vegetation is green (in particular during the vegetation growth). This parameterization might be less accurate during the senescence phase (during which τ_{SP} might be underestimated from low LAI values over some vegetation types). It is also less accurate because VWC is height dependent; this is not the case of LAI which is obtained from optical domain observations.

Our objective will be to provide a simple formulation allowing accounting for the dependence of τ_{SP} as a function of polarization and incidence angle. The formulation for the angle dependence should account for the fact $\tau_{SV}(\theta=0) = \tau_{SH}(\theta=0)$. In addition, in the retrieval process, only one variable accounting for the vegetation effects should be retrieved. Therefore, we chose to express $\tau_{SV}(\theta)$ and $\tau_{SH}(\theta)$ as a function of only one variable, namely $\tau_{S\_NAD} = \tau_S(\theta=0)$, which is estimated as a function of LAI.

The modelling of $\tau_{SP}(\theta)$ will thus be written in three equations as follows:

$$(\theta=0) = \tau_{SV}(\theta=0) = \tau_{S\_NAD} = b'_S \cdot LAI + b''_S \qquad \textbf{Eq 37}$$

$\tau_{SV}(\theta)$ and $\tau_{SH}(\theta)$ will be expressed as function of τ_{S_NAD} according to:

$$(\theta) = \tau_{S\_NAD} (\sin^2(\theta).tt_H + \cos^2(\theta)) \qquad \textbf{Eq 38}$$

$$(\theta) = \tau_{S\_NAD} (\sin^2(\theta).tt_V + \cos^2(\theta)) \qquad \textbf{Eq 39}$$

where the tt_V and tt_H parameters allow accounting for the dependence of τ_{SP} on incidence angle.

These two equations are a generalization of the equation based on the polarization correction factor C_{pol} which was developed for vegetation with a vertical structure: applying C_{pol} to the standing vegetation optical depth τ_{SP} (Eq 34 and Eq 35) corresponds to the particular case: $tt_H = 1$ and $tt_V = C_{pol}$ ($C_{pol} > 1$ for a vertical structure).

A value of $tt_P > 1$ will correspond to an increasing trend of τ_{SP} as a function of θ (as it is the case for τ_{SV} for crops with a vertical structure). A value of $tt_P < 1$ will correspond to a decreasing trend of τ_S as a function of θ. The particular case, $tt_H = tt_V = 1$, will correspond to a case where τ_{SP} is assumed to be independent of both polarizations and incidence angle: $\tau_{SH}(\theta) = \tau_{SV}(\theta) = \tau_{S\_NAD}$.

Rather than tt_H and tt_V, it may be more convenient to consider the 2 variables tt_H and $\textbf{R}_{tt} = tt_V / tt_H$.

It will be assumed that all vegetation parameters b'_S, b''_S, tt_H and tt_V are function of the canopy type

$$, b''_S , tt_H , tt_V = \text{function (canopy type)} \qquad \textbf{Eq 40}$$

In the above equations we will thus neglect the dependence of b'_S and b''_S on (i) the canopy hydric status ([66]; [67]) (ii) the change of the vegetation structure in relation with phenology ([68]). This dependence was shown to be relatively significant over crops, especially during senescence, but it is likely that it has a low impact over large mixed pixels.

The default values of tt_H and tt_V, which will be valid over most types of vegetation canopies where it is likely that the dependence of τ_{SP} on incidence angle and polarization can be neglected, are $tt_H = tt_V = 1$.

To compute **default values** of τ_P, the corrections using the tt_P formulation are applied to the optical depth of the standing vegetation (τ_{SP}) **only.** Conversely, in the retrieval process, the tt_P formulation will be applied to the whole retrieved τ_{NAD}, which includes attenuation by **both** the standing vegetation and litter. This latter option was considered to simplify the retrieval process, even though, strictly speaking, the tt_P correction should not apply to the litter optical depth.

    2) τ_L **is the optical depth of the layer of litter,** i.e. **dead vegetation laying on the ground surface** at the bottom of the vegetation layer. The following litter modelling is still under investigation and **will not be used in the operational processor** until further validations are fully done. However, for testing purposes, the litter effect can be activated or deactivated in the prototype processor using a switch.

As found in [69], litter effects can be partly accounted for by using a formulation accounting for the dependence of HR on SM, which is considered in the operational processor and was tested over both a coniferous forest [70] and various types of grass covers [69].

For some vegetation types, this layer is a litter including mainly dead vegetation materials (senescent leaves and needles in forests for instance, as well as many crops). For some other vegetation types, corresponding to fallows or natural herbaceous covers, it is sometimes difficult to distinguish between the litter and the green or senescent vegetation standing at the bottom of the canopy. In this later case, τ_L corresponds to the optical depth of a low vegetation layer including both green and dead vegetation material. However, we think that this layer should be distinguished, since its attenuation properties might have a very different behaviour from that of the standing vegetation cover [71-73].

First, this layer includes mainly dead or senescent vegetation materials which have a very high retention capacity of intercepted water [74]. Rainfall water intercepted by this layer evaporates generally at a much lower rate (daily) than that intercepted by the standing vegetation (which evaporates on an hourly basis). Thus, the water content of this layer is strongly dependent on the rainfall events and is generally closely related to the soil moisture content, contrary to that of the standing vegetation water content. Second, this layer is very dense in terms of volumetric fraction $\mathbf{F_{VOL}}$ (m^3/m^3) (i.e. volume of vegetation material (m^3) per volume of the vegetation layer (m^3) while the volumetric fraction of the standing vegetation layer is generally very low ($F_{VOL} \approx 3.\ 10^{-3}$ for a soybean crop; [75]). As the attenuation properties of the vegetation media are strongly related to its volumetric fraction F_{VOL}, we think that, for a given amount of vegetation biomass, the attenuation properties of the litter should be very different from that of the canopy. In particular, coherent scattering effects [76], which can be generally neglected in a vegetation canopy with $F_{VOL} << 1$, may be significant in a litter, which is a dense vegetation medium.

As the optical depth of the vegetation material is generally related to its total water content, we propose the following equation:

$$= c_L\ LWC \qquad\qquad\qquad \textbf{Eq 41}$$

where c_L is a coefficient characterizing the attenuation properties of the litter medium; and LWC is the amount of water included in the litter layer (kg/m^2). As litter and mulch have probably isotropic attenuation properties, this parameter will be assumed polarization independent. The coefficient c_L will mainly depend on the characteristics of the litter (density, material type, etc.) which will depend mainly on the canopy type. LWC is a function of the dry biomass of the litter and of its moisture content (%). Estimates of the dry biomass of litter can be evaluated as a function of the canopy type. Possibly, estimates of the litter moisture content (%) could be evaluated from the soil moisture value.

$$= \text{function (canopy type)} \qquad\qquad\qquad \textbf{Eq 42}$$

$$C = \text{function (canopy type, litter moisture content (\%))} \qquad\qquad\qquad \textbf{Eq 43a}$$

The value of c_L has been estimated from microwave measurements acquired over a fallow during the SMOSREX experiment [62]. τ_L was computed as a function of the retrieved value of τ_P minus the estimated value of the standing vegetation optical depth (τ_{Sp}), when there was no intercepted water. There was a clear correlation between computed values of τ_L and LWC ($R^2 = 0.61$, 10 measurement values) and the estimated value was $c_L = 0.24$. This value is about twice that of b_P, confirming possibly higher attenuation properties for litter than for standing canopy. The average value of LWC in absence of rainfall over the fallow was $0.5\ kg/m^2$.

Preliminary default value of c_L will thus be set to $c_L = 0.24$.

The amount of water included in the litter layer (LWC, kg/m^2) will be computed as a function of its moisture content ($Mg\_L\ kg/kg$), which is the ratio of the litter water content (kg) to the total litter weight (kg)) and of the litter dry biomass (B_{S}_L, kg/m^2) as follows:

$$C = [Mg\_L\ /\ (1-Mg\_L)]\ B_{S}\_L \qquad\qquad\qquad \textbf{Eq 43b}$$

A default value of the litter dry biomass estimated from SMOSREX data (fallow) is $B_{S}\_L = 0.3\ kg/m^2$.

Preliminary results obtained during the SMOSREX experiment indicated that Mg_L could be related to the soil moisture content (SM) following an approximate linear relationship:

$$\_L = a\_L\ .\ SM + b\_L$$
$$0 \le Mg\_L \le 0.8 \qquad\qquad\qquad \textbf{Eq 43c}$$

Note that we will consider that the litter moisture content (Mg_L, kg/kg) is lower than 0.8. Thus, when computing Mg_L, values exceeding 0.8 will be set to 0.8 (Mg_L=0.8) and values lower than zero will be set to zero (Mg_L=0).

Results obtained from SMOSREX will be used to define default parameters a_L = 2.33 and b_L = 0. (corresponding to a litter moisture content of 70% for a soil moisture SM = 0.3 m^3/m^3).

In summary, the optical depth of the litter layer τ_L will be computed as a function of the following vegetation parameters: c_L, a_L, b_L and B_S_L and of soil moisture (SM). The four vegetation parameters (c_L, a_L, b_L and B_S_L) will be assumed to **depend on the vegetation type** only and will be provided a priori in a table as a function of the vegetation type index. Default values of these parameters were estimated from the SMOSREX experiment (fallow vegetation).

3) τ_{IP} is the optical depth that parameterizes the effect of **intercepted water** by the standing vegetation canopy, due to rainfall or dew events. Results obtained over a senescent wheat canopy showed that for moderate amount of intercepted water (less than 1.5 mm intercepted water), the L-band measurements remained very sensitive to soil moisture, and simultaneous retrievals of both soil moisture and optical depth were possible (so-called 2-P for 2-Parameter retrieval approach) [66]. Therefore, it is likely that during dew events, during which the amount of intercepted water is relatively small (~ 1 mm), 2-P retrievals are possible. However, recent results [77] obtained over a fallow (SMOSREX experiment) showed that the effects of intercepted water might be very significant over some vegetation canopies. The optical depth may increase by a factor of two or more after rainfall events. The water can be intercepted by the green vegetation material (the water is mainly on the surface of the vegetation elements) or by senescent or dead standing vegetation elements (the water is mainly absorbed by the dead vegetation tissue). Over some natural vegetation covers, this fraction of dead or senescent vegetation elements may be significant.

An attempt to parameterize τ_{IP} would require estimations of the interception reservoir (mm) and of the fraction of intercepted water. However, estimating the fraction of intercepted water, which depends on the intensity of the rainfall events vegetation type and evaporation fluxes, would be very difficult.

Rather than attempting to parameterize the interception events, we propose to use an index **flagging** events during which interception effects are very significant (and during which it is very likely that soil moisture cannot be retrieved). Over fallow [77] showed that one of the best indices that can be used to flag interception at **local scale[8]** is the **<u>observed polarization ratio</u>** PR = (TB_V - TB_H) / (TB_V + TB_H) at rather large incidence angle ($\theta \approx 50°$). Significant interception events are associated to **low** values of PR. We propose setting a **threshold** so that low PR values will correspond to a high probability of significant interception events (see section 3.2.5.6.1). The threshold will be possibly later parameterized as a function of the vegetation type and the geographical location. However, it should be emphasized that this field of research is quite new and very few results are available to date to develop accurately this parameterization.

In summary, the vegetation type and the Leaf Area Index (characterizing the vegetation phenological stage and thus indirectly the vegetation structure) are the main parameters determining the values of the parameters used in the ***τ-ω* model**: b's, b"s, tt_V, tt_H, ω and the intensity of specific effects such as litter and interception.

See also references [61, 66, 67, 71-76, 78]

3.1.2.8 Forests

Application of the algorithm over forests is accomplished by keeping in mind the considerations indicated below.

- A large fraction of land is covered by forests. All efforts aimed at fully exploiting the potential of SMOS over these areas must be done. Although it is commonly believed that forest crowns are opaque, some experimental and theoretical studies [79, 80] [81] indicate that the situation is more complex. At higher frequencies (C band and above) there is a strong contribution of leaves to crown attenuation. At L-band, leaves are almost transparent, attenuation is mostly due to branches, and soil contribution can be still appreciable, unless if the forest is dense.

- A simple empirical approach, based on τ and ω parameters fitted over experimental data is not appropriate to forests, due to two main reasons.

    1. Only a limited amount of experimental brightness data is presently available at L-band. Therefore, a statistically reliable fitting is not feasible. Several radar signatures would be available, but a direct reuse of these signatures to derive emission parameters is not reliable, due to some basic differences between emission and backscattering processes, as demonstrated in [81].

---

[8] At SMOS scale the approach will have to be validated

2. In forests, emission/scattering processes are complex, since trunks and branches are not small vs. λ. Therefore, multiple scattering effects are appreciable, and a simple first-order approach is not reliable.

- Anyhow, the methodology to be adopted for forests must be harmonized with the general SMOS retrieval algorithm, and the complexity of the operational procedure must be kept limited.

By keeping in mind these considerations, the adopted approach is summarized below (details are available in [82]).

- Forests are initially included in the surfaces for which soil moisture retrieval is attempted over low vegetated or bare soil surfaces but with specific forest parameterization.

- From the operational point of view, the simple approach based on "albedo" and "optical depth" is kept.

- From land cover classes, 3 forest categories are aggregated: Needle leaf, broadleaf (including Tropical forests and woodland), mixed forest, woodland. The same general procedure is applied for the 3 categories, although the output parameters are specific of each single category.

- The values of albedo and optical depth are assigned by a preliminary modelling work based on the software already available at TOV, with suitable refinements and adaptation to specific cases [83], [81, 82]. The output of this basic direct modelling work consists of look-up tables, relating sets of simulated emissivities (for the SMOS configuration) to SM, for the 5 forest types indicated above. Using allometric equations available in the literature for the different forest categories, geometrical and bio-physical inputs required by the model are related to LAI_{Fmax}, , LAI_F and LAI_V. LAI_{Fmax} is the contribution of all crown components (tree trunks, branches and leaves) to the total optical thickness of an elementary surface of forest. It is obtained from the maximum yearly value of the forest LAI, whereas (LAI_{Fmax}-LAI_F) is the time dependent contribution of tree leaves to this total optical thickness, and LAI_V is the time dependent contribution of low vegetation understory to this total optical thickness [82],[84-86] . These three quantities partition the total forest optical thickness into three contributions and do not represent absolute LAIs but fractional LAI. Then, a standard RMS minimization routine is used to find the equivalent values of the parameters (optical depth and albedo) to be assigned to a simple first order model, like the one adopted for low vegetation, in order to behave most similarly to the discrete multiple scattering model. This RMS minimization is made by considering, for each forest scenario, several sets of angles and SM values. This operation is named **parameterisation.** Details are given in [82]. The output produced after this step consists in estimating the albedo and relating the nadir optical depth to LAI_{Fmax}, LAI_F and LAI_V,, with coefficients depending on forest type. These two forest parameters (**equivalent** nadir optical depth and albedo) are indicated by τ_{F_NAD} and ω_F, respectively. It is found in [82] that, due to the various orientations of branches and leaves, τ_{F_NAD} and ω_F may be assumed to be independent on polarization.

- With the two values obtained by the previously described parameterisation, the successive algorithm steps are basically similar to the low vegetation case. The basic formulas described in Sections 3.1.2.2 and 3.1.2.7 are used also for forests, leading to a unified approach. In particular, a simple formula is used to compute the nadir equivalent optical depth of standing vegetation $\tau_{F\ NAD}$, such as:

$$\tau_{F\_NAD} = \tau_{FA} + b_V \cdot LAI_V \qquad \textbf{Eq 44a}$$

  - where:

- $\tau_{FA} = b'_F \, LAI_{Fmax} - s_F \, [(LAI_{Fmax} - LAI_F) / LAI_{Fmax}]$ **Eq 44b**

A simplified version of Eq 44b that considers $s_F=0$ can be adopted. It requires only two auxiliary data files, the LAI_{Fmax} introduced above and the LAI introduced in low vegetation modelling:

$$\tau_{F\_NAD} = b'_F \cdot LAI_{Fmax} + b_V \cdot Fv.LAI \qquad \textbf{Eq 44c}$$

Note: in this simplified version F_V is the fraction of the LAI that corresponds to the understory contribution. i.e. LAI_{Fmax}, LAI_F and LAI_V are fractional LAI whereas LAI is an absolute one: Fv.LAI approximates LAI_V

b'_F, b_V and s_F values, specific of the forest categories, are obtained. It is also assumed that ω_F does not depend on LAI_{Fmax}.

The brightness temperature may be finally computed as:

$$T_{bp} = (1-\omega_F)\,(1-\gamma)\,(1 + \gamma\, r_{gp})\, T_c + (1-r_{gp})\, \gamma\, T_g \qquad \textbf{Eq 45}$$

where T_g and T_c are the effective soil and vegetation temperatures, r_{gp} is the soil reflectivity, ω_F is the equivalent albedo, and γ is the vegetation transmissivity, given by:

$$\gamma = \exp(- \tau_{F\_NAD} / \cos \theta) \qquad \textbf{Eq 46}$$

Since the contribution of forested fractions differs highly from low vegetation radiation for thick forests, a **flag** FL_TAU_FO can be established for pixels with values of τ_{F_NAD} exceeding a given threshold TH_SCENE_TAU_FO.

As previously stated, the basic algorithm for forests is similar to the one used for low vegetation. The main differences are listed below:

- A simple τ_{F_NAD} constant, without correcting factors depending on polarization and angle, may be used in Eq 46. This is a result of the variability in orientation of branches and leaves.
- ω_F may be considered constant (i.e. independent on angle, polarization and time). However, it is **not** negligible, since its value is 0.08 (see [82])

τ_{F_NAD} includes all contributions due to crown, litter and understory [85, 86]. The parameterization work led to the following values for the coefficients[85, 86][9]:

b'$_F$ = 0.11, b"$_F$ =0.03 for needle leaf, mixed forest

b'$_F$ = 0.18 b"$_F$ =0 for evergreen broadleaf, savannah

b'$_F$ = 0.13 b"$_F$ =0.05 for deciduous broadleaf

ω_F = 0.06 in all cases.

3.1.2.9　　　　　Summary of vegetation parameters

Table 9: Parameters for: (a) low vegetation cover; (b) forests cover

Surface TB of vegetation		Input/Parameter Name	Range	Units
Soil surface emission	r_{gp}	soil surface reflectivity	0 – 1	
Computing vegetation attenuation	γ_P	vegetation attenuation factor	0 – 1	
Optical depth of the different components of the canopy	τ	optical depth of vegetation (including standing vegetation, litter and intercepted water)	0 – 3	
	τ_{NAD}	value of optical depth at nadir (i.e. for incidence angle $\theta = 0$)	0 - 3	
	τ_S, τ_{S_NAD}	optical depths of the standing vegetation canopy (all the canopy but excluding the litter and the intercepted water)	0 – 3	
	τ_L	optical depth of the litter layer	0 – 3	
	τ_I	increase in optical depth due to intercepted water in the standing vegetation canopy	0 – 3	
Computing the optical depth (τ_S) of the standing vegetation canopy from LAI	b'$_S$	parameter of the relation τ_S / LAI (for θ=0)	0.01- 1	m^2/m^2
	b"$_S$	parameter of the relation τ_S / LAI (for θ=0)	0. - 3	m^2/m^2
	tt$_H$	angular correction parameter at H polarization (accounting for the dependence of τ_{SP} on incidence angle).	0.1 - 15	
	Rtt (=tt$_V$/tt$_H$)	ratio of angular correction parameters	0.05- 20	
	LAI	Leaf Area Index	0-10	m^2/m^2
Computing the optical depth (τ_L) of the litter layer	LWC	Water content of the litter	0-50	kg/m^2
	c$_L$	attenuation coefficient of litter (c = τ_L / LWC)	0.01- 1	m^2/kg
	B$_S$_L	dry biomass of litter	0-50	kg/m^2
	a_L	parameters used to compute the litter moisture content (Mg_L) from soil moisture SM	0 - 10	
	b_L		0 , 1	
	Mg_L	moisture content of litter	0-0.9	kg/kg
Modelling scattering effects within	ω_H	Single scattering albedo at H polarization	0.- 0.2	

---

[9] These coefficients are being currently improved by ESL during the Cal/Val. Probably not the final values

Surface TB of vegetation		Input/Parameter Name	Range	Units
	ω_V-ω_H	difference of albedo at H and V polarization	-0.2- 0.2	
Accounting for the effects of temperature	A_t	Weighting temperature parameter used to compute T_{gc}	0-1	
	B_t	parameter used to compute A_t as a function of τ_{S_NAD}	0.1 - 5	
	T_c	Vegetation temperature	270-330	K
	T_g	Effective soil temperature (surface + deep)	250-350	K

In the following, we listed some default values of the vegetation parameters. These values are given to help the reader to have a better understanding of the vegetation model, and make it run easily for a typical vegetation canopy with moderate biomass.

A moderate amount of vegetation corresponds approximately to a level of Leaf Area Index of 4 (the order of magnitude of LAI is roughly twice that of VWC). Using a value of 0.06 for the *b'* parameter (typical for crops) and b"=0, the default value of τ is about 0.24.

Default values we propose are thus,

- optical depth $\tau_V \approx \tau_H \approx 0.24$

- parameter $b'_S = 0.06$

- parameter $b"_S = 0.0$ (then τ_S is proportional to LAI)

- angular correction parameters: $tt_H = tt_V = 1$ (optical depth does not depend on polarization and incidence angle: $\tau_{SV}(\theta) = \tau_{SH}(\theta) = \tau_{S\_NAD}$)

- single scattering albedo $\omega_V = \omega_H = 0$

- attenuation coefficient of litter: $c_L = 0.24$ m^2/kg

- dry biomass of litter BS_L = 0.3 kg/m2

- coefficient (a_L, b_L): a_L = 2.33, b_L=0

- weighting temperature parameter $B_t = 1.7$ (leading to A_t=0.34 for default value of τ_{S_NAD} =0.24)

Table 9 (b): forests parameters

Surface TB of vegetation		Input/Parameter Name	Range	Units
Soil surface emission	r_{gp}	soil surface reflectivity	0 - 1	
Computing vegetation attenuation	γ	vegetation transmissivity	0 - 1	
Optical depth	τ_{FNAD}	Nadir optical depth	0 - 3	
Optical depth	b'_F	parameter of the relation τ_{FNAD} / LAI$_{Fmax}$	0.01 - 1	m^2/m^2
	$b"_V$	parameter of the relation τ_{FNAD} / LAI$_V$	0.001 - 1	m^2/m^2
	LAI$_{Fmax}$	Maximum yearly value of arboreous Leaf Area Index contribution	0-10	m^2/m^2
	LAI$_V$	Herbaceous (understory) Leaf Area Index contribution	0-10	m^2/m^2
Modelling scattering effects within the canopy	ω_F	Equivalent albedo	0.- 0.2	
Accounting for the effects of temperature	T_c	Vegetation temperature	270-330	K
	T_g	Effective soil temperature (surface + deep)	250-350	K

Default values we propose are:

- nadir optical depth $\tau_{FNAD} = 0.5$

- equivalent albedo $\omega_F = 0.08$

- weighting temperature parameter $B_t = 1.5$ (This value of B_t parameter is a bit lower than 1.7, used for low vegetation. This is an approximate way to consider that litter and understory contribute to the overall τ_{FNAD}, but their temperature is close to soil temperature).

3.1.2.10 Specific issues for nominal case

3.1.2.10.1 Litter

Litter can be present in vegetation canopies, which are not (or rarely) ploughed: prairies or non-agricultural canopies, natural covers, forests, etc. Very few experimental studies have been made over these vegetation types, and modelling of the effects of that layer overlaying the soil should be developed. However, as noted above (see 3.1.2.7), it is likely that the effect of litter is significant.

For instance, very high values of the b_p parameter ($b_p \approx 0.4$) were obtained over natural vegetation covers such as prairies by researchers of USDA and INRA (while usually $b_p \approx 0.12$ over crops). Such high values could be probably related to the attenuation effect of litter that was implicitly accounted for by b_p. Over forests, measurements in Les Landes coniferous forest showed the vegetation water content (VWC) of litter to be close to 10 kg/m^2 in some stands. For such a large amount, it is likely that soil emission is totally attenuated at L-band.

These two examples show the importance of improving our knowledge of litter. Research activities were recently carried out in the framework of the SMOSREX experiment (over a fallow) and two experiments over forests (INRA Bordeaux experiment in Les Landes coniferous forest in 2004, and Zurich ETH experiment over deciduous forest in Jülich, 2004-2005).

For forests, the litter model adopted in section 3.1.2.8, considers the litter as a continuous layer overlying the soil [71]. The thickness may be related to the same LAI_max static parameter used to compute τ_{FNAD} (assuming dry leaves to be the main component), while litter permittivity may be estimated by assuming a given ratio between soil moisture and litter moisture.

In the forward model for brightness temperature, litter effects are then included following the steps indicated below:

- compute permittivity of soil
- compute permittivity of litter as a function of soil moisture, dry weight matter density and assumed ratio between soil moisture and litter moisture
- estimate litter layer thickness as a function of LAI_max and vegetation type
- using simple formulas given in [34], compute the reflectivity of the ensemble soil + litter for flat interface
- apply roughness correction
- include standing vegetation effects

3.1.2.10.2 Dry Sand

In itself, sand is simply a soil type and could be considered as a purely nominal case. However, due to its own characteristics, it has almost no bound water and hence has specific dielectric constant behavior. Moreover, sand has specific water capacities and can be very dry, leading to large penetration depths. Hence, the equations given in **Error! Reference source not found.** are bound to be less accurate as sand proportion increases and should be corrected.

It is often considered that the dielectric constant of sand can be expressed at 1.4 GHz [35]:

$$\varepsilon_{dry\text{-}sand} \approx 2.53 - 0.05\,j \qquad\qquad \textbf{Eq 47}$$

A specific model might be developed from this expression. However, since it is not currently available, in the meantime the one given in equation **Error! Reference source not found.** is to be used for sand areas.

Refer to section 3.7 for future developments.

3.1.3 Open water

3.1.3.1 General case

Most land surfaces include extended water surfaces, which may be the ocean for coastal pixels, or inland features such as rivers, canals, lakes, ponds, flooding etc. To derive a sensible value for soil moisture, these contributions have to be taken into account.

The emission by water bodies is estimated by assuming the validity of the Fresnel Equations [Eq 11] and deriving the dielectric constant of an assumed flat water body. It must simply be noted that in the Fresnel equation the magnetic permeability μ_s should be replaced by μ_w.

The **real** (dominant) and imaginary parts of the complex dielectric constant for free water $\varepsilon_w = \varepsilon'_w - j\,\varepsilon''_w$ at a given radiometer frequency f are given by the modified Debye equation [87],

$$\varepsilon'_w = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w\infty}}{1 + (2\pi r \tau_w\, f)^2} \qquad \textbf{Eq 48}$$

$$\varepsilon''_w = \frac{2\pi\ r\tau_w\ f(\varepsilon_{w0} - \varepsilon_{w\infty})}{1 + (2\pi r \tau_w\, f)^2} + \frac{\sigma_i}{2\pi\varepsilon_0 f} \qquad \textbf{Eq 49}$$

Note that in the following equations (Eq 50a Eq 52e) the **temperature, T, is in °C**.

σ_i is the ionic conductivity for saline water (in S/m) function of temperature and salinity:

$$\sigma_i(S, T) = \sigma_i(25, S) \cdot e^{-\phi(S, \Delta)} \qquad \textbf{Eq 50a}$$

Where $\sigma_i(25, S)$ is the ionic conductivity of sea water at 25°C and is given by:

$$\sigma_i(25, S) = S \cdot \left(ow_{23} + ow_{24} \cdot S + ow_{25} \cdot S^2 + ow_{26} \cdot S^3\right) \qquad \textbf{Eq 50b}$$

And the function ϕ depends on S and $\Delta = 25 - T$

$$\phi(\Delta, S) = \Delta \cdot \left(ow_{27} + ow_{28} \cdot \Delta + ow_{29} \cdot \Delta^2 - S \cdot \left(ow_{30} + ow_{31} \cdot \Delta + ow_{32} \cdot \Delta^2\right)\right) \qquad \textbf{Eq 50c}$$

For pure water S=0, thus the ionic conductivity is also null, $\sigma_i(0, T) = 0$

The magnitude of the high frequency dielectric constant $\varepsilon_{w\infty}$ was determined by Lane and Saxton [88] to be 4.9.

There are **separate algorithms** for calculating the static dielectric constant ε_{w0} and the relaxation time $2\pi r \tau_w$ of fresh and saline water.

The static dielectric constant of **fresh water**, ε_{w0}, is a function of temperature as described by Klein and Swift [89]:

$$\varepsilon_{w0}(T) = ow_1 + ow_2 \cdot T + ow_3 \cdot T^2 + ow_4 \cdot T^3 \qquad \textbf{Eq 51a}$$

The relaxation time of pure water, $r\tau_w$, is given by Stogryn [90]:

$$2\pi r\tau_w(T) = ow_{14} + ow_{15} \cdot T + ow_{16} \cdot T^2 + ow_{17} \cdot T^3 \qquad \textbf{Eq 51b}$$

For **saline water** with a salinity SAL or SSS = S, the static dielectric constant of water, ε_{sw0}, is given [89] as

$$\varepsilon_{sw0}(S, T) = \varepsilon_{sw0}(0, T)\, a_{ST}(S, T) \qquad \textbf{Eq 52a}$$

with

$$\varepsilon_{sw0}(0, T) = ow_5 + ow_6 \cdot T + ow_7 \cdot T^2 + ow_8 \cdot T^3 \qquad \textbf{Eq 52b}$$

$$a_{ST}(S, T) = ow_9 + ow_{10} \cdot S \cdot T + ow_{11} \cdot S + ow_{12} \cdot S^2 + ow_{13} \cdot S^3 \qquad \textbf{Eq 52c}$$

The relaxation time of saline water, $r\tau_{sw}$, is given by Stogryn [90]:

$$2\pi r\tau_{sw}(S, T) = 2\pi r\tau_w(T)\, b_{ST}(S, T) \qquad \textbf{Eq 52d}$$

$$b_{ST}(S, T) = ow_{18} + ow_{19} \cdot S \cdot T + ow_{20} \cdot S + ow_{21} \cdot S^2 + ow_{22} \cdot S^3 \qquad \textbf{Eq 52e}$$

Coefficients OW1 to OW32 are supplied in TGRD UPF.

Idealised forward/inverse modelling indicates that a 1% (absolute) underestimate in the weighted field of view occupied by water can give rise to a 0.01 m^3m^{-3} error in soil moisture retrieval, in cases of high soil moisture (0.4 m^3m^{-3}) and dense vegetation cover (optical depth 0.6).

3.1.3.2 Rivers

Vector rivers data is available from ESRI's 'Digital Chart of the World' dataset: http://www.lib.ncsu.edu/gis/dcw.html

For most rivers, there is no associated width, and indeed any estimated width would be subject to local weather and tidal conditions; however wide rivers are coded as lakes with an associated area, and in these cases the vector data can be converted to raster to generate open water area estimates.

3.1.3.3 Time dependent water areas

Abnormal retrieval in some areas may allow flooding conditions to be flagged, if other conditions can be disregarded. Potential confounding environmental conditions include:

- The seasonal behaviour of large rivers.
- The presence of very flat beaches, which give rise to highly variable areas of water coverage.
- Large rain events causing significant ponding.
- Areas of extended gravimetric irrigation and / or rice growing areas…etc.
- Wetlands which have specific but related issues (e.g., mangrove stands ...) While some water bodies are rather stable in time, others **fluctuate** significantly like some rivers (e.g. Niger) due to the rainfall patt**ern or ot**her factors (e.g. freezing for the Ob). Some lakes have stable dimensions; others fluctuate with the seasons (e.g. Tchad lake). To go to the extreme, estuaries fluctuate as well (tidal effects) as well as deltas (Okavongo). This may have a significant impact and cannot be addressed with a fixed inland water/land map. It may be noted that ECOCLIMAP flags tidal flats.

Coastal pixels might induce some errors (variable water / wet sand / dry sand limits) and will have to be addressed by **flagging**. This is currently on hold but might have to be tackled depending on the commissioning phase outcome. In that case, it will be a variable water fraction area.

Similarly, **flooding** will have an impact and is not necessarily known from auxiliary data. By flooding we consider here areas which are regularly (seasonally) flooded, the special events are excluded.

If we consider taking into account correctly water bodies, we need to have an **evolving water/ land mask**, which has yet to be found or established. There might be possibilities with MODIS data, but this will have to be addressed. The fall back option is to **identify** areas prone to such events and **flag** them.

Pending further developments (see section 3.7), a flood flag will be set depending on the amount of past local rain.

3.1.4 Non nominal cases

3.1.4.1 Very dry soils, rocky outcrops and other specific surfaces

It has been found that for very dry soils, the behavior of emissivity was not fully in line with the theory described in 3.1.2.2. Consequently, non-linearities are to be expected and corrective factors or adequate flagging will have to be imposed. We are currently investigating the best way to account for very dry soils (the sand particular case is considered in 3.1.2.10.2 and 3.7.1 and is thus not covered here). Several cases can be considered:

3.1.4.1.1 Very dry soils

Very dry soils do have a specific behavior linked to the different roles of bounded versus free water. To account for this we can adapt the dielectric model with one caveat. Wang model shows a discontinuity in the derivative which may make problems. We are currently working on the issue and should the concept be validated a new formulation would be implemented. The principle should not affect the algorithms as the changes can be included either in the dielectric constant formulation or in the surface roughness model.

Moreover, very dry soil might exhibit extreme penetration depth and thus complicate the estimation of the equivalent temperature. As very dry soils are usually i) without vegetation, ii) of little interest for water fluxes, we believe this specific case should only be of concern for very limited applications.

3.1.4.1.2 Rocks and rocky outcrops

Rocks and rocky areas are not well modelled for the time being. They are assumed to behave as very dry soils. Field measurements do not show significant effects from rocks [13]. It is also worth noting that rocks and the like are usually on barren areas or in mountains regions etc… and thus concern only a limited number of cases. Indeed, problems may arise only when a significant amount of surface is covered with rocks (boulders, steep high mountains, cliffs), or when the dry soils or rocky outcrops have very specific signatures. In all those latter cases, the issue will only complicate existing issues and such cases will probably have to be flagged. The algorithms will then be directed towards dielectric constant values estimation.

In [87] permittivity values are given for rocks at 400 MHz and 35 GHz. They range from 2.4 to 9.6. Approximate expressions do exist for rocks (see Weiner's model for powdered rocks for instance) but it does not seem worth the effort to implement them in the level 2 algorithm for the reasons given above. However, a default dielectric constant ε_{rock} should be provided. We suggest:

$$\varepsilon_{rock} = 5.7 - j*0.074 \qquad \textbf{Eq 53}$$

3.1.4.1.3   Other specific soil surface cases

In some instances, the surface will be affected by other factors such as mineral deposits, salted residues (for instance salt lakes or degraded soils from saltwater intrusion) or surface with very specific dielectric constants.

With current knowledge, this can only be addressed with the dielectric approach. Actually, below 10 GHz the ionic conductivity of saline water has a marked effect on the loss factor, and this is used in SMOS for salinity retrievals. However, the exact form of the dependence of the dielectric constant on soil salinity is not well understood, due to the very sparse measurements available.

3.1.4.2   Frozen soils and ice

- Frozen soils cover large areas at high latitudes (and sometimes altitudes). At mid latitude, frozen soil can also be expected in winter, especially for the morning orbit. Experience shows that the dielectric properties of frozen soil are very close to those of dry soil, while vegetation is almost fully transparent [91]. It is often considered that for frozen soils the dielectric constant can be written [92]

$$= 5 - 0.5\, \mathbf{j} \qquad \textbf{Eq 54}$$

- It can thus be expected that the algorithm will deliver a "very dry bare soil" output when soil is frozen. The presence of frozen soil will be identified by this "very dry bare soil" result from the retrieval when other variables such as air temperature, vegetation cover, and retrieved soil temperature are consistent. It should also be borne in mind that frozen ground often shows extreme spatial heterogeneity, complicating the matter. A more sophisticated expression is given for frozen soils in [93] but it was deemed too complex (in terms of necessary input data to be used in the context of the L2 SM retrieval algorithm).

- We consequently believe that the algorithm used over frozen ground might either be the **standard one** (nominal case) with possibly the adding of a flag (when temperatures are low etc) or dielectric constant retrieval one. Effectively, when everything is frozen things should work nominally. Problems may occur when the area is partially frozen (and patchy surface either dry or wet!). Then the frozen surface is modelled with the default frozen ground model and the complementary area undergoes the decision tree retrieval routine.

- The areas of **permanent** ice/dry snow are known, and will be masked out, so that only the dielectric constant is retrieved. (e.g., Greenland, Antarctica, etc). For other areas or in the case of partial ice (mountains, cold lakes) the idea is that above a given threshold the dielectric constant could be retrieved.

  It can be noted however that ice is rather transparent, with $\varepsilon_{ice}"$ being very small ($\varepsilon_{ice}"= 0.1$ in [93] for pure ice) as given in [94]:

$$\approx 3.17 - \mathbf{j}\, \varepsilon_{ice}" \qquad \textbf{Eq 55}$$

3.1.4.3   Snow

Snow covers about 40% of the Northern hemisphere land mass seasonally but has very different dielectric properties depending on its history. Fresh, dry snow is transparent to microwave radiation; however as snow melts its dielectric constant increases dependent upon snow grain size and liquid water content and may be totally opaque (at $T_e \cong 273$ K) when wet. Consequently, the effects of snow are too complicated to be incorporated into the currently proposed algorithm, and areas with significant

snow coverage other than dry snow must be considered as retrievable only in terms of an equivalent dielectric constant. The issue will be in identifying and flagging the snow-covered areas.

See section 3.7 for future developments.

Dealing with snow other than dry is a topic for further research. As a **preliminary** approach, we suggest defining 3 categories for non-permanent snow cover: dry, wet, and mixed or intermediate. They will be defined through comparing an estimate of the snow temperature T_SNOW to a couple of thresholds. If the snow is dry, it will be assumed transparent and ignored; if it is wet, it will be assumed opaque and subject to a possible retrieval of the dielectric constant of the snow-covered zone. In the intermediate case, a default equivalent dielectric constant will be retrieved for the whole land area. The basic input will be ECMWF until more efficient data from NSIDC SSM/I is available. Thus, ESL and cal/ Val teams are expected to test NSIDC data as part of Cal Val studies. Should they prove to be more suited and / or more accurate, the data source will be changed accordingly. To provide an alternative to ECMWF forecast data, a specific snow map fraction on the DFFG is provided to hold such new snow data associated with a fall back mechanism to ECMWF snow forecast standard use in case of unavailability.

3.1.4.4   Sea Ice

Obviously, sea ice should not be part of the SM processor, but it was identified that neither the SM nor OS processor were covering this type of surface. After some iterations it has been decided that Sea ice will be processed by the ocean processor. However, some land classified pixels may contain sea ice, so it has to be considered in the decision tree and related models.

A sea ice surface is seen in Level 2 landcover as saline water with added rules to handle non-permanent conditions effects (NPE) that may transform this saline water into ice (see section 3.2.3.2.2). Therefore, sea ice is modelled as standard ice (see section 3.1.4.2).

Provided that, 1st) ECMWF information on sea surface temperature (SST) and/or sea ice fraction (CI) is given, and 2nd), the DGGs being fully or partly ocean covered are transmitted to L2 processing, then sea ice will belong to the L2 process with no special action.

3.1.4.5   Urban

Urban areas are the most complex. They include variable mixtures of bare soil and vegetation areas, with buildings. Buildings can be considered similarly to rocks or soils depending on the material used for roofing or even more complex with metallic material ... Moreover, the structures are organized in space with geometrical shapes. And finally, roads (sometimes with trees) and RFI (see 3.1.6) might also influence the signal.

IGBP maps should enable to flag all the large towns. Some like Los Angeles Ca cover several SMOS pixels. Smaller towns and villages may probably and hopefully be innocuous in the retrieval. This assumption will be validated after launch. The current classification seems to refer only to purely man-made surfaces as urban. Consequently, instead of having the generally admitted 3% of land surfaces "urban" we have almost none. This might have to be improved for the decision tree.

However, this is still a placeholder. As we do not have models yet for cities, the cities will be assumed to be **barren soil** for a start, and the surface assumed to be similar to rocks. As much as possible, the concerned areas will be restricted to dense urban areas (including airports), while more sparsely populated suburbs will be considered as vegetated regions.

Because of the uncertainty in material properties, knowledge of the urban fraction for a given area does not allow the effect to be modelled but allows the possible effects to be predicted. It can be shown that if about 11% of a field-of-view is covered with ideally modelled very dry bare soil, then over a range of scenarios (covering soil moisture of 0.1 m^3m^{-3} and 0.4 m^3m^{-3} and vegetation optical depths 0.0, 0.2 and 0.6), soil moisture retrieval exceeds 0.04 m^3m^{-3}. In the case of a highly emissive black-body covering part of the field-of-view, 16% can be covered before the equivalent overall retrieved soil moisture exceeds 0.04 m^3m^{-3}. A likely range for an urban threshold setting a flag indicating a retrieval outside the limits of the user requirements is therefore around 11%.

For high urban coverage, i.e. above a "high" threshold (few cases to expect), the retrieval will switch to the dielectric constant configuration.

3.1.4.6   Topography

The process of retrieving soil moisture and vegetation opacity relies on the use of angular signatures. Obviously, it is necessary to have a reference angle. So, an inclined surface may behave quite differently as a function of azimuth viewing with respect to the same but "horizontal" surface. At SMOS scales, we will never encounter such inclined surfaces but in mountainous area the

pixel will present different facets for varying slopes and azimuths, inducing effects which may eventually render the inversion impossible. Added to this, are the shadowing and adjacency effects.

Two previous studies [95-99] tried to cover the issue of topography. Currently it seems that up to a certain level the almost ever-present topography can be totally neglected (gently rolling hills to pre-mountains). There is then a range of topography characteristics for which the algorithms should be able to retrieve some values but with larger error bars or little significance (old and eroded mountains, mountains with plateaus, etc…). It corresponds to what we call "soft topography".

Finally, very rugged mountains (strong topography) will cause the signal to be useless.

Several aspects must be noted at this point:

- The effect of topography should not be confused with altitude as these two features have very different impacts on the signal. As an example, the Tibetan plateau or the Grand Canyon are rather high but with negligible topographic effects while the centre of the Pyrenees or Scotland may be rather low but with very significant topography effects.

- Mountains are also very often characterized by geomorphologic features (general orientation) which may induce other effects (azimuthally anisotropy).

- Finally, mountains are most often characterized by altitudinal and exposition features with gradients in moisture, vegetation type and density, rock proportions, snow and ice quantities. By their spatial distribution highly correlated to the topography itself, these features will also contribute.

The proposed approach for topography [100], is to process once and for all a global DEM so as to have for every node a descriptor of the topography (topography index). From the values of the indicator, points will be either processed (normal case) or processed with a flag affixed (caution flag) or flagged as mountainous and then processed only for equivalent dielectric parameter.

It is also expected that when real SMOS data are available, the thresholds will be refined after analysis of actual measurements.

Currently, we have defined two approaches to qualify topography. One is based on the slope distribution factor, the other one on the variograms.[101]

The second approach seems the most promising and was tested over France to fully assess the method. As satisfactory, we processed a global DEM (GTOPO30) coupled with a high-resolution one where available (the Shuttle Radar Topography Mission SRTM is not available for high latitudes). From this a 1 km map is produced giving a topography index with 3 values (too much topography, topography that can be accounted for with larger error bars on the retrievals, no noticeable topographical effects) and a mean altitude value.

The idea is that when topography is relatively high, we could afford a much larger proportion of surface before switching to MD models (see next section).

This map could eventually be updated once the satellite is delivering data and our approach fully validated.

3.1.4.7   The cardioid model

In the cases of vegetated soil as well as open water, the basis of physical modelling consists of writing the reflectivity (or emissivity) for a smooth surface as a function of the complex dielectric constant $\varepsilon = \varepsilon' - j\,\varepsilon''$. In turn, the dielectric constant is written as a function of physical parameters, including surface soil moisture for the vegetated soil or salinity for open water.

For cases where ε cannot be expressed in the same way (e.g. iced surfaces), it is still possible to retrieve, from SMOS data, information about the dielectric constant.

It has been shown [102] that, to a very good approximation, ε can be written:

$$\varepsilon' = A\_card\,(1 + \cos(U\_card))\,\cos(U\_card) + B\_card$$
$$\varepsilon'' = A\_card\,(1 + \cos(U\_card))\,\sin(U\_card)$$

Eq 56a

When A_card is constant and B_card is taken equal to 0, this parameterized expression reduces to a cardioid. Hence the name of "modified cardioid".

Or conversely:

$$A\_card = m\_card^2 / (m\_card + \varepsilon' - B\_card) \qquad U\_card = \tan^{-1}(\varepsilon''/(\varepsilon'-B\_card))$$
$$\text{with: } m\_card = ((\varepsilon'-B\_card)^2 + \varepsilon''^2)^{1/2}$$

Eq 56b

The optimal value for B_card is very close to 0.8.

These expressions are relevant because angular dependent radiometric data allow retrieving accurately the value of the magnitude A_card. On the other hand, the retrieval accuracy on the polar angle U_card is extremely poor. Indeed, the emissivity is almost independent of U_card, to the extent that almost any *a priori* value can be stipulated for the angle U_card. In this situation, while retrieving both ε' and ε" would result in very large uncertainties, the modified cardioid approach can be understood as a **regularisation** of the retrieval problem.

Therefore, in cases considered above, SMOS data can still be used to derive an estimate for the magnitude A_card, which will be referred to as the **dielectric constant index**.

This may be useful, as any additional independent information on the dielectric constant can then be used to infer the full complex ε.

For the cases of vegetated soil or open water, values for A_card and U_card may be computed readily from the complex dielectric constant (which is available from the retrieval when retrieved values for SM or SSS are introduced in the direct model), using the above equations (Eq 56 & Eq 57), if necessary. They can then be used as initial values in case a complementary retrieval using the modified cardioid formulation is attempted.

In the decision tree, the notation for the cardioid forward model is MD.

This model is a particular case in the sense that its implementation is not completely identical depending whether it is used for direct simulation or retrieval. In order to apply it:

Once values for ε' and ε" from TGRD relevant LUT have been obtained for each relevant fraction;

- For every fraction such that a **fixed** contribution is needed in the aggregated forward model, this contribution is obtained using directly these ε' and ε" values and then taking care of surface roughness and possible vegetation layer.
- For the non-nominal surfaces over which the cardioid model **retrieval** is carried out:
  - For A_card prior value and U_card reference value two cases shall be considered:
    - Only one non-nominal surface is concerned (i.e. one fraction): then their ε' and ε" are used to obtain the prior value for A_card and the reference value for U_card using equations (Eq 56 & Eq 57). This case is referred later on as **MD** retrieval.
    - Several different non-nominal surfaces are considered (i.e. a group of fractions): then the A_card and U_card default values provided in TGRD UPF are used. This case is referred later on as **MDd** retrieval.
  - In the fwd model for iterative retrieval, A_card is a free parameter; ε' and ε" are computed from A_card, U_card and the constant B_card using Eq 56. The following steps (surface roughness, vegetation layer) are carried out in the usual way. Depending on the decision tree, the vegetation optical thickness may be either a fixed or a floating parameter.
  - Using the retrieved A_card and the constant U_card and B_card, ε' and ε" can be computed and should appear in the output product.

3.1.4.8 Roughness parameterization for non-nominal surfaces

Soil moisture for non-nominal surfaces is not defined and the soil moisture dependent roughness function HR(SM) becomes not relevant. However, we may consider roughness effects on those surfaces that belong to the LAND_COVER_CLASSES table (e.g. rocks, barren, urban areas etc.) and on those that result from non-permanent effects (e.g. surfaces becoming frozen, icy or snow covered, etc.)

This soil moisture dependence can be easily cancelled by setting a constraint in the TGRD' LAND_COVER_CLASSES table such that HR_MAX and HR_MIN have the same value, even possibly 0 when no roughness at all is to be considered.

Consequently, equation Eq 27 and the two sisters equations Eq 28 a & b can be safely used for any kind of surfaces considered in the document. The constraints on HR_MAX and HR_MIN are summarized in the following Table 10; the first yellow row just reminds the standard use of HR(SM) for nominal soil and forest surfaces.

Table 10 : HR_MIN, HR_MAX constrained values for non-nominal surfaces

Surface type		Roughness Parameters		Observations
		HR_MIN	HR_MAX	
nominal, forest		LCC	LCC	When SM is Free, HR (SM, …) is updated for its variation.
barren, urban		LCC	LCC	LCC constraint HR_MIN=HR_MAX
pure, saline water		LCC	LCC	LCC constraint HR_MIN=HR_MAX=0
wetland	perm.	LCC	LCC	LCC constraint HR_MIN=HR_MAX=0 for the time being
	nperm.	LCC #238	LCC #238	HR_MIN, HR_MAX from LCC code #238 (for flooding surface)
frozen soil		UPF Or LCC	UPF Or LCC	HR_MIN, HR_MAX=HR_FRZ from TGRD UPF Or, another possible option HR_MIN=HR_MAX=averaged underlying LCC **HR_MAX** values.
snow		UPF	UPF	HR_MIN=HR_MAX=HR_WET_MIXED_SNOW from TGRD UPF
ice	perm.	LCC	LCC	LCC constraint HR_MIN=HR_MAX
	nperm.	LCC #006	LCC #006	HR_MIN, HR_MAX from LCC code #006 (for water becoming ice)

3.1.5 Other contributions to the radiometric signal

Corrective terms in the radiative transfer equation (RTE) refer to ionospheric (Faraday) rotation and sky and atmospheric contributions. Faraday rotation is taken care of in the geometrical transformation from TOA to antenna (sections 3.2.2.1.3& 3.2.2.1.8); this section deals with the remaining contributions. It must be recalled that they should **not** be considered as corrections to be applied before the retrieval, but as **corrective terms and factors** to be included in the forward models used in the retrieval.

As seen above (Eq 9a), the RTE can be written:

$$TB_P = TB_{atm} + \exp(-\tau_{atm})\,[TB_{atm} + TB_{sk}\exp(-\tau_{atm})]\,R1 + \exp(-\tau_{atm})\,R2 \qquad \textbf{Eq 57}$$

Where aggregated terms R1 and R2, which do not depend on either atmospheric or sky contributions, are dealt with in sections 3.1.2 to 3.1.4.

This section is focused on atmospheric and sky contribution, i.e. obtaining first **TB_{atm} & τ_{atm},** next **TB_{sk}.**

3.1.5.1 Atmospheric contributions

3.1.5.1.1 Physics of the problem

This description makes use of the analysis reported in an ESA study [103], updated using more recent cross section data [104].

3.1.5.1.1.1 Radiative transfer equation

The RTE above assumes that upwelling and downwelling radiations from the atmosphere are equal. This has been verified to be correct within better than 0.01 K for L-band [105].

There are 4 atmospheric components to be considered: dry atmosphere, water vapor, clouds and rain. Ideally, the quantities to be known in Eq 57 (τ_{atm}, TB_{atm}) are the sums of the 4 corresponding contributions.

In every case, the basic quantity from which atmospheric contributions can be estimated is a **lineic absorption coefficient κ**, generally expressed in dB/km.

3.1.5.1.1.2 Dry atmosphere

The radiatively active component in dry atmosphere is **molecular oxygen**. Oxygen molecules have a permanent magnetic moment; therefore, absorption and radiation in the microwave region occur due to magnetic interactions with the incident field. This interaction produces a family of rotation absorption lines in the vicinity of 60GHz (known as the oxygen complex) and an additional isolated line at 118.8GHz [106]. Due to pressure characteristics of the lower part of the Earth's atmosphere, pressure broadening causes the complex of lines to blend together to a continuous absorption band centered around 60GHz.

The oxygen absorption and radiation change due to changes in the meteorological parameters, and depend on the pressure $P(z)$ and the temperature $T(z)$ of the gas as a function of height z.

A model for the absorption by oxygen for lower frequencies is described in [34, 87]. For frequencies below 45GHz, the contribution from the 118.75GHz oxygen absorption line can be neglected, and thereby we only have the contribution from the 60GHz absorption line. Then the lineic absorption from oxygen at f=1.413 GHz can be written in dB/km as:

$$\kappa_{O2} = 1.1 \cdot 10^{-2} f^2 \left(\frac{P}{1013} \right) \left(\frac{300}{T} \right)^2 \gamma \left(\frac{1}{(f - f_0)^2 + \gamma^2} + \frac{1}{f^2 + \gamma^2} \right)$$

Eq 58a

where

- f is the frequency (1.413 GHz)
- $f0$ is the absorption line frequency (60 GHz)
- P is the pressure in millibars (mb) or hectopascals (hPa)
- T is the physical temperature in K
- γ is the line width parameter written in GHz as:

$$\gamma = \gamma_0 \left(\frac{P}{1013} \right) \left(\frac{300}{T} \right)^{0.85}$$

Eq 58b

Where the line width γ_0 is pressure dependent:

According to [34]: $\gamma_0 = 0.59$ above $P = 333$mb; $\gamma_0 = 1.18$ below 25mb, $\gamma_0 = 0.59 \, [1 + 3.1 \cdot 10^{-3} \, (333 - P)]$ in between. However, more recent spectroscopic measurements[104] are better described when choosing $\gamma_0 = 0.59$ over the whole pressure range.

3.1.5.1.1.3 Water vapour

In the microwave region, water vapor has rotational absorption lines at 22.235 GHz and at 183.31 GHz. Furthermore, there are also some absorption lines above this region, which contribute to the microwave absorption spectrum. For calculation of absorption at L-band one can [34, 87], group the contributions from the 183.31GHz and all the absorption lines above in a residual term through the use of low frequency approximation. The resulting absorption coefficient κ_{H2O} can then be written as a sum of the contribution from the 22.235 GHz absorption line κ_{22} and a residual term κ_r:

According to [107]:

$$\kappa_{22} = 2 f^2 \rho_v \left(\frac{300}{T} \right)^{5/2} e^{-644/T} \left(\frac{\gamma_1}{(494.4 - f^2)^2 + 4 f^2 \gamma_1^2} \right)$$

Eq 58c

Where

- ρ_v is the water vapour density (gm^{-3})
- γ_1 is the line width parameter (GHz):

$$\gamma_1 = 2.85 \left(\frac{P}{1013} \right) \left(\frac{300}{T} \right)^{0.626} \left(1 + 0.018 \frac{\rho_v T}{P} \right)$$

Eq 58d

Concerning the residual term, according to [34, 87]:

$$\kappa_r = 2.4 \cdot 10^{-6} f^2 \rho_v \left(\frac{300}{T} \right)^{3/2} \gamma_1$$

Eq 58e

And finally:

$$\kappa_{H2O} = \kappa_{22} + \kappa_r$$

Eq 58f

3.1.5.1.1.4   Clouds

When electromagnetic radiation interacts with particles such as those in snow, clouds, fog and rain, it involves absorption and scattering. However, if only drops, which have a diameter much smaller than the wavelength, are considered – which is the case for 1.4GHz - then scattering at L-band is unimportant, and the absorption coefficient can be calculated from the Rayleigh approximation.

The particles are assumed to be randomly distributed within the volume, and therefore the contribution of the individual particles can be summed assuming an incoherent process.

Furthermore, it is also assumed that the particles are spherical, which is a reasonable assumption for most atmospheric water and ice droplets. The scattering and absorption characteristics of a spherical particle are governed by three factors: electromagnetic wavelength, index of refraction, and particle radius.

Clouds are complex phenomena, which consists of water either in liquid or in frozen form. The amount of water and the phase of the water in the cloud depend on the altitude, the temperature and indirectly on the pressure. Clouds are described by cloud base, cloud top, the mass density of the liquid water in the cloud and principal composition of the cloud. The water content of a cloud is according to [34, 87] typically less than $1g/m^3$.

Radiative effects of ice clouds are negligible at L-band. Concerning liquid water clouds, according to [103] and [34, 87], empirical expressions have been developed by [108] for the lineic absorption coefficient. It appears that the only cases where the overall radiative effect at L-band might not be negligible (>0.1K) concerns deep cumuli. However, there is no reliable auxiliary data allowing estimating a depth for these clouds. In addition, they are mostly associated with rain events, which are dealt with next.

3.1.5.1.1.5   Rain

Physically, rain occurrence is similar to clouds. However, the problem is complicated by several factors:
- Due to the size of raindrops, the Rayleigh approximation is no longer strictly valid, hence a dependence appears with the granulometry of rain, which is variable and not accurately known;
- Large raindrops are not spherical;
- While ice particles do not contribute to atmospheric extinction, there is often a melting zone (just below the 0°C isotherm) which is very poorly predicted and may not be negligible in terms of radiative effects;
- Finally, the rain is often expressed in rainfall intensity, whereas the relevant quantities are lineic densities (liquid water content) in the atmosphere.

For all these reasons, it does not seem practical to correct for rain. According to [103], rain in the atmosphere produces a non-negligible radiative contribution (>0.1K) when the rain intensity exceeds about 10 mm/hr; this is estimated to happen less than 0.2% of the time over all latitudes, up to less than 0.65% of the time over equatorial areas (these figures may be pessimistic for a 06h local time).

Therefore, rain occurrences are a matter for **flagging** rather than correcting. As stated above, the heavy clouds should be associated with rain events.

3.1.5.1.2   Mathematical description of algorithm

3.1.5.1.2.1   Radiative transfer for gaseous components

From the physics, it is concluded that atmospheric contributions must be computed for oxygen and water vapor.

Numerical simulations show that, for L-band, the upwelling and downwelling radiative contributions are extremely close one to each other and can be assumed equal to a single value TB_{atm} in equation Eq 57. Therefore, what is needed is:

$$\tau_{atm} = \tau_{O2} + \tau_{H2O} \; ; \quad TB_{atm} = TB_{O2} + TB_{H2O} \qquad \textbf{Eq 59}$$

Contributions to absorption come from the whole thickness of the atmosphere. However, for oxygen it is not necessary to consider altitudes higher than a level $ZM \approx 30$ km, where absorption becomes completely negligible. For water vapor, the altitude range to be considered is limited to $ZM \approx 10$ km.

Over the required altitude range, the exact computation requests knowledge of altitude profiles for T and P; then, the atmosphere is divided in slices δz. For each slice and for each component, the elementary optical thickness $\delta\tau_G$ (where G is replaced by either O2 or H2O) is computed from the lineic absorption coefficient (expressed in dB km^{-1}) κ_G:

$$\delta\tau_G = 1 - 1/10^{\kappa_G\,\delta z/10}$$

Eq 60a

Where the effect of **incidence angle** θ on optical thickness is introduced:

$$\delta z(\theta) = \delta z_{NADIR}\,/\cos(\theta)$$

Eq 60b

The total optical thickness τ_G is obtained by summing the $\delta\tau_G$ over the relevant altitude range:

$$\tau_G = \sum_{Z=0\to ZM}\delta\tau_G(z)$$

Eq 60c

The radiative contribution TB_G is (taking the upwelling case) computed as:

$$TB_G = \sum_{Z=0\to ZM} T(z)\,\delta\tau_G(z)\exp[-\sum_{Z'=Z\to ZM}\delta\tau_G(z')]$$

Eq 60d

This formulation yields the upwelling contribution. The downwelling contribution is found very close, with differences well below 0.01K.

Since the attenuation through an elementary layer is very small, and the physical temperature variation at this scale is linear, the estimate for the physical temperature T(z) in (Eq 60d) can be taken as the **average** between T values for the bottom and the top of the elementary layer.

3.1.5.1.2.2   Empirical laws for computing atmospheric terms

Three ways are contemplated for computing τ_{O2} and TB_{O2}:

1. Carry out the **integrations** as indicated in equations (Eq 60). The estimated necessary altitude ranges ZM are 20 km for O2, 10 km for H2O; the necessary resolution along the vertical is better than 100m.

2. **Tabulate** the τ_G and TB_G as functions of some parameters (e.g. surface atmospheric temperature T0, the surface pressure P0, some parameter describing the structure of the temperature profile, surface humidity…) and then interpolate from these tables.

3. Build **empirical laws** to compute the τ_G and TB_G.

For the land surface, the required accuracy is estimated to be about 0.2K (circa 0.1% for soil moisture). Then method 1 is not necessary and the algorithm proposed below **uses method 3**.

The most efficient (and physically meaningful) way to do this consists in writing the emission of each component as the product of optical thickness by an **equivalent layer (physical) temperature**, which is conveniently defined by its difference DT_G with the surface air temperature T0:

$$TB_G = (T0 - DT_G)\,\tau_G$$

Eq 61

For dry atmosphere, a **quadratic** fit to results obtained using the whole radiative transfer computation has been found necessary:

$\tau_{O2} = 10^{-6}$ x (k0_tau_O2 + kT0_tau_O2 x T0 + kP0_tau_O2 x P0
   + kT02_tau_O2 x T0^2 + kP02_tau_O2 x P0^2 + kT0P0_tau_O2 x T0 x P0) /

$\cos(\theta)$

Eq 62a

DT_{O2} = k0_DT_O2 + kT0_ DT_O2 x T0 + kP0_ DT_O2 x P0
   + kT02_ DT_O2 x T0^2 + kP02_ DT_O2 x P0^2 + kT0P0_ DT_O2 x T0 x P0

Eq 62b

where
- θ is the incidence angle (radian)
- T0 is the near surface air temperature (Kelvin)
- P0 is the surface pressure (millibar)
- τ_{O2}=is obtained in neper; DT_{O2} is obtained in Kelvin

For the water vapour contribution, a **linear** fit is found adequate:

$\tau_{H2O}=10^{-6}$ x (k0_tau_H2O + k1_tau_H2O x P0 + k2_tau_H2O x WVC) / $\cos(\theta)$
$\tau_{H2O}=\max(\tau_{H2O}, 0)$

Eq 63a

$$DT_{H2O} = k0\_DT\_H2O + k1\_DT\_H2O \text{ x } P0 + k2\_DT\_H2O \text{ x } WVC \qquad \textbf{Eq 63b}$$

where
- **WVC** is the total precipitable water vapour content (kg m^{-2}), available from ECMWF data.
- τ_{H2O} is obtained in neper; DT_{H2O} is obtained in Kelvin.

From values obtained for the DT and τ quantities:
- TB_{O2} and TB_{H2O} are obtained using Eq 61 respectively for O2 and H2O;
- Atmospheric terms TB_{atm} and τ_{atm} to be inserted in the RTE aggregated forward model Eq 57 are obtained using Eq 59.

The numerical values for coefficients in Eq 62a,b and Eq 63a,b are supplied in TGRD UPF. They have been optimized for the ranges P0=[400 1100] mbar; T0= [230 320] K. Note opposite signs between kT0_tau_O2 and kP0_tau_O2, which tend to stabilize the oxygen optical thickness.

In any case, the needed auxiliary data i.e. T0, P0, WVC will remain the same.

3.1.5.1.2.3 Error budget estimates (sensitivity analysis)

The errors induced by the empirical adjustments are well below 0.1K in any case. The major error source will be due to estimates of absorption cross sections, which in turn reflect the uncertainty on spectroscopic measurements. This uncertainty is estimated around 5%, i.e. up to 0.25 K for high incidence angles. This is a permanent bias rather than an uncertainty.

3.1.5.1.3 Practical considerations

3.1.5.1.3.1 Calibration and validation

Since the uncertainty on absorption cross sections cannot be overcome, the resulting error will have to be corrected within the overall SMOS validation process. However, the variation with incidence angle offers a possibility to discriminate among other effects.

Assuming one succeeds in determining correctly the absorption cross sections, the resulting uncertainty would be permanently eliminated.

3.1.5.1.3.2 Quality control and diagnostics

The approximations have been built using the detailed radiative transfer equations.

The following orders of magnitude can be mentioned
- τ_{O2} varies between 0.006 and 0.01 for low altitudes at nadir
- DT_{O2} is around 30 to 32 K
- τ_{H2O} is mostly negligible and may reach at most 0.0003 for low altitudes at nadir
- DT_{H2O} is around 10 to 13 K

The **oxygen** overall contribution is by far the largest atmospheric contribution. It may reach up to 5 K and beyond, as described in [103], for high incidence angles.

3.1.5.1.4 Assumption and limitations

Assumptions are related to laboratory knowledge of spectral properties of atmospheric gases.

Limitations concern the presence of liquid (cloud or rain) water in the atmosphere, for which a **flagging** approach is suggested rather than a correction.

3.1.5.2 Galactic noise contamination

3.1.5.2.1 Physics of the problem

This section has been written using material prepared by J Boutin for the SSS level 2 processor, which, in turn was done with the help of [109] and [110] papers. This common part of the two SMOS L2 processors should thus be fully compatible even though the requirements for land are less stringent than for sea and the reflected signal is weaker in most cases.

At L-band, radiation from celestial sources is strong and spatially variable; they have been reviewed by *Delahaye et al.* [111], *Le Vine and Abraham* [109], and associated corrections needed to interpret L-band radiometric measurements have been thoroughly described by [109]. Radiation originates from three types of sources. The hydrogen line emission corresponds to a

hyperfine atomic transition in neutral hydrogen: the radiation reaches a maximum around the plane of the galaxy, most of the time less than 2 K. The cosmic background is a remnant signal of the origin of the universe and is almost constant in space and time (2.7 K). In addition to the almost constant cosmic background, a very variable (in space) continuum radiation (up to more than 10 K) is due to emissions from discrete radio-sources. Hence, over land the most significant contribution will come from some of the discrete radio-sources and it can be safely assumed that the contributions will be small.

As in the case of atmospheric emission, the cosmic background adds a contribution to the radiometric temperature that depends on the incidence angle linked to the reflection of the signal on the land surface

The two other types of sources add a signal that varies according to the incidence and azimuth angle of the measurement.

3.1.5.2.2 Mathematical description of algorithm

3.1.5.2.2.1 Data conversion

The common practice in passive microwave remote sensing of the Earth is to consider equivalent brightness temperatures. Hence, for L band radiometry, it is common to present data from radio-astronomy surveys in the form of equivalent black-body temperatures, i.e., as if they were from an equivalent thermal source with total power P:

$$P = k_{BC} \; TB \; BD \qquad \qquad \textbf{Eq 64}$$

where k_{BC} is the Boltzmann constant and BD is the bandwidth of the receiver used for the survey, and P is a total power integrated over a frequency range as in the case of the hydrogen line emission.

a) Hydrogen Line emission:

The line emission has a relatively narrow spectrum. However, the line is shifted by the motion of the hydrogen relative to the observer (Doppler shift) and spread by thermal energy of the gas (collisions and vibrations). Leiden / Dwingeloo survey [112] in the Northern hemisphere and IAR (Instituto Argentino de Radioastronomia) survey in the southern hemisphere [113] cover the velocity range from -450 to +400km/s which corresponds to a frequency range of 4.025MHz about the centre frequency of 1.42GHz of hydrogen at rest. The integrated power reported in radio-astronomy survey, P, is given in Kelvin kilometres per second (K-km/s). In order to convert it to a brightness temperature that will be recorded by a radiometer having a bandwidth ΔB, it is necessary to convert it in Kelvin-MHz using the line emission bandwidth and then to calibrate it with respect to the radiometer bandwidth. The standard form for Doppler shift is given by:

$$\nu = \nu_0 (1 - v / c) \qquad \qquad \textbf{Eq 65}$$

with ν_0 the centre frequency (1.42GHz), ν the frequency associated with the velocity v and c the light speed, a velocity range from -450 to + 400km/s corresponds to a frequency bandwidth of 4.025MHz.

Thence the integrated power reported in radio-astronomy survey corresponding to a velocity range of 850km/s, Pint, can be converted in Kelvin-MHz using:

$$\text{Pint (K-MHz)} = \text{Pint (K-km/s)} \times 4.025/850 = \text{Pint (K-km/s)} \times 4.735 \; 10^{-3} \qquad \textbf{Eq 66}$$

Since the SMOS radiometer bandwidth BD_S is close to 20 MHz, well above 4MHz, this value can be converted to get an equivalent Tb for SMOS, as follows:

$$\text{Tb} = \text{Pint (K-MHz)} / BD_S = \text{Pint (K-km/s)} \times 4.735 \; 10^{-3} / BD_S \qquad \textbf{Eq 67}$$

b) Continuum radiation and cosmic background

These radiations are usually given in terms of effective brightness temperature, TB, i.e. they include the correction for the bandwidth of the survey (e.g. Reich and Reich maps [110]). Thence, as these radiations are supposed to be homogeneous over the frequency range of SMOS bandwidth, there is no need to correct TB obtained from radio-astronomy surveys.

3.1.5.2.2.2 Galactic noise reflected towards the radiometer:

In the following, we will consider the effective brightness temperature of the galactic radiation, TB_{sky}, as the sum of the hydrogen emission line plus the continuum radiation plus the cosmic background.

First, it is necessary to determine the location in the celestial sky from which incident radiation will be reflected from one point in the field of view into the antenna. For any given node, we know θ and ϕ (they are respectively the incidence and the azimuth

(0 towards the north, positive eastward) angles of one radiometer measurement toward the satellite, the latitude and longitude as well as the universal time. The specular direction, given by the vector (θ, π+ϕ), points to the location in the celestial sky to consider.

Next, we need to convert (θ, π+ϕ), given in the instrument angle reference (Table 11), into the standard astronomical reference local geographic coordinates, elevation, el, and azimuth Φ (0 toward the south, counted clockwise) using:

$$el = \pi/2 - \theta; \quad \Phi = \phi \qquad \qquad \textbf{Eq 68}$$

Usually celestial maps are given in celestial coordinates system (declination δ, and right ascension, α). It is therefore necessary to derive δ and α from the latitude, lat, longitude lon, local mean sidereal time Θ_L, angles el and Φ. This can be done using:

$$\alpha = \Theta_L - H$$

$$\text{with: } \tan(H) = \frac{\sin(\Phi)}{\tan(el)\cos(lat) + \cos(\Phi)\sin(lat)} \qquad \textbf{Eq 69}$$

$$\sin(\delta) = \sin(lat)\sin(el) - \cos(lat)\cos(el)\cos(\Phi)$$

where H is the sidereal angle (see for instance Appendix C of [109]); Θ_L is the local mean sidereal time at the longitude observed and UTC time U_T and can be computed using the L1c data as described in appendix C of [109].

$$\Theta_L = \Theta_{G0} + \Omega_E U_T + lon$$

$$\Theta_{G0} = 100.46062 + 36000.77\,U_0 + 0.000388\,U_0^2 - 2.6\cdot10^{-8}\,U_0^3$$

$$U_0 = (JD - 2451544.5)/36525 \qquad \qquad \textbf{Eq 70}$$

$$JD = 367Y - INT[1.75(Y + INT[(M+9)/12])] + INT(275M/9)] + C_0$$

$$C_0 = 1721013.5 + D + [HH + MM/60 + SS/3600]/24$$

where Y is the year, and M, D, HH, MM, SS respectively month, day, hour, minute and seconds. JD is the julian date and U_0 the number of julian centuries since the reference epoch (1 JAN 2000). INT means selecting the integer part. Ω_E is the Earth rotation rate.

To have the time, the simplest is probably to extract the acquisition time of the first view of the node and then adding for each view the SNAPSHOT_ID times the elementary integration time. Using a mean time for all the views could also be considered. Similarly, the mean average angle between the two polarizations is also quite adequate.

In the following, we will distinguish two polarizations for T_{sky}. At present, existing galactic maps do not distinguish between V and H pol but there is a polarization dependency, though not yet fully quantified.

a) Assuming a flat land

The galactic contribution reflected towards the radiometer, TB_{sk}, to be introduced in Eq 9, can be computed from TB_{sky}:

$$TB_{sk} = TB_{sky}(lat, lon, U_T, \theta, \phi, p) = TB_{sky}(\delta, \alpha, p) \qquad \qquad \textbf{Eq 71}$$

where p is one of the polarizations (H or V).

b) Taking into account the roughness of the land:

If the sky were homogeneous, it is expected that the introduction of the roughness would have a small effect in most cases: for instance, when the reflection coefficient is modified by about 2.5% (at nadir)) and for a galactic noise of 5K, neglecting the roughness effect would introduce an error of less than 0.1K.

Ideally, it would be necessary at this level to introduce bistatic reflection coefficients, σ_0; in theory, the galactic noise over the whole sky should be convoluted with these scattering coefficients. However, since they are expected to decrease rapidly away from specular reflection, the integration could be done over a narrow interval. This is approximately taken care of by the integration that is necessary anyway due to the finite width of the synthetic antenna beam (see next section).

Keeping this in mind and considering the magnitude of the very maximum galactic brightness temperature we might encounter, (10-15 K) we believe that simplified approach using the direct model values could be used as a proxy (Eq 71) and that no significant spreading will occur.

 In other words, we assume that the reflectivity coefficients estimated through the direct model are deemed sufficient to compute the reflected galactic contribution.

3.1.5.2.2.3 Integration over the antenna beam:

Finally, it is necessary to integrate the reflected brightness temperature over the antenna pattern to obtain the sky contribution to the signal exactly as for a ground element. When the antenna pattern is axially symmetric, according to [109] it is possible to make the integration on δ and α and hence to **precompute** (TGRD) customized galactic maps integrated over an average apodization window after reflection on the surface. As the SMOS lobe varies across the FOV and is not symmetric, this is not accurate but deemed sufficient for our purpose. Views with sky contributions above a TH_SKY threshold will be flagged.

3.1.5.2.3 Error budget estimates (sensitivity analysis)

The main uncertainty is expected to come from inaccuracies of the galactic noise maps. In [110] the authors estimate the accuracy on their maps (due to the calibration of the instrument) to be 0.5K.

In addition to a constant bias, uncertainties are likely to appear on these maps near the equatorial galactic plane. In order to estimate these uncertainties, the SSS ESL have compared the maps derived from the Stokert survey, commonly called the Reich and Reich map, and the ones derived from the Effelsberg survey. Both maps include the continuum radiation and the cosmic background; Stockert survey was performed with a 34mn angular resolution instrument while Effelsberg used a 9mn angular resolution instrument. Stockert map for the northern hemisphere and Effelsberg maps are available on the http://www.mpifr-bonn.mpg.de/survey.html site; the Stokert map for southern hemisphere was provided by ESA. Stockert maps are global but region around Cassiopeia is excluded (no data) and strong sources are suspected to be underestimated; Effelsberg survey is concentrated close to the equatorial plane (Cygnus excluded).
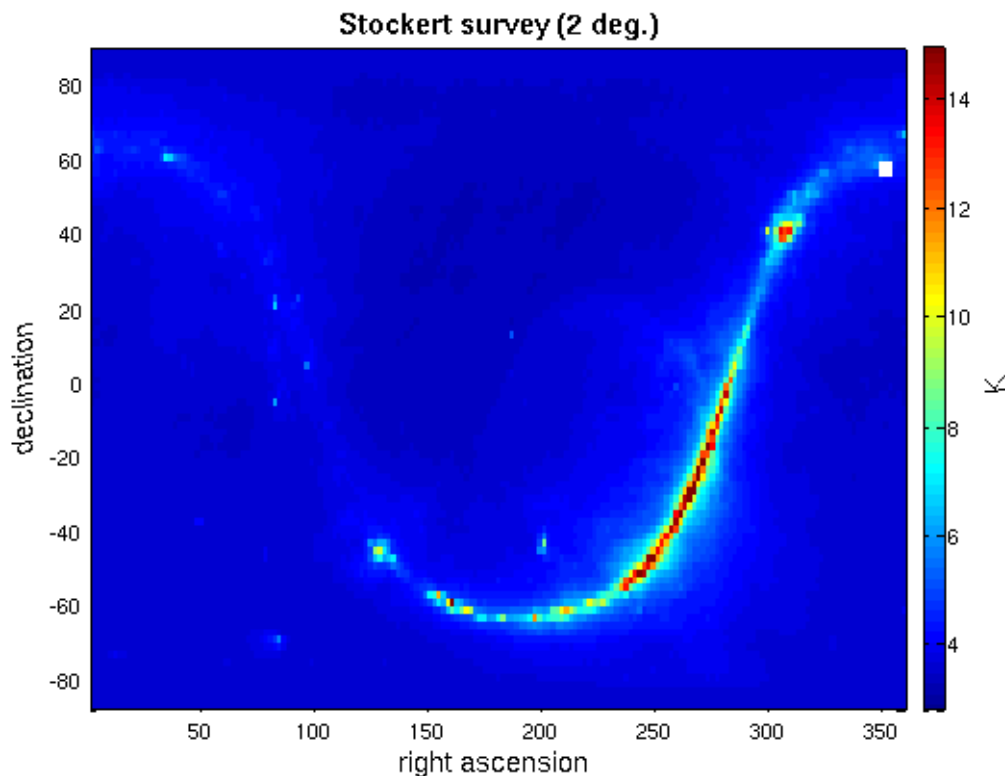

Figure 5: Stockert map (continuum radiation + cosmic background)

3.1.5.2.4 Practical considerations

3.1.5.2.4.1 Calibration and validation

As suggested before, it may be necessary to introduce a calibration factor proportional to TB_{sky} during the Cal/Val phase to correct for calibration and saturation problems of the existing surveys.

3.1.5.2.4.2 Quality control and diagnostics

Looking towards North (azimuth=0) with an incidence angle equal to the elevation of the observer, one looks towards the celestial North pole which location is invariant.

3.1.5.2.5 Assumption and limitations

The model above (flat land) described in 3.1.5.2.2.1 a) assumes a specular reflection over the surface. As galactic noise is inhomogeneous spatially, especially close to the galactic plane, this may be not completely justified. However, this limitation is expected to be of second order with respect to uncertainties on galactic noise maps.

Depending on the reliability we can put on galactic noise maps, it is not expected to have too much sky radiation affecting land pixels. However, a **flag** (actually a confidence descriptor N_sky, since this concerns individual views) will be built for cases where the relevant area of the galactic map exhibits a particularly strong source.

3.1.6 Spurious Events

Spurious events will also have to be covered. We have identified a list below which might not be exhaustive, but which might represent the most significant cases.

3.1.6.1 Radio Frequency Interferences (RFI)

Although the L-band used is protected and the satellite bandwidth is even more restricted to avoid out of band "spill over", spurious manmade signals might still be present. We are especially concerned by bodies and entities not obeying regulations (GPS-L3 transmission, Military radars, telecommunication and television relays not filtering properly harmonics etc…). No fool proof method for avoiding RFI exists. However:

- Very high levels of RFI induce large TB values that can be eliminated easily through TB **range testing**. The range (threshold) might be adjusted to local values expected for TB.
- Somewhat smaller effects can still hopefully be eliminated when checking **continuity of TB** (actually the 1rst Stokes parameter) over a range of incidence angles, as described in section 3.2.2.1.4.
- In addition, persisting **outliers** detected following the retrieval may indicate lower levels of RFI.

It must also be noted that RFI contamination outside the alias free zone could be folded back into the useful scene. Possible RFI signatures on Stokes parameters 3 and 4 are being investigated from airborne measurements in view of later improvements for the full polarization mode.

These criteria are used to clean the input TBs before their use in the retrieval optimization loop as detailed section 3.2.2.1.5. The count of disqualified TBs is kept updating (add) the current RFI counters in AUX_DGGRFI files daily. At a given date, a specific Current RFI map provides the total number of such eliminated TBs and the total number of observed TBs per DGG. The ratio of number of detections over the number of observations, R_{RFI}, provides a convenient estimate of the probability of RFI occurrence events since the beginning of the mission up to the date of the current RFI map. As integrated information (cumulative sums) the difference of counters between two current RFI maps at two different dates gives also access to probabilities of RFI occurrences at any dates and over any time window.

These time series of RFI counters in Current files provide a valuable integrated past story and evolution of RFI contamination detected by the algorithms directly from SMOS observations. Finally, through the accumulation of ground data and ground surveys, we have a **map of potential RFI sources** to be used as an input file for the processor. These maps allow obtaining statistics of RFI, i.e. the probability of RFI estimated through the fraction R_{RFI} of cases where RFI has been detected for each DGG during the last N days before the processed orbits. Bigger N, better the statistical stability of R_{RFI}, but also more dampened and delayed response to new/disappearing RFI events. An adequate range for N is 7 days to 15 days; we suggest using N=12 days which corresponds to a minimum of 4 orbits observations for a DGG at the equator ensuring enough observed TBs (more than 800) for a good stability of R_{RFI} computation and not too much dampening and time delays with the events.

We consider that when the DGGs RFI status are regularly detected as not clean at least during the previous N days prior to the retrieval day then even if it seems clean or cleaner at the retrieval day, we still have a low confidence on the DGGs status. For example, the RFI level could have dropped just below our detection level or dropped due to particular SMOS observation geometry but still having contamination in brightness temperatures.

In cases where the current DGGs are identified as a **potential** RFI contaminated DGG by the DGGRFI maps although no anomaly is detected in L2, the currently processed DGG is **not** eliminated. Rather, the retrieval post-standard deviation (DQXs) are enhanced by a factor C_{RFI} function of the RFI statistics R_{RFI}.

$$C_{RFI} = 1 + C1_{RFI} * R_{RFI}{}^{C2_{RFI}} / (1 + R_{RFI}{}^{\wedge} C2_{RFI}) \qquad \qquad \textbf{Eq 72}$$

It is the way we chose to forward to the user the knowledge of the RFI past story they do not have access to; we make the retrieval values artificially more uncertain. Note that added to the above general rule, R_{RFI} is also provided to the users and at some places in the algorithms the C_{RFI} factor is also used in the same restrictive spirit.

In every of three cases ((i) L2 RFI detection consistent with current RFI map, (ii) L2 RFI detection not consistent with current RFI map, (iii) no L2 RFI detection although RFI is indicated in current map), the RFI current map and/or statistics associated to it should be **updated**. (see § 3.2.5.2.1)

3.1.6.2 Sun glint

In some specific cases, sun glint will affect the signal (very similarly to what happens over the ocean, though to a lesser extent). The contribution will amount to several degrees and thus could be taken into account in a corrective fashion. The input will be the sun geometry and surface characteristics (spatial spreading and intensity) and could probably be accounted for in the inversion process.

The concept here is to use the L1c sun glint flag indicating the point where sun glint is specular. This point could be excluded from the retrieval or better only eliminated when posing a problem with the retrieval. The area around this point affected by sun glint is very variable (linked to the vegetation type structure and status). Here we only propose to eliminate the sun point and "reduce" the sun tails, as simulations have shown that sun tails are still visible after the sun cancellation process

3.1.6.3 Sun in secondary lobes

In some very specific cases, the Sun might affect the signal through the side-lobes contribution. Again, it is a purely geometrical feature and will strongly depend upon the quality of the antenna patterns.

3.1.6.4 Radio sources

For the sake of completeness we could also mention and flag **strong galactic radio sources**, as described in 3.1.5.2

3.1.7 Target independent issues

3.1.7.1 Polarization modes

In dual polarization mode, the forward models allow to compute both horizontal and vertical up-welling temperatures TB_H and TB_V. The vertical TB corresponds to a radiated electric field parallel to the plane defined by the nadir and the line of sight (see Figure 2); the horizontal one is perpendicular to this plane.

In full polarization mode, the forward model should also provide the 3rd and 4th up-welling Stokes parameters. Over land surface, they are assumed to be zero.

In other words, the full polarization option should not induce any additional burden concerning the forward model. The impact on processing is detailed in section 3.2.4.5.

3.1.7.2 Uncertainties in forward models

For the nominal case, many campaigns have shown the forward models to be mostly reliable and robust. However, their accuracy when applying them to SMOS data is limited for at least two main reasons:
- These models are based on **approximations**. For example, the small-scale roughness is described phenomenologically. For vegetation, the parameterization through albedo and optical thickness is a strong simplification.

- Moreover, the direct models are built for **uniform scenes**. However, SMOS scenes, on a 40 km scale, will never be uniform, even when limiting oneself to homogeneous cases. While expected effects are limited (although not negligible) for soil contributions, consequences may be quite significant for vegetation cover, due to the highly non-linear effect of vegetation parameters.

Some simulations have been made concerning the impact of non-uniformity and are included in the error budget below. It must also be stated that for a number of topics, we are quite aware that the assumptions made in modelling will have to be validated with real data. For instance, the roughness contribution is still subject to debate and will have to be eventually clarified.

3.2 Description of retrieval algorithm

3.2.1 L1c input

3.2.1.1 Geographical coverage and apodization

The land – or soil moisture- processor is intended to run over land surfaces with specifically prepared L1c data. It could be possible to limit the considered areas to the coast lines, but this would not allow good retrievals for all mixed pixels. It is thus better to include a margin over the oceans so that the retrieval can be performed efficiently even over coastal areas. For the sake of simplicity and to enable future improvements after the commissioning phase, it is recommended that the area covered by the L1C land product satisfies the following criteria:

- Over the oceans and salted areas, the L1c should extend over 200 km beyond actual land areas;
- The inland waters should all be present (some large lakes may freeze at high latitudes, some other have variable limits etc…);
- The land mask should be extended to be able to process sea ice at high latitudes, up to the furthest extent of the ice limits. It is recommended that only two extreme latitudes (one for each hemisphere) be considered.

It is finally recommended that the apodization function used by the L1C land processor is the exact Blackman function. The rationale is that amongst the families of functions we need to select one with a good trade off between spatial resolution and side lobes. Over land side-lobes are not as important as resolution (it is the opposite over the ocean in a way). A study was performed by Anterrieu and Waldteufel to address the choice of apodization window and it was found that Blackman was perfectly satisfactory, and it was not required to use more sophisticated functions.

It was also proposed at several stages to consider strip adaptive processing. We do not believe, and this was endorsed by the Science Advisory Group (SAG), that such an approach was useful as it provides the user with the worst spatial resolution, maximises the number of mixed pixels (either with water snow ice, urban RFI etc.).

3.2.1.2 Polarization mode

The SMOS instrument can be set in two different acquiring modes: a dual polarization mode or a full polarization mode. The relative advantages of one mode versus the other are not easy to assess as long as real SMOS data are not available. Consequently, retrieval algorithms have to be compatible with both modes. [AD13] covers specifically the full polarization mode.

The brightness temperatures forward models presented in section 3.1 were meant for dual polarization (H/V) mode. However, as the **3rd and 4th Stokes parameters (ST3 and ST4) above natural land surface are assumed to be in most cases very close to zero**. It may be safely assumed that these models are also adequate for full polarization mode. In another hand, we also have to provide a way to control the use of SMOS full polarimetric observations of these quantities in the retrieval if the observations tell us that the above highlighted assumption is wrong.

Nevertheless, each mode has its own specificities in terms of data organization and data interpretation, and these aspects have to be documented.

The main difference between the two modes, is that, in dual polarization mode, observations are independently acquired (i.e. they come from different snapshots) which is not the case in full polarization mode. A typical dual polarization sequence of TBs and their associated snapshot IDs (SID) is: $[TB_X1, SID1]$, $[TB_Y2, SID2]$, $[TB_X3, SID3]$, $[TB_Y4, SID4]$, ….

In full polarization mode, single polarized TBs are acquired similarly to those in dual polarization mode while cross polarized TBs are acquired using a sequence that sweeps all the configurations where one arm acquires a polarization while the two others acquire the other. As a result, two TBs are reconstructed, the cross-polarized TB$_{XY}$ and an associated (in turn) TBxx or TB$_{YY}$; both of them belong to the same snapshot and therefore share the same geometry. A typical sequence of TBs and their associated snapshot IDs is [XX1 SID1], [(XX2 XY2) SID2], [YY3 SID3], [(YY4 XY4) SID4] … It may also be noted that in full polarization mode, the polarizations are obtained in a somewhat more synchronous fashion. In full pol each polarization is acquired in a pseudo interleaved mode over the whole integration time, while in dual pol the two polarizations are acquired sequentially. The smearing effects might also be reduced in full pol mode.

Some consequences for the full polarization mode:

- Algorithms are a bit more complicated. It is necessary to keep track of the pairing information since some measurements (those associated to SID2 and SID 4 in the example above) have to be interpreted as inseparable pairs. If something happens to one of the components, it must be forwarded to its companion.

- For a same interval of time more observations are acquired (roughly 1/3 more); increasing the processing load. However, because a pair shares the same geometry, subparts of the processing that involves computations based on geometry (modelling, surface intercepted fractions under the antenna footprint) should not be an extra burden and a single call for the whole pair is sufficient. Implementation should take care of this.

Another expected difference is that the radiometric quality is mixed; it is the same as dual polarization for the single polarized acquiring, but it is degraded for the cross + single polarized acquiring pair. This may have an impact on the tuned values of some parameters that control the algorithm (thresholds and coefficients). The main consequence will probably mean that two different optimal setup configurations for TGRD UPF have to be found and used separately for each mode.

In all the following sections, we will consider dual polarization as the reference and when necessary we will specifically address full polarization.

NOTE: after the six months of the commissioning phase the full polarimetric mode has been selected as the regular mode. The first analysis showed that the TBXY observations exhibit unexpected features. While the imaginary part of TBXY, linked to A4, is not used in our modelling, the real part of TBXY was expected to provide extra measurements to TBXX and TBYY to be used in the surface parameters retrieval. The observed features on Re(TBXY) are not very consistent with ST3=0 or may come from reconstruction problems for cross polarimetric observations or from the instrument or both ...

Nevertheless, it will be safer for the time being to not use A3 in the **modelling** and consider only the subset of mono-polarized TBs, TBXX and TBYY.

<u>Note</u> that, filtering rules used for TBs screening still apply to all the TBs, only retrieval is concerned.

So, the retrieval part of the algorithms can be configured in the three following alternate modes, activated or deactivated on demand through a switch. Indeed, future improvements and better understanding may rehabilitate the use of Re(TBXY).
- Normal mode: all measurements can be used: TBXX, TBYY and Re(TBXY)
- Dual-in-Full mode: only TBXX and TBYY measurements can be used, including those coming from mixed snapshots.
- Pure-Dual-in-Full mode: only pure TBXX and TBYY measurements can be used. I.e. not belonging to mixed snapshots.
-

3.2.2 Input and pre-processing

3.2.2.1 L1c pre-processing

3.2.2.1.1 Level 1c output

As of December 2012, the applicable document for level 1c output is [AD 4 - SO-TN-IDR-GS-005 L1 Specs V5.4]. The following table lists some quantities of interest that are used by L2 processors.

Table 11: Inputs from L1c

Field	Tag name	Description	Units	Bytes	Size
L1c #	*From L1 DF T 25*	**SNAPSHOT info list**			
		Time Data			
#01	Snapshot_Time	UTC Time at which the scene was taken. Start of integration time period. This reference time is applicable to all PVT and AOCS data contained in the "Auxiliary data" part. Expressed in EE CFI transport time format. .	N/A	12	signed/unsigned integer (4bytes) Vector array of 3 elements First element (days) is signed integer, remaining two (seconds and microseconds) are unsigned
#02	Snapshot_ID	Unique identifier for the snapshot. Formed by aggregation of orbit and time within orbit. Contents of this field are formed by: Absolute_orbit_number*10000 + Seconds_from_ANX.	N/A	4	4 bytes unsigned integer
#03	Snapshot_OBET	Unique identifier for the snapshot. Formed by the OBET at T_SYNC extracted from L0. Represents start of integration time in OBET format.	N/A	8	8 bytes (OBET Format is specified in Section 4.2.1.2)
		Auxiliary Data (referenced to SNAPSHOT_TIME UTC value)			
#04 to #06	X, Y, Z_Positions	Orbit State Vector X Position in Earth Fixed Reference at snapshot_Time (field 02)	m	12	3x4 bytes signed IEEE float
#07 to #09	X, Y, Z_Velocities	Orbit State Vector X Velocities in Earth Fixed Reference at snapshot_Time (field 02)	m/s	12	3x4 bytes signed IEEE float
#10	Vector_Source	Source of the Orbit State Vector record: • SOLUTION_NULL = 0 • SOLUTION_INITIALISATION = 1 • SOLUTION_PVT_FROZEN = 2 • SOLUTION_NKF_SOLUTION= 3 • SOLUTION_SPS_SOLUTION= 4 • SOLUTION_PREDICTED = 5 (not used) • SOLUTION_RESTITUTED = 6(not used)	N/A	1	1 byte
#11	Q0	Real number component of quaternion obtained rotating from the J2000 inertial reference frame to the satellite body frame (OS L2P only)	N/A	4	4 bytes signed IEEE float
#12 to #14	Q1, Q2, Q3	First, second, third components of quaternion obtained rotating from the J2000 inertial reference frame to the satellite body frame. (OS L2P only)	N/A	12	3x4 bytes signed IEEE float
#15	TEC[10]	Total Electron Count content applicable to snapshot data. TEC value is corrected for SMOS altitude as in [RD.26]. (1 TECU = 10^{16} el/m^2)	TECU	8	8 bytes double
# 16 to #18	Geomag_F, Geomag_D, Geomag_I	Full or Total Intensity (F), Declination (D), Inclination (I), of Geomagnetic field vector applicable to snapshot data, obtained mixing PVT and IGRF model.	nT deg	24	3x8 bytes double

---

[10] We might have several TEC values per node (I per view). It is thus suggested to consider only the median value and compare it to the threshold so as to eventually set the TEC flag in case of high TEC

Field	Tag name	Description	Units	Bytes	Size
		(D) is the angle between magnetic north and true north. D is considered positive when the angle measured is east of true north and negative when west. Positive in eastward direction. (I) is the angle between the horizontal plane and the total field vector, measured positive into Earth. Positive in downward (towards Earth Surface) direction. (OS L2P only)			
# 19	Sun_RA	Right Ascension of Sun illumination direction in Earth Fixed Reference	deg	4	4 bytes signed IEEE float
#20	Sun_DEC	Declination of Sun illumination direction in Earth Fixed Reference	deg	4	4 bytes signed IEEE float
#21	Sun_BT	Direct Sun estimated Brightness Temperature that has been removed from snapshot	K	4	4 bytes signed IEEE float
#22	Accuracy	Snapshot overall accuracy measurement, based on Corbella equation and computed as the difference of the mean snapshot Brightness Temperature and the averaged physical temperature of the LICEF receivers	K	4	4 bytes signed IEEE float
#23	Radiometric_Accuracy	Error accuracy measurement in the Brightness Temperature value at boresight: • First one is the pure polarization • Second one is the crosspolarization The second element is only used to store the boresight accuracy for full pol snapshots, set to 0 in all other cases.	K	4	4 bytes signed IEEE float
L1c B	*From L1 DF T 26*	*Dual polarization products*			
		Grid Point Data			
#01	Grid_Point_ID	Unique identifier for Earth fixed grid point, linking it to Auxiliary Earth Grid file. For ISEA 4-9, maximum of 2.7M pixels	N/A	4	4 bytes (for ISEA 4-9, maximum of 2.7M pixels)
#01	Grid_Point_Latitude	Latitude of the DGG cell's centre identified by Grid_Point_ID	N/A	4	4 bytes signed IEEE float
#01	Grid_Point_Longitude	Longitude of the DGG cell's centre identified by Grid_Point_ID	N/A	4	4 bytes signed IEEE float
#01	Grid_Point_Altitude	Altitude of the DGG cell's centre identified by Grid_Point_ID	N/A	4	4 bytes signed IEEE float
#01	Grid_Point_Mask	Flag indicating land/sea USGS content, coastline distance and ice content (OS only)	N/A	1	unsigned byte
#02	Counter	Counter of Brightness Temperature Data values for current point (variable number depending on point across track position). Size of array to be read with data.	N/A	1	1 byte (maximum value of 255)
		BT Data (repeated COUNTER times)			
#03	Flags	L1c flags applicable to the pixel for this particular integration time. Flags identified below this table. (AF_FOV flag (only OS), EAF_FOV flag, BORDER_FOV, polarization, RFI flag (only OS), SUN_FOV, SUN_GLINT_AREA,	N/A	2	2 bytes

71

Field	Tag name	Description	Units	Bytes	Size
		SUN_GLINT_FOV, SUN_POINT, SUN_TAILS) SUN_TAILS)			
#04	BT_Value	Brightness temperature value over current Earth fixed grid point, obtained by DFT interpolation from L1b data.	K	4	4 bytes signed IEEE float
#05	Radiometric_accuracy	Error accuracy measurement in the Brightness Temperature presented in the previous field, extracted in the direction of the pixel. Coded in 2's complement. LSB=$X/2^{16}$. Meaning that value=(unsigned short)$*X/2^{16}$ K, where X is Radiometric_Accuracy_Scale given in SPH	K	2	2 bytes unsigned short
#06	Incidence_Angle	Incidence angle value corresponding to the measured BT value over current Earth fixed grid point. Measured as angle from pixel to S/C with respect to the pixel local normal (0° if vertical). Coded as an unsigned short. LSB=$90/2^{16}$. Meaning that value=(value coded as unsigned short)$*90/2^{16}$degrees	deg	2	2 bytes unsigned short
#07	Azimuth_Angle	Azimuth angle value corresponding to the measured BT value over current Earth fixed grid point. Measured as angle in pixel local tangent plane from projected pixel to S/C direction with respect to the local North (0° if local North) Coded as an unsigned short. LSB=$360/2^{16}$. Meaning that value=(value coded as unsigned short)$*360/2^{16}$ degrees	deg	2	2 bytes signed short
#08	Faraday_Rotation_Angle	Faraday rotation angle value corresponding to the measured BT value over current Earth fixed grid point. It is computed as the rotation from antenna to surface (i.e. inverse angle) Coded as an unsigned short. LSB=$360/2^{16}$. Meaning that value=(value coded as unsigned short)$*360/2^{16}$ degrees	deg	2	2 bytes signed short
#09	Geometric_Rotation_Angle	Geometric rotation angle value corresponding to the measured BT value over current Earth fixed grid point. It is computed as the rotation from surface to antenna (i.e. direct angle) Coded as an unsigned short. LSB=$360/2^{16}$. Meaning that value=(value coded as unsigned short)$*360/2^{16}$ degrees	deg	2	2 bytes signed short
#10	Snapshot_ID	Unique identifier for the snapshot. Formed by aggregation of orbit and time within orbit. Contents of this field are formed by: Absolute_orbit_number*10000+Seconds_from_ANX	N/A	4	4 bytes unsigned integer
# 11	Footprint_Axis 1	Elliptical footprint major semi-axis value. Coded as an unsigned short. LSB=(SPH Table 24 field #12)/2^{16}. Meaning that value=(value coded as unsigned short)* (SPH Table 24 field #12)/2^{16} km	km	2	2 byte unsigned short
#12	Footprint_Axis2	Elliptical footprint minor semi-axis value	km	2	2 byte unsigned short
L1c C	*From L1 DF T 27*	*Full polarization products*			
		Grid Point Data			
#01	Grid_Point_ID	*Same as dual polarization*	N/A	4	4 bytes (for ISEA 4-9, Maximum of

Field	Tag name	Description	Units	Bytes	Size
					2.7M pixels)
#01	Grid_Point_Latitude	*Same as dual polarization*	N/A	4	4 bytes signed IEEE float
#01	Grid_Point_Longitude	*Same as dual polarization*	N/A	4	4 bytes signed IEEE float
#01	Grid_Point_Altitude	*Same as dual polarization*	N/A	4	4 bytes signed IEEE float
#01	Grid_Point_Mask	*Same as dual polarization*	N/A	1	unsigned byte
#02	Counter	*Same as dual polarization*	N/A	1	1 byte (maximum value of 255)
		BT Data (repeated COUNTER times)			
#03	Flags	Same as dual polarization	N/A	2	2 bytes
#04	BT_Value_Real	Brightness temperature value over current Earth fixed grid point, obtained by interpolation from L1b data. Contains real components of HH, HV or VV polarization measurements (see Note2).	mK	4	4 bytes signed IEEE float
#05	BT_Value_Imag	Brightness temperature value over current Earth fixed grid point, obtained by interpolation from L1b data. Contains imaginary components of HH, HV or VV polarization measurements (see Note2).	mK	4	4 bytes signed IEEE float
#06	Radiometric_accuracy	*Same as dual polarization*	K	2	2 bytes unsigned short
#07	Incidence_Angle	*Same as dual polarization*	deg	2	2 bytes unsigned short
#08	Azimuth_Angle	*Same as dual polarization*	deg	2	2 bytes signed short
#09	Faraday_Rotation_Angle	*Same as dual polarization*	deg	2	2 bytes signed short
# 10	Geometric_Rotation_Angle	*Same as dual polarization*	deg	2	2 bytes signed short
#11	Snapshot_ID	*Same as dual polarization*	N/A	4	4 bytes unsigned integer
# 12	Footprint_Axis 1	*Same as dual polarization*	km	2	2 byte unsigned short
#13	Footprint_Axis2	*Same as dual polarization*	km	2	2 byte unsigned short

Useful information is included in the L1c **product headers**.

- Main Product Header MPH (AD 4 § 4.1.1): product confidence field #22 will be accounted for in the global quality index. Using the product confidence field #23 is strongly suggested.
- Specific Product Header SPH (Table 9): note ASCENDING_FLAG (field #09); MODE (field #11).

We do not foresee significant differences in the ascending and descending orbit processing. However, we might encounter different TEC levels, different RFI levels, or different natural surface effects (dew, rain) all linked to diurnal cycles and requiring specific attention. This will only be answered after the commissioning phase.

After 3 years of operations part of this answer comes for the handling of DGG_CURRENT files content which is now separated per pass basis, both in their uses and their updates. The processing of ascending orbits uses/updates ascending specific data stored in DGG_CURRENT files and reciprocally for descending ones.

Note1: several MODE values indicate that the L1c record is not to be considered for L2 SM processing. In addition, the field #17 in SPH, num_MDP, is expected to supply the number of measurement records. For land surfaces, the case where num_MDP is zero might happen and should not generate an exception.

Note2: in the table above (directly extracted from AD 4) notations for polarizations make the use of H, V or HH, VV, HV though they refer to measurements in satellite X/Y reference frame. In the context of these L1 AD, note that H means X-axis and V means Y-axis.

3.2.2.1.2 Views – Pixels – Nodes

In this entire document and in particular in the following sections, concepts of nodes, views and pixels are extensively used. It seems thus useful to explicit these terms so as to avoid any possible confusion.

A **view** is defined as a collection of information attached with one **single** SMOS TB measurement. It consists of the following information:
- A brightness temperature, which can be a real number in dual polarization mode or a complex number in full polarization mode.
- An observation geometry that consists in the incidence angle and the azimuth angle relative to the observed surface location with respect to the instrument.
- A transformation geometry that consists of two angles. One related to the Faraday rotation, and the other related to the transfer from the surface reference frame (H/V polarization) to antenna reference frame (X/Y polarization)
- A radiometric property that provides the radiometric accuracy of the brightness temperature, the semi-major and semi-minor axis of the 3dB ellipse contour of the synthetic antenna footprint related to the observed point.
- An identification which indicates the polarization of the brightness temperature, (either X or Y in dual polarization mode, or XX, YY, XY or YX in full polarization mode) as well as the identification number of the associated snapshot. Note that we make no difference between XY and YX cross polarizations.

In the context of the L1C data file, views are directly mapped to the BT Data records described in Table 11 above, in either fields #03 to #12 for the dual polarization mode, or in fields #03 to #13 for the full polarization mode.

This concept of view can also be used at surface level (for observation or for modelling); in which case X should be replaced by H and Y by V.

Usually, when referring to views, not all attributes are necessarily considered. Quite often, only a subset is necessary, depending on the context.

A **SMOS pixel consists** in the brightness temperature and its spatial resolution on the Earth. The pixel can be considered then as the subset of a view made up of the TB itself and the 3dB surface characterized by the ellipse semi axes.

A **node**, or a DGG node, L1C node and sometimes called a L1C pixel, is defined as the list of observed views attached to a given location on the Earth defined by its DGG id number in the ISEA grid.

In the following, the content of fields #02 ("counter") in L1c files is renamed **M_AVA0**: this is indeed the number of views initially available for any given node.

3.2.2.1.3 Angles

Figure 6 describes the viewing geometry, as defined in [114].

Note that in *standard SMOS coordinates*, SZ will be replaced by "*minus SX*" and SX replaced by "*SZ*".

Also, note that in this figure the angle between the instantaneous orbital plane and the reference geographical meridian (or parallel) plane is missing.

For subsequent processing, it is necessary to obtain angles depicted on this figure. The logical way to do this is as follows:
- Compute geographical polar angles θ_g and ϕ_g from spacecraft position and DGG node coordinates
- Compute, from these and AOCS, polar angles in the antenna reference frame θ_a and ϕ_a, as well as angle ψ;
- From θ_a and ϕ_a compute **director cosines (DC)** ξ & η: $\xi = \sin\theta_a \, \cos\phi_a$; $\eta = \sin\theta_a \, \sin\phi_a$
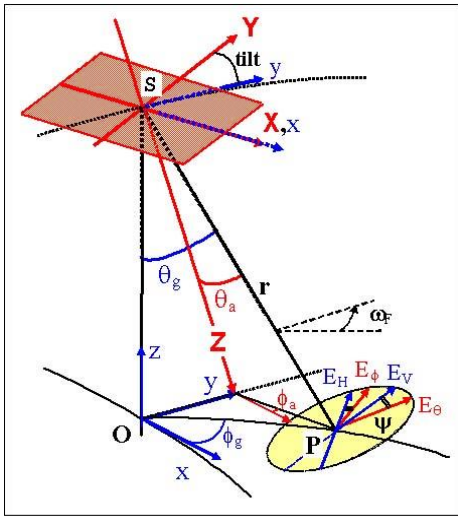
The expressions for angle ψ as well as the Faraday rotation angle ω_{Fa} are given below for information; actually, the L1c records will supply values for ω_{Fa} and $\alpha_r = -(\phi_a + \psi)$, which are needed for geometrical transformations (fields #08 & #09 or #09 & #10 in Table 11 B & C), as well as for incidence angles θ (fields #06 & #07).

On the other hand, it remains necessary to compute in L2 the **DGG node director cosines**, which are necessary for applying the weighting function. It is expected that the EE CFI will be used in this respect.

Figure 6: Viewing geometry for a particular L1c view11

3.2.2.1.4   Applying spatial resolution requirements

Among the **M_AVA0** views read from L1c data for a given DGG pixel, some have to be discarded, reducing the number of available views to **M_AVA**; some others will be kept with a reduced weight.

While L1c TB are restrained to the extended alias free zone of the field of view, it is still necessary to filter out areas that do not meet the spatial resolution criteria. This is illustrated by Figure 7 which shows contours for these size and shape (dotted line) requirements:

- the area of the half maximum contour of the WEF must be smaller than 55×55 km^2 (TH_SIZE);
- the elongation (major axis to small axis ratio) of this contour must be smaller than 1.5 (TH_ELON).
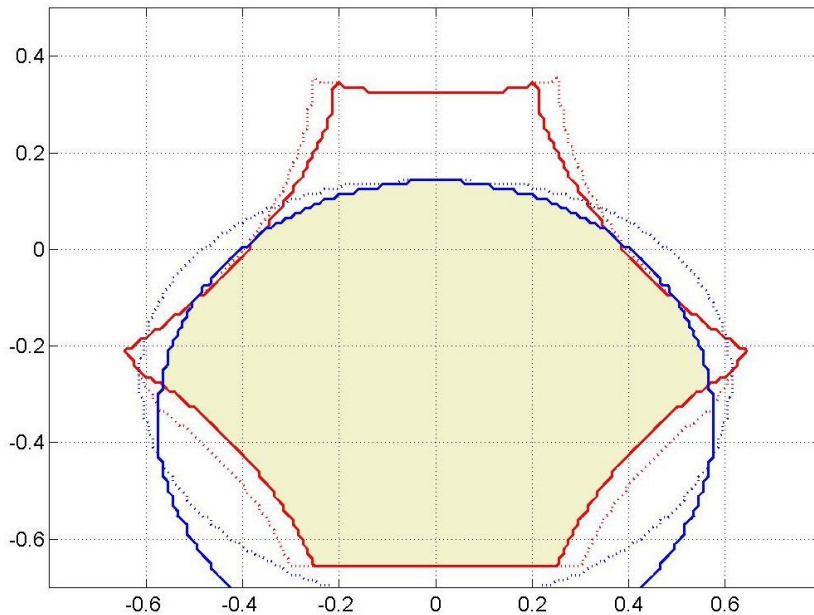
This spatial resolution mask was built using a 771 km flight altitude, which is the average value over the orbit (altitude range spans from 751 to 788 km).

Filtering out the non-compliant views will actually make use of FOOTPRINT information provided by L1c: **half lengths** of major axes (axis$_1$) and minor axes (axis$_2$) of the 3dB contour of the pixel, which is close to an ellipse.

Hence L1c views will be **eliminated** whenever:

- $\sqrt{4 \cdot \text{axis}_1 \cdot \text{axis}_2} > TH\_SIZE$

or

- $\dfrac{\text{axis}_1}{\text{axis}_2} > TH\_ELON$

Figure 7: Impact of spatial resolution requirements & alias-free margin over the valid FOV (in blue pixel size, dotted ellipse elongation constraint):

---

[11] Note that this angle definition has two caveats: i) the reference frame XYZ in red is not direct; ii) ESA will issue a set of official angles and the diagramme and conventions used in this document will have to be adapted.

On Figure 7 is shown also the impact of a margin over the alias free zone. It is necessary to account for such a margin; to this end, use will be made of the L1c BORDER_FOV flag.

In full polarization mode, a mono-polarized view and its companion cross-polarized view share the same geometry: if one is discarded the other too.

3.2.2.1.5   Filtering L1c views

In this subsection and the following one, the "TB" notation refers to brightness temperatures at **antenna level**, i.e. the BT in Table 11 (L1c inputs).

The TBs are screened in the order in which they are stored (referring to the Counter field #2 in Table 11 B & C). It is assumed that the first one then corresponds to the highest incidence angle. In full polarization, whenever one of the TB components is eliminated its companion TB is also eliminated (irrespective of its status).

Some tests are applied on what is called "paired views"; they consist in pairing views having two opposite polarizations (TB_X, TB_Y) or (TB_Y, TB_X) and made from adjacent snapshots (i.e. no gap between X and Y or between Y and X). Full polarization is handled similarly with the only difference that three views are considered [(TB_{YY}, TB_{XY}), TB_{XX}] or [(TB_{XX}, TB_{XY}) TB_{YY}].

TB range testing for strong RFI sources detection: two elimination criteria for screening L1c views are introduced:
- The first test compares each antenna TB (fields #04 or #04 and #05 in Table 11 B & C) to its expected range: defined by [TBscene_min(Tsurf_min), TBscene_max(Tsurf_max)] for both polarizations (X and Y). In full polarization mode the ranges are defined by [TBXY_Re_min, TBXY_Re_max] and by [TBXY_Im_min, TBXY_Im_max].
- The second test compares the amplitude of paired views, $\sqrt{TB_X^2 + TB_Y^2}$, to the expected range defined by $\sqrt{2}$ [TBscene_min(Tsurf_min), TBscene_max(Tsurf_max)].

The TBscene_min(Tsurf_min) and TBscene_max(Tsurf_max) are function of Tsurf_min and Tsurf_max, the minimum and maximum of the temperatures reported within the working area among STL1 and SKT:
- TBscene_min(Tmin) = Emissivity_Min * (Tsurf_min – (Tscene_Margin_Low + DTB_Scale*DTB))
- TBscene_max(Tmax) = Emissivity_Max * (Tsurf_max + (Tscene_Margin_High + DTB_Scale*DTB))

Note: for the amplitude test which uses a pair of views, the maximum DTB of the two is taken.

Emissivity_Max will be set to 1, the maximum physical value with Tscene_Margin_High representing the uncertainty on ECMWF temperatures field set to 5 K. Emissivity_Min will be set to 0.3, which corresponds to the water emissivity at horizontal polarization and incidence angle equal to 55°. The Tscene_Margin_Low representing the uncertainty on ECMWF temperatures field is set to 5 K. DTB_Scale accounts for the radiometric uncertainty interval extension and is set to 2.0 i.e. one standard deviation is accepted.

Finally, the static ranges for the TBXX real parts and imaginary part are set to -50 K for their minimum values and +50 K for their maximum values.

Note that all these range parameters, Emissivity_Min, Tscene_Margin_Low, Emissivity_Max, Tscene_Margin_High, DTB_Scale, TBXX_RE_MIN, TBXX_RE_MAX, TBXY_Im_min abd TBXY_Im_max are kept configurable and provided in the TGRD UPF.

L1C flags: A further rejection criterion is based on L1c flags; the SUN_POINT flag **eliminates** the view.

When TB values are outside the above range tests they are understood as contaminated by strong RFI.

Since softer RFI may have passed the above range test, a short **additional module** aimed at refining L2 RFI detection is proposed for the L2 processor. This procedure takes advantage of incidence angle continuity whenever possible, and of the fact that, while TB_X and TB_Y vary strongly with incidence angle, this is not the case for the first Stokes parameter under the assumption that the surface is sufficiently homogenous and/or uniform and its emissivity follows a Fresnel angular law. I.e. the only TBs variability comes from the Fresnel law and not from antenna footprint sampling.

 It consists in:
- considering only L1C **paired views** when their number is larger than a given threshold **TH_AVA_Min** (this condition ensures at the same time a significant range of incidence angles)**;**
- considering only paired views (TB_X, TB_Y) or (TB_Y, TB_X);
- computing the mean value ⟨TBS1⟩ of the **halved first Stokes parameter TBS1** = 0.5*(TB_X + TB_Y) of every paired view;

- **eliminating** paired views for which abs(TBS1 - ⟨TBS1⟩) > CA_TBS1 + CB_TBS1 x DTB$_X$, where CA_TBS1 and CB_TBS1 are constant (provided in TGRD UPF), and DTB$_X$ the radiometric uncertainty on TB$_X$;
- This test shall be applied only when the surfaces are reasonably homogeneous. That guaranties that the meaning of ⟨TBS1⟩ comes only from the small angular variation of the first Stokes and not from a drastic change in brightness temperatures seen under various AFP surface fractions interception. The highest TB contrast comes when water TBs is mixed with TBs of other surfaces and to a lesser extent when snow melting is involved.
- To avoid false RFI detections (mainly on coast transition) and thus to prevent removing usable TBs, the 1st Stokes detection test is activated only when the following test is true; a suggested value for the threshold is 0.1275 (85% / 15%) to begin with:
- (FWO + FWL + FSW)*(1- FWO - FWL - FSW) < TH_HOMOGENOUS_1ST_STOKES

Finally, a further second step RFI detection may be done after the retrieval. This RFI detection discards and counts the possible presence of outliers in the residuals coming from the minimization process done by the retrieval (see section 3.2.4.2)

At the end of this filtering process, M_AVA is decreased by the number of eliminated views. The RFI occurrences are also reported in two counters N$_{RFI-X}$ and N$_{RFI-Y}$. These counters are used to update the DGG_CURRENT_RFI LUT (see section 3.2.5.2.1). In full polarization, RFIs for TB$_{XY}$ are reported through the counter of its mono-polarized companion TB.

The following table summarizes the updates to consider for M_AVA, N$_{RFI-X}$ and N$_{RFI-Y}$ for all the form of filtering used in algorithms: spatial resolution requirements (previous section), TBs range and RFI filtering (this section) and outliers filtering (section 3.2.5.1).

Table 12: Summary on L1C views filtering and RFI reporting

#	Test	TBs used	TBs eliminated	M_AVA increment	N$_{RFI-X}$ increment		N$_{RFI-Y}$ increment
1	Amplitude Dual Pol Paired views	TBX **and** TBY	TBX **and** TBY	-2	+1 if min or max test fails	**a n d**	+1 if min or max test fails
	Amplitude Full Pol Tripled views	TBXX **and** TBYY	TBXX **and** TBYY **and** TBXY	-3	If min or max test fails, then +2 if TBXY shares the same snapshot with TBXX **or** +1 otherwise	**a n d**	If min or max test fails, then +2 if TBXY shares the same snapshot with TBYY **or** +1 otherwise
2	TB range Dual Pol Single views	TBX **or** TBY	TBX **or** TBY	-1	+1 if min or max test fails	**or**	+1 if min or max test fails
	TB range Full Pol Single views	TBXX **or** TBYY **or** TBXY	TBXX **or** TBYY **or** TBXY	-2 if companion or -1 if not	If min or max test fails, then +2 if TBXY shares the same snapshot with TBXX **or** +1 otherwise	**or**	If min or max test fails, then +2 if TBXY shares the same snapshot with TBYY **or** +1 otherwise
3	SunPoint flag Dual Pol Single views	TBX **or** TBY	TBX **or** TBY	-1	0 but set FL_SUN_POINT_C to true		
	SunPoint flag Full Pol Single views	TBXX **or** TBYY **or** TBXY	TBXX **or** TBYY **or** TBXY	-2 if companion **or** -1 if not	0 but set FL_SUN_POINT_C to true		
4	Spatial resolution Dual Pol Single views	TBX **or** TBY	TBX **or** TBY	-1	0	**or**	0
	Spatial resolution Full Pol Single views	TBXX **or** TBYY **or** TBXY	TBXX **or** TBYY **or** TBXY	-1	0	**or**	0

#	Test	TBs used	TBs eliminated	M_AVA increment	N_{RFI-X} increment		N_{RFI-Y} increment
5 [12]	1st Stokes anomaly Dual Pol Paired views	TBX **and** TBY	TBX **and** TBY	-2	+1	**and**	+1
	1st Stokes anomaly Full Pol Tripled views	TBXX **and** TBYY	TBXX **and** TBYY **and** TBXY	-3	+2 if TBXY shares the same snapshot with TBXX **or** +1 otherwise	**and**	+2 if TBXY shares the same snapshot with TBYY **or** +1 otherwise
6	Outlier Dual Pol Single views	TBX **or** TBY	TBX **or** TBY	-1	+1	**or**	+1
	Outlier Full Pol Single views	TBXX **or** TBYY **or** TBXY	TBXX **or** TBYY **or** TBXY	-1	+1 also for a TBXY if it shares the same snapshot with a TBXX	**or**	+1 also for a TBXY if it shares the same snapshot with a TBXY

Note 1: Companion TB means the TB that shares the same snapshot ID. e.g. TBXX and TBXY are companions if they share the same snapshot ID.

Note 2: As a general rule, if we need to increment the RFI counter for TBXY, we increment N_RFI_X if TBXY shares a snapshot with TBXX, and we increment N_RFI_Y if TBXY shares a snapshot with TBYY. If TBXY is missing its companion TB (this should not happen but nevertheless), always increment N_RFI_X if we need to increment the RFI counter for this TBXY.

Note 3: In the Sun Point Flag Test, it is assumed that the companion TB will fail its own Sun Point Flag Test.

Note 4: Test No. 1 is the first test to be performed. Tests No 2, 3 (for full polarization only) and 4 are then performed on ALL TBs including those eliminated by previous tests. Test No.5 is then performed on the remaining TBs. Finally, Test No. 6 is performed on the TBs that pass all the previous tests.

3.2.2.1.6   Enhancing radiometric uncertainties

According to L1 documentation ([AD 4] § 3.4.3.3), the following flags are provided in the L1c products: AF_FOV, EAF_FOV, BORDER_FOV, SUN_FOV, SUN_POINT, SUN_TAILS, SUN_GLINT_FOV, SUN_GLINT_AREA flags.

BORDER_FOV and AF_FOV(=FALSE) are used to **enhance** corresponding radiometric uncertainties. Similarly, SUN_TAILS and SUN_GLINT_AREA flags are used to **enhance** corresponding radiometric uncertainties. Enhancing factors C_BORDER, C_EAF, C_SUN_TAILS, C_SUN_GLINT_AREA are provided in TGRD UPF.

The purpose of these enhancements is to have the possibility to decrease the weight of particular brightness temperatures when the observation is reported as being potentially less accurate or distorted by the L1 upstream processing.

3.2.2.1.7   Filtering L1c pixels

Once the faulty or dubious L1c views have been filtered out from the initial number M_AVA0, the number M_AVA of remaining views for a given SMOS grid point is estimated.

The **initial validation index MVAL0** is a weighted sum of the number of available measurements that expresses roughly their information content:

$$\text{MVAL0} = C_{val}\, \text{DTB\_F} * \text{sum}(1/\text{DTBa})\ /\ C_{RFI}$$ **Eq 73**

Where
- The sum is to be carried out over every view and polarizations **TB_X** and **TB_Y** (and **TB_XY** in full polarization mode)
- The DTBa are radiometric uncertainties over each TB at antenna level

---

[12] Only if half 1st Stokes anomaly is applicable.

- DTB_F is a **scaling** factor
- C_{val} (= C_{val_2} or C_{val_4}) is a coefficient depending on the polarization mode.
- C_{RFI} is the enhancement factor due to frequent RFI contamination (see section 3.1.6.1)

The scaling factor is adjusted in such a way that MVAL0 is roughly equal to the maximum available number of views along the track. Away from track, it decreases rapidly because the along track size of the FOV decreases (as well as the range of available incidence angles).

MVAL0 is next compared to the **eliminatory threshold TH_MMIN0**:

If MVAL0 < TH_MMIN0, the **L1c pixel is eliminated**.

Numerical values for DTB_F, C_{val}, **TH_MMIN0** are provided in TGRD UPF.

3.2.2.1.8 Computing modelled antenna brightness temperatures

The forward modelling presented in section 3.1 provides top of atmosphere simulated brightness temperatures for horizontal and vertical polarizations defined in the Earth reference frame. These modelled TOA TB_H and TB_V, need to be transformed in components relevant at the SMOS antenna frame in order to be compared with SMOS brightness temperatures observation TB_X and TB_Y. This transformation takes into account a rotation of the electric fields, due both to geometrical considerations and to the Faraday rotation induced by ionosphere.

Here the definition of angles is believed to follow the conventions described in Earth Explorer CFI Software Mission Convention Document [115]; we introduce the mathematical expressions for the angles to be used in the transport from ground to antenna reference frames [116] [117]:

$$\theta_a = \text{Arccos}\left[\sin\text{tilt}\ \sin\theta_g\ \sin\phi_g + \cos\text{tilt}\ \cos\theta_g\right]$$

$$\phi_a = -\text{Arcsin}\left[\frac{(-\sin\text{tilt}\ \cos\theta_g + \cos\text{tilt}\ \sin\theta_g\ \sin\phi_g)}{\sin\theta_a}\right]$$

$$\psi = \pi - \text{Arcsin}\left[\frac{(\cos\text{tilt}\ \sin\theta_g - \sin\text{tilt}\ \cos\theta_g\ \sin\phi_g)}{\sin\theta_a}\right]$$

for $\pi/2 \leq \phi_g \leq 3\pi/2$ (modulo 2 π):

ϕ_a to be replaced by $\pi - \phi_a$ **Eq 74**

ψ to be replaced by $\pi - \psi$

The rotation angle α_r supplied by L1c is the "direct" one, i.e. this rotation transfers the TOA signals to the antenna reference frame. The Faraday angle ω_{Fa} as supplied in L1c is a correction, i.e. it is the angle that corrects the Faraday rotation. The final rotation angle is then defined by: $\boldsymbol{\alpha = \alpha_r + \omega_{Fa}}$.

Theoretical description of Faraday rotation

The Faraday rotation is due to the effect of ionospheric electrons on the propagation of electromagnetic waves.

Mathematical description of algorithm

The Faraday angle ω_{Fa} for each view is provided by L1c data (field #09 or #10 in tables 26 or 27, depending on polarization mode), using auxiliary TEC_n (Total Electron nadir columnar Content) values. Therefore, the description below needs only be implemented when introduced in the direct model in the case where TEC_n is retrieved.

So, the **SM processor will use directly the ω_{Fa} provided in L1c data**, whereas SSS processor will use the following equations.

Making use of the magneto-ionic theory and using the quasi longitudinal approximation as well as assuming a plane parallel to ionosphere result in the following expression for L-band ([25, 118],):

$$\omega_{Fa} \approx 6950 * TEC_n * (\mathbf{B.U_{LS}}) / \cos(\theta_g)\ (°)$$ **Eq 75**

Where:
- TEC_n is the total **vertical** electron content (TEC units; 1 TEC unit = 10^{16} m^{-2}); it is obtained from L1c field #15 for each view. The range of TEC_n is about 5 to 50.
- If the TEC_n is retrieved, then a unique value for the DGG node is defined as initial value by selecting the median of the values for every view. Otherwise, every individual TEC_n value can be selected for computing ω_{Fa}.

- (**B.U$_{LS}$**) is the scalar product of the magnetic field vector **B** by the unitary vector **U$_{LS}$** giving the direction of the line of sight (from target to spacecraft).
- The magnitude |B| of **B** (tesla) is obtained from L1c field #16 (expressed in nanotesla). The range of |B| is about 2 to 5 10^{-5} Tesla.
- The vectors **B** and **U$_{LS}$** must be expressed in the same Euclidian reference frame.
- Concerning **B**: The L1c provides (fields #17 and #18) the declination dec_B and inclination inc_B of B in a local geographical frame Oxyz (Ox towards East, Oy toward North, Oz upwards)
- In L1c data, dec_B is understood as the angle of **B** away from geographic North Oy, counted positive eastwards (clockwise); inc_B is understood as the angle of **B** away from the local horizontal plane Oxy, counted positive downwards.
- Every individual |B|, dec_B and inc_B values can be selected for computing **ω$_{Fa}$**.
- Concerning **U$_{LS}$**: let us define polar geographical coordinates $θ_g$ (elevation away from the Oz axis) and $φ_n$ (azimuth from origin Ox, counterclockwise).

Then:
- **B** = [cos(inc_B) sin(dec_B), cos(inc_B) cos(dec_B), -sin(inc_B)]
- **U$_{LS}$** = [sin($θ_g$) cos($φ_n$), sin($θ_g$) sin($φ_n$), cos($θ_g$)];
  (Note that $φ_n$ differs from the relative azimuth defined in a frame linked to the spacecraft and introduced in the SM ATBD).
  It is expected that the EE CFI may provide directly $φ_n$ and $θ_g$. Alternatively, $θ_g$ could be inferred from the incidence angle (provided by the L1c) through adding the Earth centre angle; $φ_n$ could be computed from the DGG node coordinates, assuming the coordinates of the sub-satellite point are provided by the EE CFI.

The **ω$_{Fa}$** Faraday angle value is positive clockwise.

Assumptions and limitations

A single average magnetic field vector is used rather than altitude dependent values when carrying out an integration over the line of sight. The optimal value corresponds to altitudes, which may vary between 350 and 400 km, depending on the ionospheric altitude profile. Considering the variation of B with altitude, resulting errors are not significant.

The TEC value is assumed constant over the area (up to about a 500 km size at ionospheric altitudes) concerned by a SMOS dwell line. This assumption may not be fully satisfactory in regions of strong ionospheric gradients.

Then, following [25], two transformation matrices are defined depending on polarization modes in order to express the Stokes parameters A on the antenna as a function of the up-welling (TOA i.e. top of the atmosphere) TB quantities.

Dual polarization mode:

$$\begin{bmatrix} A1 \\ A2 \end{bmatrix} = \begin{bmatrix} \cos^2(a) & \sin^2(a) \\ \sin^2(a) & \cos^2(a) \end{bmatrix} \begin{bmatrix} TB_H \\ TB_V \end{bmatrix} = [MR2] \begin{bmatrix} TB_H \\ TB_V \end{bmatrix}$$

Eq 76a

Remark: For $a=\pm45°$, MR2 becomes singular (not invertible). Therefore, the reciprocal transformation, from antenna to Earth reference, can't be done around this rotation angle.

Full polarization mode:

$$\begin{bmatrix} A1 \\ A2 \\ A3 \\ A4 \end{bmatrix} = \begin{bmatrix} \cos^2(a) & \sin^2(a) & -\cos(a)\sin(a) & 0 \\ \sin^2(a) & \cos^2(a) & \cos(a)\sin(a) & 0 \\ \sin(2a) & -\sin(2a) & \cos(2a) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} TB_H \\ TB_V \\ TB_3 \\ TB_4 \end{bmatrix} = [MR4] \begin{bmatrix} TB_H \\ TB_V \\ TB_3 \\ TB_4 \end{bmatrix}$$

Eq 76b

Correspondence to L1c notations (Table 11):

Dual polarization: A1 = TB$_{XX}$, A2 = TB$_{YY}$

Full polarization: A1 = real(TB$_{XX}$), A2 = real(TB$_{YY}$), A3 = 2 real(TB$_{XY}$), A4 = - 2 imag(TB$_{XY}$)

The **Eq 76** a & b are used after the modelling for every observed views geometry has produced M_AVA surface modelled [TB$_H$ TB$_V$]^t vectors. These vectors are then converted to antenna-modelled vectors [TB$_{XX}$ TB$_{YY}$]^t using [MR2] matrices

computed for every view a angle. Since antenna TBs are modelled twice more than observed, a last step remains to be done select and keep only the relevant modelled antenna polarizations (i.e., those that match the observed ones).

The same logic shall be applied for full polarization but using MR4 for antenna TB modelling except we will consider only the use of A3. Indeed, A3 is the only which depends of surface parameters through TB_H-TB_V.

3.2.2.2 Summary of auxiliary data

Full details concerning auxiliary data are to be found in the TGRD. This ATBD section simply presents the summary TGRD tables.

These tables mainly belong to 2 main categories, depending whether their content does not or does vary with time. Subcategories are as follows:

- Tables always prefixed with DGG describe maps on the SMOS **Discrete Global Grid** (DGG) which is the ISEA4-9 grid used for L1c products.
- Tables always prefixed with DFFG describe maps on the **Discrete Flexible Fine Grid** which is the fine working grid used in this document. [AD 12]
- Other general-purpose tables, which are prefixed neither with DGG nor with DFFG.

Note 1: those prefixes mean only how the auxiliary data are used by the algorithms; they are not necessarily linked to the original format and scale of the auxiliary data source that may be different.

Note 2: only 30% of maps tables are of interest for SM (land) => the true size can be largely reduced by using an optimized computer representation (sparse matrix).

Finally, the **user's parameter files** (TGRD UPF) include every numerical value for constants used in the processor, allowing thus easy tuning.

3.2.2.2.1 Static auxiliary data

Table 13: Overview of TGRD precomputed tables

Sec-tion	Table Name	Description	Source	Sampling Requirements	Time Sampling	Generation Method	Quality Check	Notes
center	**Instrument Model Tables**							
3.1.1.1	MEAN_WEF	Array of weights used to compute the incidence angle independent mean WEF for a working area, related to a DGG node.	ESL	Distance in $\rho_{Earth} <$ 1-E-2 meter	None	Analytical methods	N/A	Series of Rho_Earth dependant 1D vectors. Used at DFFG scale.
3.1.1.2	WEF	Array of weights used to compute the incidence angle dependent WEF for a working area, related to a given DGG node.	ESL	Distance in $\rho_{DC} \leq$ 1-E-5	None	Analytical methods	N/A	Series of Rho_DC dependant 1D vectors. Used at DFFG scale.
center	**Miscellaneous Tables**							
3.1.2.1	SKY_RADIATION	Galactic radiation sources TB contribution	L1, External fixed data	< synthetic antenna pattern width ($\approx$ 2.5°)	None	Build from available (0.5°) data		Used at both DGG and DFFG scales.
3.1.3.4	LAND_COVER_CLASSES	Landcover Class Code/ecosystems classes along with soil and vegetation properties	ECOCLIMAP IGBP ESL	Flexible	None	File or table directly provided by ESL	N/A	Used at DFFG scale
center	**DFFG Tables**							

Sec-tion	Table Name	Description	Source	Sampling Requirements	Time Sampling	Generation Method	Quality Check	Notes
3.1.3.1	DFFG_INFO	Provides detailed infor-mation for each DFFG cell, including fraction percentage, landcover class code of fraction	ESL	N/A	None	Generated based on ECOCLIMAP, IGBP Water Fraction, topography index, and other sources See TGRD for details	Consistency check	At DFFG scale.
3.1.3.2	DFFG_XYZ	Provides X, Y, Z Earth reference coordinates in the Earth fixed frame for each DFFG cell	ESL	N/A	None	Geographical coordinates conversion	N/A	At DFFG scale.
3.1.3.3	SOIL_PROPERTIES	Fractions of clay and sand, information on soil bulk density, soil temperature vertical interpolation parameters, transition moisture point and field capacity	Food Agri-culture Organization (FAO), included in ECOCLIMAP package. ISLSCP II	Flexible	None	Resample from 1/12° x 1/12° maps to DFFG @ 4 km x 4km	N/A	At DFFG scale
3.1.3.5	DFFG_LAI_MAX	Max annual LAI	ECOCLIMAP phenology or ESL source	Flexible	10 days	File or table directly provided by ESL	N/A	At DFFG scale.
DGG Tables								
3.1.4.2	DGG_XYZ	SMOS DGG node indexes to Earth reference coordinate (X,Y,Z)	SMOS L1 Auxiliary File	ISEA4-9, 15 km on average	None	Geographical coordinates conversion		At DGG scale
DGG Default Tables								
3.1.5.1	DGG_DEFAULT_FRACTIONS	The pre-computable MEAN_WEF fractions of decision tree classes all around the Earth. This is a placeholder for now	ESL	DGG ISEA4-9, 15 km on average	None	At DGG scale		Pre-computed frac-tions can be used in absence of non-permanent cover types.

3.2.2.2.2   Time varying auxiliary data

Table 14: Overview of TGRD time updated tables

Sec-tion	Table Name	Description	Source	Sampling Requirements	Temporal Sampling	Generation Method	Quality Check	Use and Notes

Sec-tion	Table Name	Description	Source	Sampling Requirements	Temporal Sampling	Generation Method	Quality Check	Use and Notes
Tables Used at Both DFFG and DGG Scales								
3.2.1	ECMWF_ FORECAST	All SM+OS ECMWF products	ESA/E CMWF	0.225° × 0.225°	3 hours	Direct read	ECM WF docs	For surface & atmospheric models, and flags. Content is used both at DFFG & DGG scale and by online mapping in DPM.
DFFG Tables								
3.2.2.1	DFFG_LAI	Leaf Area Index	MODIS MOD15	Flexible	Weekly or 10 days	When the new MODIS product is available, fill its missing LAI values from the previous product. The first-time missing LAIs are filled with phenology.	N/A	Generated at DFFG scale
DGG Currents Tables								
3.2.3.1	DGG_ CURRENT_ TAU_NADIR _LV	Every non-missing value is used as the Tau_Nadir_LV reference value for retrieval for Low Vegetation (LV) fractions (FNO, FWL) of every DGG.	SMOS L2	ISEA 4-9	Every day	When SML2P retrieval of Tau_Nad_LV is possible and accurate, post-processing will update this table with the retrieved values.	Ground truth, time consistency	At DGG scale
3.2.3.2	DGG_ CURRENT_ TAU_NADIR _FO	Every non-missing value is used as the Tau_Nadir_FO reference value for forest fraction FFO.	SMOS L2	ISEA 4-9	Every 4 weeks	When SML2P retrieval of Tau_Nadir_FO is possible and accurate, post-processing will update this table with the retrieved values.	Ground truth, time consistency	At DGG scale
3.2.3.2. 4	DFFG_ CURRENT_ FLOOD	Area where flood may occur	SMOS L2	ISEA 4-9	Every 4 weeks	When area is flooded after high rains and according to channelling, the post-processing will update this table	Ground truth, time consistency	At DGG scale
3.2.3.3	DGG_ CURRENT_ ROUGHNES S_H	Every non-missing value is used as the roughness parameter HR reference value for retrieval.	SMOS L2	ISEA 4-9	Every 4 weeks	When SML2P retrieval of the roughness parameter HR is possible and accurate, then post-processing will update this table with the retrieved values.	Ground truth, time consistency	At DGG scale
3.2.3.4	DGG_ CURRENT_ RFI	RFI statistics	SMOS L1/L2	ISEA 4-9	Every day	Prior knowledge? + When SML2P post-processing diagnostic analysis suspects RFI	Ground truth, time consistency	At DGG scale
3.2.3.5	DGG_ CURRENT_ FLOOD	Flood flags	SMOS L2	ISEA 4-9	Every Day	When SML2P post-processing diagnostic analysis suspects Flood, it is recorded and updated in this table	Ground truth, time consis-tency	At DGG scale

3.2.2.2.3   Auxiliary Users Parameters File

Table 15: Overview of TGRD user parameters tables

Section	Table Name	Description	Source	Sampling Requirements	Temporal sampling	Generation Method	Notes
3.3.1	USER_PARAMETERS	Algorithm configuration data	ESL	N/A	×	File or table provided directly by ESL	To be modified based on the input from ESLs.

3.2.2.3   Auxiliary data pre-processing

This consists of checking auxiliary surface data and interpolating them on a reference grid (DFFG) centered on each L1c node of the SMOS fixed grid (DGG grid), in order to compute fractions through applying the weighting function.

In addition, topography data acquisition and pre-processing will be useful, as it will be done once.

It is probable that the land sea mask will also be processed only once.

3.2.2.3.1   Auxiliary data check

It will also be presumably necessary to address the validity of the auxiliary data. Thus, both ranges and availability will have to be checked.

3.2.2.3.2   Auxiliary data resampling

The description of algorithms requires data to be available either at the DGG and/or DFFG scale depending of their use. We will assume in the rest of this document that these data are available to the right grid scale when needed. For example, if an ECMWF parameter is needed for a given DFFG cell or a given DGG node, we will assume we have it.

Consequently, while resampling or interpolating are necessary tasks, they are not detailed in this document. Details concerning each source of auxiliary data and its specific generation method are addressed in TGRD. The purpose of this section is to provide indications and/or recommendations.

Concerning spatial resampling or interpolation, we have only to consider the following possibilities concerning the spatial resolution of auxiliary data:

- It is coarser than the DFFG, but not needed at DFFG scale. This is only the case for atmosphere contribution that is accounted at the DGG scale. For those data, we recommend a bi-cubic scheme. That concerns surface pressure, Water vapour column content and T2m for atmosphere.
- It is coarser than the DFFG and used at DFFG scale. This is the case for most ECMWF data as well as soil texture data. We recommend disaggregating the auxiliary data cell by replicating their contents to the DFFG (using a "nearest neighbour" scheme).
  A special mention is needed for snow that will be redistributed by filling the highest latitude part of each fraction (unless we can use altitude as well!) except if the use of a specific DFFG snow map is possible with available information (not missing) then a direct use can be performed (same DFFG grid).
- It is close to the DFFG resolution but defined on a different grid. This could be the case for LAI or topography … We need to resample those auxiliary data to the reference grid through an interpolator. We recommend using a bi-cubic interpolation.

There are many possible interpolating schemes. An overview for spatial interpolation can be found in a DEIMOS technical note ([31]; section 2.5, Resampling Approach). This document gives formulas and advantages / drawbacks from both computer and mathematical standpoints.

Concerning time interpolation, it is not foreseen to operate a coupled space and time interpolation i.e. to interpolate evolving auxiliary data values at each space and time SMOS measurement. Besides, we should also deal with the multi-temporal characteristic of the L1c product that is made up of multiple snapshots measurements at different observing times.

For now, we are only concerned with updating ECMWF forecast data. For some smooth parameters such as pressure, perhaps time interpolation could be done. However, for others, like snow, the meaning or the validity of such an approach would be

doubtful. Therefore, we recommend using a **nearest neighbor time scheme** applied to each half orbit as a whole, based on the UTC time when crossing the equator[13].

In a sequence of 3 hourly forecast steps, we have to select the forecast step time the closest to the half orbit UTC time when crossing the equator. Assuming that a SMOS half orbit is ~50' long, the worst case, in time lag between the forecast step time and the SMOS observing time, occur when the UTC time of the orbit cross the equator coincide with a forecast half step time (step time plus 1h30). Using that scheme leads to a maximum time lag of 1h55' between observation and forecast times.

3.2.2.3.3   Pre-processing in order to obtain DGG quantities

- Default fractions FM$_0$ used by the decision tree
  They are static except for non permanent (NPE) frost, snow zone and possibly water events; Figure 8 (used in [119]) shows the concerned areas as defined using ECOCLIMAP climatology [32] and ISLSCP-I data.
  It is possible to compute **default fractions** over the DGG assuming neither (NPE) snow nor frost. Since these default fractions will always remain valid over areas shown as blank on these maps, they can be computed offline and thus provided directly through tables in TGRD, given the L1c node number.
  This option will save pre-processing CPU time. The fractions will be computed on the fly only if auxiliary data report snow or frozen situations or possibly flood conditions.
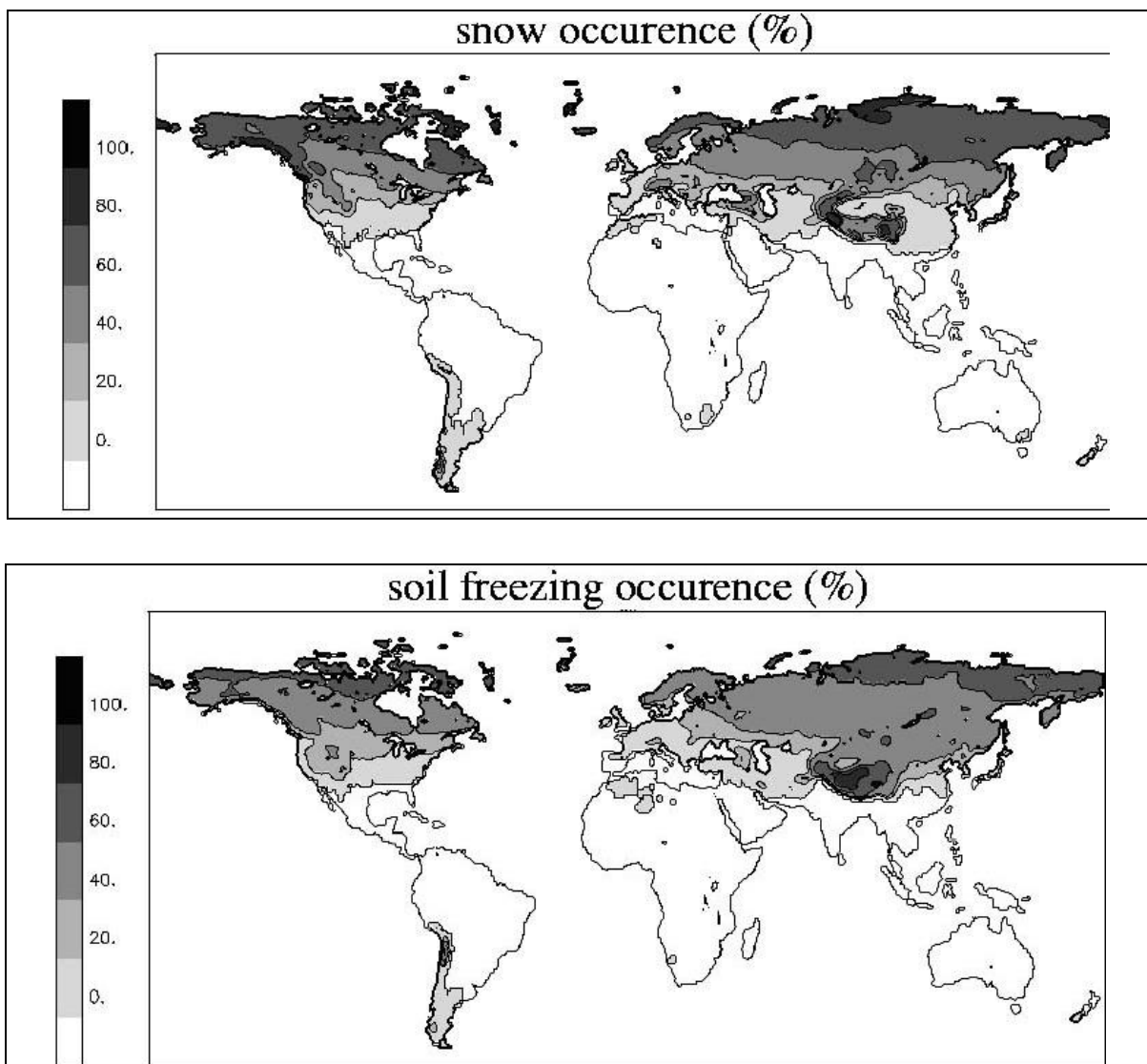
Figure 8: Spatial frequency distribution of (top) snow and (bottom) soil freezing

---

[13] Not necessarily in line with the current operational ECMWF pre-processor

- Default reference parameter values

  As a consequence of the aforementioned, it is thus possible to apply the decision tree offline for each fraction over the SMOS grid assuming neither NPE snow nor frost. In those cases, all the reference values coming from static auxiliary data and model selection will always remain valid over areas shown as blank on these maps and can be provided directly through TGRD tables, given the L1c node number.

  This option will save some CPU time, since applying the decision tree is lighter than computing the fraction. It might be extended to parameters that enter directly the forward models: optical depths, soil structure…

- Local DFFG working areas

  Cases where a single fraction is present can be identified from the default fractions computation presented above, over zones where NPE features do not occur.

  In other cases, **local subsets of DFFG arrays** are needed around the DGG node being processed. The local subset is a square of size = WEF_SIZE, made up of DFFG cells, that defines the DFFG **working area**, WA$_{DFFG}$, illustrated Figure 9

  The suggested value (WEF_SIZE=123 km) corresponds to twice the largest extent of the 3dB footprint occurring in SMOS soil moisture observations.

  One side of the square can be taken parallel to the local meridian.

WEF_SIZE=123 km
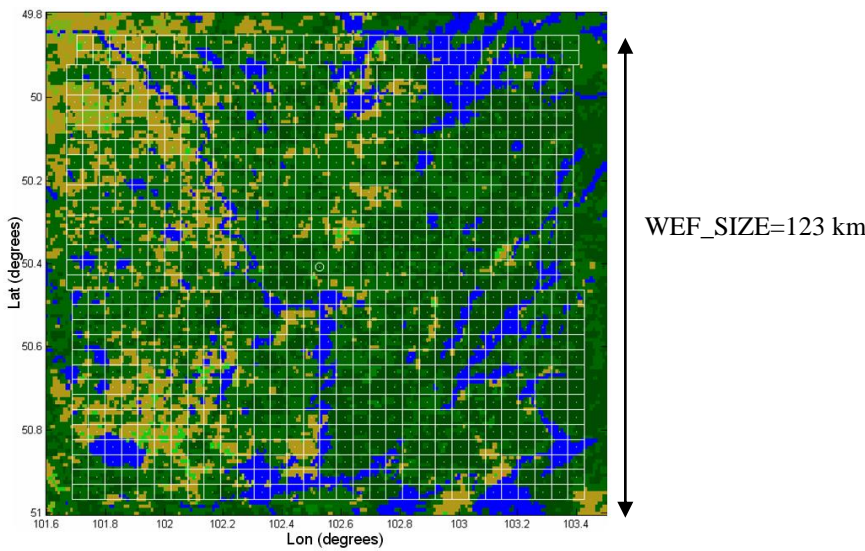
Figure 9: DFFG pixels in a Working Area over the aggregated landcover classes

3.2.2.4 Obtaining the incidence angle dependent weighting function WEF

In the general case, weighting functions are necessary over land surfaces for 3 purposes:

- Compute fractions **FM0** of various land types, which are used to drive the decision tree.
- Compute reference values (to be used as a priori or default) for each mean fraction or group of fractions **FM**.
- Apply incidence angle dependent fraction (**FV**) weighting to each L1c view for building the aggregated forward model used in the iterative retrieval and to be used during retrieval.

Concerning tasks 1 and 2, use is made of an average MEAN_WEF function for which an analytical formulation is given below in 3.2.2.5 and a tabulation will be described in the TGRD. The present section addresses point 3, where an **incidence angle dependant WEF** function is needed.

The weighting functions are to be applied to the cells of DFFG working area and parameters analyzed over it.

3.2.2.4.1 Rigorous formulation of the WEF

For each snapshot, depending on whether the pixel is dominated by land surface or ocean (or possibly other considerations), an apodization window is selected during the L1 processing. This quantity is close to the synthetic antenna pattern, which drives the angular resolution of the SMOS interferometer.

The synthetic antenna pattern, also called Equivalent Array Factor EAF, is given by [120]:

$$\text{AF}_{eq}\left(\xi, \xi', \eta, \eta'\right) = \frac{\sqrt{3}}{2} d^2 \sum_m \sum_n W\left(u_{mn}, v_{mn}\right) \cdot \tilde{r}\left(-\frac{u_{mn} \cdot \xi + v_{mn} \cdot \eta}{f_0}\right) \cdot e^{j2\pi\left(u_{mn}\left(\xi-\xi'\right)+v_{mn}\left(\eta-\eta'\right)\right)} \quad \textbf{Eq 77}$$

86

where

- W is the apodization function
- r is the fringe-washing factor (**FWF**) which accounts for the spatial decorrelation between antennas.
- u,v are the baseline coordinates in the frequency domain
- d is the antenna element spacing (= 0.875)
- f_0 is the central frequency (1413 MHz)
- ξ, η are the central director cosines (DC) coordinates; ξ', η' are running DC coordinates.

For the nominal processing, were it not for the FWF factor, the EAF would be the same everywhere in the DC plane, i.e. would not depend on ξ and η but only on ξ' and η' (For strip adaptive processing, a specific EAF must be computed for each node of the fixed grid: see [121]).

While its central part is centro-symmetric on antenna boresight ($\xi = \eta = 0$), the EAF has significant side-lobes which are either positive or negative and are no longer symmetrical off boresight.

Furthermore, going from DC to polar coordinates, for nominal processing, the EAF becomes elongated as the angular distance to boresight increases.

The weighting function WEF is obtained through intersecting the EAF with Earth surface. To this end, two further steps are needed, which account for:
- a "smearing effect" due to integration along the track;
- the variation of the integrating element with incidence angle, as the fine grid area does not lie on a plane.

3.2.2.4.2 WEF approximations

The results of the Soil Moisture retrieval Study [122] suggest that, for well-behaved APF functions (Blackmann or better), **truncating** the exact APF to the main lobe does not generate significant errors for representative scenes. This is a worthwhile option to be considered since it restricts the domain over which the APF and WEF (see below) should be computed.

Other approximations have been tested [123] through numerical retrieval simulation. As a result, it is found possible to:
- ignore the FWF factor;
- approximate the APF by a centro-symmetric function, for which a simple analytical formula can be fitted;
- ignore both smearing effect and Earth sphericity on the scale of the fine grid area.

Then, the following simplified expression WEF_A for the WEF is proposed:

$$WEFA(\rho_{DC}) \approx \frac{\text{sinc}(C_{WEF1} \cdot \rho_{DC})^{C_{WEF2}}}{1 + C_{WEF3} \cdot \rho_{DC}^{C_{WEF4}}} \text{ if } C_{WEF1} \cdot \rho_{DC} \leq \pi$$

$$WEFA(\rho_{DC}) = 0 \text{ otherwise}$$

Eq 78

where

- $\rho_{DC} = \text{sqrt}((\xi' - \xi)^2 + (\eta' - \eta)^2)$ is the distance in the DC coordinates
- $\text{sinc}(x) = \sin(x)/x$ ($x \neq 0$); $\text{sinc}(0) = 1$

Values for $C_{WEF}1$ to 4 are supplied in TGRD UPF.

Note: For computational efficiency, the WEF_A should be computed inside the main lobe. Otherwise, it results in negative values of the sinc that will generate complex values results when applying the exponent, as it is shown in Eq 78..

WEF_A will be provided as an auxiliary table in TGRD using the above formula for its generation method:
- First, it will increase the computational efficiency; the WEF_A function can be tabulated and we suggest to sample ρ_{DC} with a 10^{-5} step on the interval $[0, \pi/C_{WEF1}]$. In that case, the accuracy is better than 5×10^{-4}, which is sufficient.
- Second, it will provide a flexible way to adapt to different centro-symetric shapes without changing the processor code.
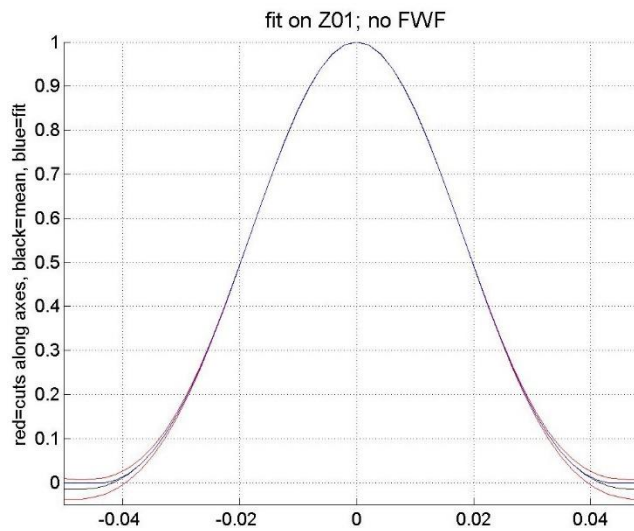
Figure 10: Fit quality over main APF lobe (cut)   **Figure 11: Fit quality (image)**
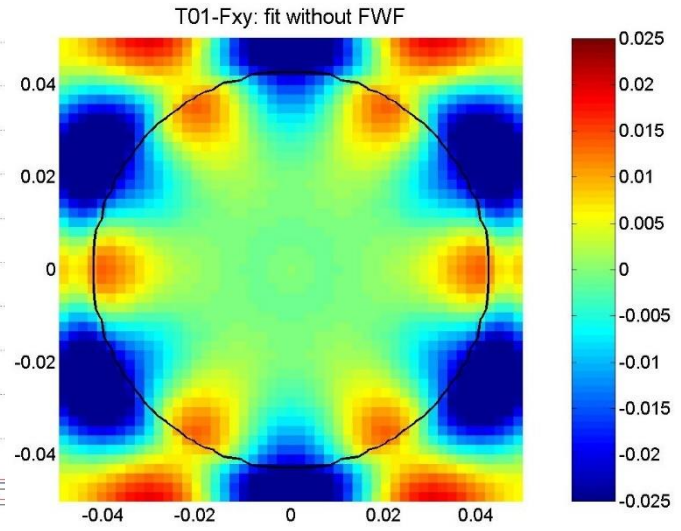
Figure 10 (cut) & Figure 11 (image) illustrate the quality of the fit over the main APF lobe. The RMS difference is about 0.6% (with maximum value of APF normalized to 1).

The WEF_A provides directly the weighting function.

For every DGG node, the DC ξ and η must first be obtained. Then, for every DFFG cell within the WA_{DFFG} defined around an L1c node, the remaining task is to compute ρ_{DC}.

3.2.2.4.3 Incidence angle dependent fractions
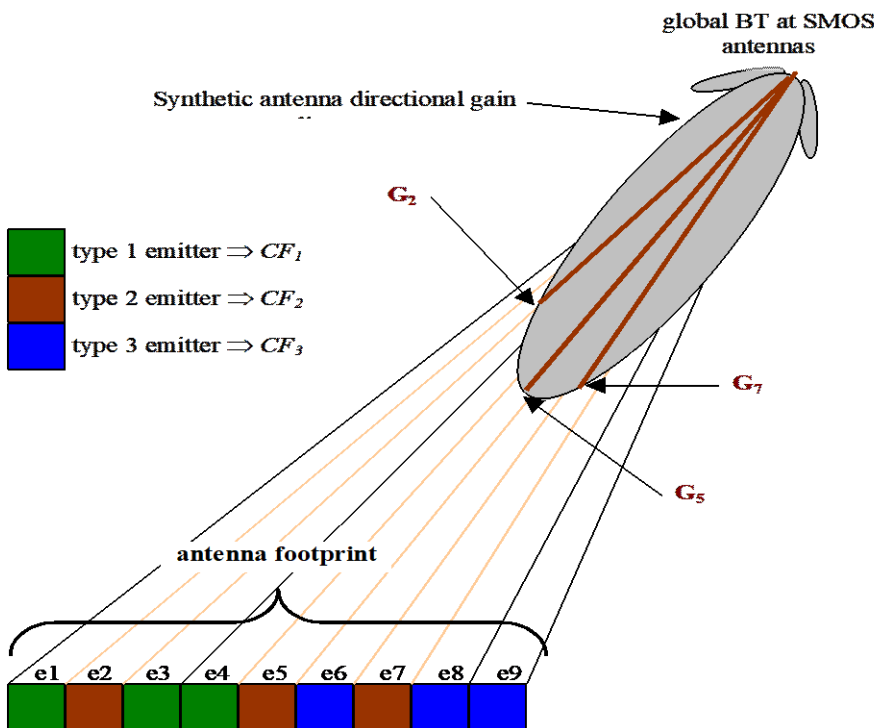
Cover fractions are not straightforward surface ratios because the SMOS observed TB is obtained from integrating radiance through the (directional) synthetic antenna pattern, as shown in the adjacent figure. This figure can be seen as a simplified vertical slice of the SMOS observation of the scene depicted Figure 3.

The example illustrates that the full measured TB is a weighted sum of intra-footprint TB for each class of emitters. The sum should be carried out over the WA_{DFFG} surrounding the concerned DGG node. Due to the geometry of observation, the weights differ for each incidence angle.

Figure 12: Cover fractions

The DFFG integrating zone (WA_{DFFG}) must be such that it includes all cases where the WEF takes non-zero values. The estimated required size is a square area with size WEF_SIZE (see 3.2.2.3.3).

3.2.2.5　　　　　Computing average fractions

However, prior to the retrieval, **angle independent fractions** FM_0 must be computed in order to allow running the decision tree and selecting the processing options.

The MEAN_WEF uses basically, as its main component, a centro symmetric analytical approximation identical to the one used in the WEF (see Eq 78). In order to represent the **average** SMOS pixel, it is scaled in such a way that its half maximum contour is 40 km wide, which corresponds closely to the average size of the SMOS pixel over the FOV.

On the other hand, it is necessary that when identifying fractions likely to be present no one be ignored, in such a way that reference values and default contributions can be computed if necessary, for some particular incidence angles. To this end, a flat circular disk, with diameter equal to WEF_SIZE, has been added to the MEAN_WEF. This "background component" ensures that no fraction is left out, since it accounts for the maximum size of the actual pixel in any direction.

The height of this additional background has been adjusted in such a way that the total weight outside the main component does not exceed 2% of the overall integral. Then, any **new** land cover contribution from the background component will never be large enough to modify the results of applying the decision tree.

Figure 13 and Figure 14 show a map (the background being slightly enhanced) and a cut (semi log scale) across the MEAN_WEF.
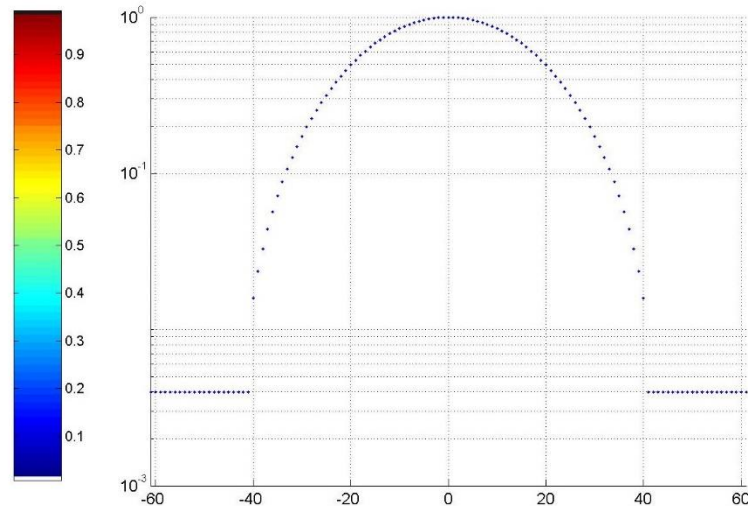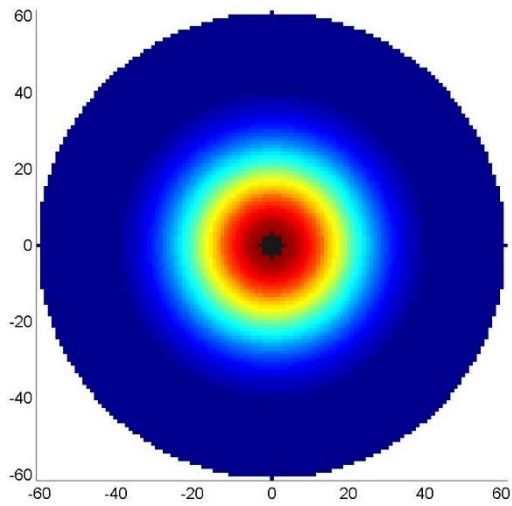
Figure 13: MEAN_WEF: image　　　　　　　**Figure 14: MEAN_WEF: semi log cross-cut**

The MEAN_WEF must be normalized to unity integral and applied to land cover in order to compute mean fractions.

The above description corresponds to the MEAN_WEF being defined over a DFFG geographical grid expressed in km. The following formulation can be used for its actual computation:

$$MEAN\_WEF(\rho_{earth}) = C_{MWEF2} + WEF_A\left(\frac{\rho_{earth}}{C_{MWEF1}} \cdot \frac{\pi}{C_{WEF1}}\right) \text{ for } \rho_{earth} \in \left[0, C_{MWEF1}\right]$$

$$MEAN\_WEF(\rho_{earth}) = C_{MWEF2} \text{ for } \rho_{earth} \in \left]C_{MWEF1}, \frac{WEF\_SIZE}{2}\right] \qquad \textbf{Eq 79}$$

$$MEAN\_WEF(\rho_{earth}) = 0 \text{ otherwise}$$

Same as for WEF, the MEAN_WEF will be used as a tabulated form of the above equation and thus defined in TGRD.

The values for C_{MWEF1} (40 km) and C_{MWEF2} (0.02) are supplied in TGRD UPF.

Both WEF and MEAN_WEF are thus used to compute aggregated fractions from the land cover array over a fine grid area. It will be seen that the land cover includes two classifications: a complementary one, and a superimposed supplementary one, which accounts for possible NPE conditions as well as the topographic mask.

- Since the complementary classification covers the whole area, rules are necessary to "blend in" the supplementary classification. This topic is addressed in the decision tree section.

- In a later step, the WEF will be used to compute reference parameter values for every quantity relevant and for each view for building the forward models. This is also addressed in the decision tree section.

3.2.2.6 Geometric vs. radiometric fractions or mean usage

When we are working inside the DFFG pixel, all is defined with respect to **geometric fractions**. These fractions sum to 100% over the entire fractions list inside a DFFG pixel. They correspond to what is usually termed by "pixel". However, in our case, the sensor has a gain pattern meaning that some areas (boresight) will contribute more to the signal than others. To account for this, the signal emitted by the surface has to be weighed by the antenna gain pattern. It can be neglected at the center but has to be taken into account the more we go away from boresight.

Consequently, when considering DFFG, the only meaningful quantities are geometric ones since they are defined before introducing any WEF. When considering WA, weighting by WEF quantities becomes a natural step.

So, when we are working externally to a DFFG pixel, e.g. at the working area level, all is defined with respect to **radiometric fractions** which are the product of the geometric fractions times the normalized antenna gain patterns, either the "true" one provided through the WEFs or the average one provided through the MEAN_WEF.

A **mean radiometric fraction** is just the sum of all the radiometric fractions of a given class. It sums to 100% over all the DFFG pixels of the working area and all the classes fractions of the DFFG pixels.

Similarly, the concept of **geometric mean** and **radiometric mean** is applied when computing average values. Geometric mean are specifically used for DFFG non-permanent effect rules (NPE) whereas radiometric mean are used in all the other cases including working area NPE rules, prior value computation and reference value computation.

Prior values and references values are computed from surface distributed parameters within the working area as radiometric means before being used in models: e.g. SM MEAN_WEF radiometric mean for the model associated to the fraction to retrieve or SM WEFs means for the defaults models associated to the default fractions as for all other parameters like omega, QR, NRH, NRV etc …

There are two exceptions: effective temperatures and optical thicknesses are computed first locally to the DFFG pixels using their parameters (SM, w_0, bw0, b', b", etc …) local to the DFFG pixel and then are averaged using MEAN_WEF and/or WEFs.

3.2.3 Decision tree

3.2.3.1 Content of the decision tree section

- The decision tree procedure begins with determining the weighted mean aggregated fractions FM_0, which are to be considered in the decision tree, as well as those FM which contribute to modelled radiometric contributions.
- A battery of tests is defined, based on a series of thresholds concerning the magnitude of various fractions, and allows defining the branches for the 1st stage of the decision tree.
- The fraction (s) and model (s) selected for retrieval or default contributions are selected for each branch.
- From auxiliary data, reference values of either fixed parameters or a priori constraints are obtained for every relevant fraction of the pixel.
- Options (stage 2 set of branches) are finally chosen for the retrieval, concerning the number and nature of parameters to be actually retrieved as well as a priori standard deviations.

3.2.3.2 Computing aggregated fractions

We now consider a specific DGG node, and the WA_{DFFG} that surrounds the DGG node being considered (as defined in section 3.2.2.5).

3.2.3.2.1 Content of DFFG pixels

An initial **land cover classification surface fractions** on the DFFG is given by a LUT described in TGRD. This LUT is fixed and has been generated once and for all for a given DFFG resolution from the aggregation of landcover detailed classification, eventually complemented with other information such as high-resolution water information and topography (see TGRD generation methods for details).

Let's consider a given DFFG pixel cell that belongs to WA_{DFFG}. This pixel contains a list of sub-pixel surface fractions (%) that correspond to the list of **FM complementary** classes presented in Table 16. Each FM class is also associated with a detailed landcover classification code that will be used later as an index to the LAND_COVER_CLASSES LUT to access surface

properties for computing reference values. One then refers to those classes as **complementary**, in the sense that together they cover the whole surface of the DFFG pixel.

Besides, this list is completed by the topography features of the DFFG cell: fractions of moderate and strong topography, FTM and FTS. Opposite to the previous fractions, FTM and FTS are considered as **supplementary**, in the sense that in order to introduce them, one has to override the previous complementary fractions in such a way that the resulting set becomes complementary.

Table 16 shows first the fractions **FM0** used to drive the decision tree, next the fractions **FM** used to compute reference values and TB contributions.

FM0 and FM classes are referred to as indicated in columns A and C.

Table 16: Aggregated fractions FM$_0$ and FM

FM0 class	Aggregated land cover	FM class	Complementarity	
A	B	C	D	E
FNO	Vegetated soil + sand	FNO	Comple-mentary	Sum of comple-mentary fractions equals unity
FFO	Forest	FFO		
FWL	Wetlands	FWL		
	Open fresh water	FWP		
	Open saline water	FWS		
FWO	Open water			
FEB	Barren	FEB		
FTI	Total Ice fraction	FTI		
	Ice & permanent snow			
	Sea Ice			
FEU	Urban	FEU		
FTS	Strong topography		Supple-mentary	Supple-mentary fractions are super-imposed
FTM	Moderate topography			
FRZ	Frost	FRZ		
FSW FSM	No permanent dry snow	FSN		
	No permanent wet snow			
	No permanent mixed snow			

3.2.3.2.2 Accounting for external conditions to update the DFFG pixel content

The initial aggregated classes and surface fractions describe the fixed features of WA$_{DFFG}$. These features are subject to change due to eventual non-permanent (NPE) surface conditions. We handle this situation with the three following actions to apply **for each DFFG pixel** of any WA$_{DFFG}$:

- First, we need to complete the initial set of classes with categories corresponding to frozen and snow-covered zones since they are handled with specific models. As for topography, NPE fractions are considered as supplementary.

- Next, we eventually need to reallocate fractions and/or need to change other surface properties depending on the above events and possible others. For example, if the temperature drops under a certain level, FWP becomes FEI, if the low vegetation becomes frozen then its optical thickness is forced to 0, …

- In certain cases, some actions have also to be taken on reference values

- Always satisfy the constraint that all the final fractions of the DFFG pixel are complementary.

- The next subsections indicate the actions to be performed and the triggers that activate them. It is important to follow the order of the subsections.

3.2.3.2.2.1 Soil to frozen

When the ECMWF surface soil temperature (STL1) drops below a given threshold, all types of soil surface classes are assimilated to the frozen class. The mean optical thicknesses of the affected classes weighted by their initial fractions feeds the optical thickness of the frozen class.

Trigger: $STL1 < TH\_SOIL\_FRZ$

Actions on reference values: $TAU\_FRZ=(FNO*TAU\_FNO+FFO*TAU\_FFO+FEU*TAU\_FEU) /(FNO + FFO + FEU)$

Actions on fractions values: $FRZ = FNO + FFO + FEU; FNO=FFO=FEU =0$

3.2.3.2.2.2   Sea ice

When ECMWF reports directly the presence of sea ice (CI), the saline water and pure water are decreased proportionally.

Trigger: $FSI \neq 0$

Actions on reference values: $TAU\_FSI=(FWP*TAU\_FWP + FWS*TAU\_FWS)/ (FWP + FWS)$

Actions on fractions values: $FWS=(1-FSI)*FWS, FWP=(1-FSI)*FWP$

3.2.3.2.2.3   Pure water to ice

When the pure water bodies temperature drops below a given threshold, fractions of classes related to pure water are assimilated to the ice class. The mean optical thicknesses (if any) of the pure water classes (FWL, FWP) weighted by their initial fractions add to the optical thickness of the ice class.
If the ECWMF cell is water dominant (LSM < 0.5) then the temperature to use is the sea surface temperature (SSTK), otherwise if the ECWMF cell is land dominant (LSM >= 0.5) we use the two meters temperature (2T).

Trigger: $(LSM>0.5$ **and** $2T< TH\_PWATER\_FRZ)$ **or** $(LSM<0.5$ **and** $SSTK<TH\_PWATER\_FRZ)$

Actions on reference values: $TAU\_FEI=(FEI*TAU\_FEI + FWL*TAU\_FWL + FWP*TAU\_FWP)/ (FEI+FWL+FWP)$

Actions on fractions values: $FEI=FEI+FWL+FWP; FWL=FWP=0$

3.2.3.2.2.4   Saline water to sea-ice

When the saline water bodies temperature drops below a given threshold, the saline water class fraction is allocated to the sea ice class. The mean optical thickness above the saline water replaces the optical thickness above the sea ice class.
If the ECWMF cell is water dominant (Land Sea Mask LSM < 0.5) then the temperature to use is the sea surface temperature (SSTK), otherwise if the ECWMF cell is land dominant (LSM >= 0.5) we use the two meters temperature (2T).

Trigger: $(LSM>0.5$ and $2T< TH\_SWATER\_FRZ)$ **or** $(LSM<0.5$ and $SSTK<TH\_SWATER\_FRZ)$

Actions on reference values: $TAU\_FSI=(FSI*TAU\_FSI + FWS*TAU\_FWS) / (FSI + FWS)$

Actions on fractions values: $FSI=FSI + FWS; FWS=0$

3.2.3.2.2.5   Mixed snow cover

When mixed snow exists, it is assumed it will remain on several surface types (FNO, FFO, FEI, FSI, FRZ) while it melts and disappears rapidly on others that consequently remain unchanged (FWS, FWP, FWL, FEB, FEU). When the fraction of mixed snow (FSM) within the DFFG cell is not 0%, then FSM is distributed uniformly on all fractions but only the first list will contribute to the final redistributed fraction of mixed snow.

Trigger: $FSM \neq 0$

Actions on reference values: $TAU\_FSM=(FNO*TAU\_FNO + FFO*TAU\_FFO + FEI*TAU\_FEI + FSI*TAU\_FSI + FRZ*TAU\_FRZ +) / (FNO + FFO + FEI + FSI + FRZ)$

Actions on fractions values: $FNO = FNO*(1-FSM), FFO = FFO*(1-FSM), FEI= FEI*(1-FSM), FRZ=FRZ *(1-FSM)$

3.2.3.2.2.6   Wet snow cover

When wet snow exists, we consider it remains on some surface (FNO, FFO, FEI, FSI, FRZ) while it melts and disappears rapidly on others that consequently are not affected (FWS, FWP, FWL, FEB, FEU). When the fraction of mixed snow (FSW) within the DFFG cell is not 0%, then FSW is distributed uniformly on the all fractions but only the first list will contribute to the final redistributed fraction of wet snow.

Trigger: FSW ≠ 0

Actions on reference values: TAU_FSW=(FNO*TAU_FNO + FFO*TAU_FFO + FEI*TAU_FEI + FSI*TAU_FSI + FRZ*TAU_FRZ +) / (FNO + FFO + FEI + FSI + FRZ)

Actions on fractions values: FNO = FNO*(1-FSW), FFO = FFO*(1-FSW), FEI= FEI*(1-FSW), FRZ=FRZ *(1-FSW)

3.2.3.2.2.7   Low vegetation becomes transparent

When the water content of the standing vegetation freezes, there is no more absorption and the optical thickness becomes negligible for every surface except forests. SKT is the ECMWF Skin Temperature.

Trigger: SKT < TH_VEG_FRZ

Actions on reference values:

TAU_FNO=TAU_FWL=TAU_FWP=TAU_FWS=TAU_FEB=TAU_FEI=TAU_FSI=TAU_FEI=TAU_FRZ=0

Actions on fractions values: none

3.2.3.2.3   Obtaining the final integrated radiometric fractions for WA$_{DFFG}$

The final integrated fractions FM and FM0 for the working area WA$_{DFFG}$ are simply computed using a weighted mean of the local DFFG pixel fractions over the set of DFFG pixels that belongs to WA$_{DFFG}$. The MEAN_WEF or WEF values are the weights to be used.

Finally, similarly to the previous section, we will apply the following rule to handle the **forest winter exception case** (formerly case 12 of DT). But this time it applies to the global fractions associated with WA$_{DFFG}$.

Branches and thresholds of the decision tree are listed in tables 17 and 18 below.

When practically no vegetation exists in WA$_{DFFG}$ except for the forest fraction and the main fraction to retrieve is a nominal (quasi bare soil) one, while the second main fraction is forest, then the retrieval is performed on the forest fraction instead of the main nominal one. The rationale being that it should enable getting a better estimate of Forest vegetation opacity. To force the decision tree to behave like this, we move the minimum amount of FNO to FFO to insure that DT#12 case is abandoned in favour of DT#13 case.

This rule aims to have a better estimate of the optical thickness of forest than the proxy using the maximum LAI. As a result, the retrieved optical thickness will be the forest one. While it is obtained at the expense of the retrieved SM quality, this kind of retrieval is also expected to be done only once, in order to feed DGG_CURRENT_TAU_NAD_FO for the current DGG that will change from a missing value (-999) to a plain one forever or until manually changed.

Trigger: (TAU_NOFFO < TH_TAU_F1) **and** (TH_F1 < FFO < TH_F2) **and** (FNO > TH_NO) **and** (DGG_CURRENT_TAU_NAD_FO == -999)

With TAU_NOFFO being the mean optical thickness of all the 10 fractions but forest. TAU_NOFFO = (FNO*TAU_FNO + FWL*TAU_FWL + FWP*TAU_FWP + FWS*TAU_FWS + FEB*TAU_FEB + FTI*TAU_FTI + FEU*TAU_FEU + FRZ*TAU_FRZ + FSN*TAU_FSN) / (FNO + FWL + FWP + FWS + FEB + FTI + FRZ + FSN)

Actions on reference values: none.

Actions on fractions values (follow the order):

- FNO = FNO - (TH_F1 - FFO)
- FFO = FFO + (TH_F1 - FFO) = TH_F1

3.2.3.2.4   Open water

It now appears that the direct models for fresh and saline water will be **slightly different** (section 3.1.3). Then, it becomes necessary to compute and store two distinct FM fractions, FWS and FWP, respectively for saline and fresh water. At the same time, a single FM$_0$ fraction with FWO = FWS + FWF is needed to drive the decision tree.

In the future, these static water fractions may be improved to deal with flooded areas by using information from an auxiliary DGG_CURRENT_FLOOD map, and introducing an additional flag FL_FLOOD_PRONE (see section 3.7.3).

3.2.3.2.5   Sea Ice

Concerning **time-varying sea ice cover**, no dedicated provision is made as no pure sea nodes were to be processed. Should they be available outside the coastal limit they will have to be processed. For landcover sea ice categories it is considered as "sea". It will thus be necessary to rely on ECMWF fields.

The idea is to define a new fraction for SEA ICE (FSI) derived from ECMWF fields. This fraction is then used jointly with the permanent ice and snow fraction (FEI) as an ICE fraction FTI (called FTI in a similar fashion to the case of fresh and saline water).

In consequence, at high latitudes, for a given node, the landcover ice fraction FEI is added to the sea ice fraction from ECMWF FSI and the general structure of the decision tree is not affected.

3.2.3.2.6   Topography

The TGRD indicates that several topography indexes will be available. The idea is that **strong** topography is often associated with rocks, snow, ice and high slopes. Above a relatively low threshold, the pixel is processed with the cardioid.

Softer topographies have less slopes, rocky outcrops, etc) and more vegetation and "nominal" surfaces. A higher topography fraction can be thus tolerated but a flag is still needed. Eventually a correcting factor (H-Q type as defined in [124]or used in [15] see 3.1.2.1) could even be considered.

From the topography LUT, three classes are identified, featuring (i) weak or non-existent, (ii) moderate, and (iii) strong topography. Hence, based on thresholds TH_TS and TH_TM to be compared to the mapped topography index, **two** fractions will be computed, FTS and FTM, for strong and moderate topography respectively. The FTM **includes** FTS. These are FM0 fractions only, as long as no direct model is available for strong topography.

3.2.3.2.7   Snow

The given snow cover parameter is a **percentage**.

Snow cover percentage obtained from two possible sources:

ECMWF snow depth (equivalent meter of water), SD. Using the SCR ECMWF IFS parameter, the minimum snow mass that ensures complete coverage, it defines the threshold value for 100% snow-covered grid box. SCR is provided in TGRD UPF and has a default value set to 0.015 m.

Snow cover percentage = min(1, SD / SCR)*100

DFGG snow map which provide directly the fraction of snow at the DFFG cells.

If possible, the preference is given to the DFFG snow fractions map (if activated), next to the ECMWF derived fractions.

If the snow percentage is smaller than 100%, it is to be applied to **each** pre-existing aggregated land fraction, selecting for each of them the **zone with the highest absolute latitude**.

There are three possible cases (see section 3.1.4.3): dry, wet and mixed, depending on the value of the relevant surface temperature T_SNOW when compared to 2 thresholds TH_TDRY and TH_TWET:

- T_SNOW < TH_TDRY: dry snow
- Dry snow is transparent in L-band, it is thus ignored in the processing: FSN=0 and FSD = snow cover percentage. FSD is used for flagging purposes.
- TH_TDRY ≤ T_SNOW < TH_TWET: mixed snow
- FSN = FSM = snow cover percentage TH_TWET ≤ T_SNOW: wet snow.
- FSN = FSW = snow cover percentage

Suggested values (TGRD UPF) are TH_TDRY = - 12°C, TH_TWET = - 2°C.

The ECMWF field relevant to estimate T_SNOW is "TSN" in the GRIB (see TGRD). T_SNOW will be estimated over the **whole** SMOS pixel, in such a way that snow cover belongs to the **same** category (dry, wet of mixed) everywhere; there is thus a single (FM and FM0) FSN snow fraction. Only wet and mixed snow types are considered, while the dry snow is only indicated for flagging purposes.

A future alternative for assessing snow category and fraction to feed the DFFG snow maps are SSM/I data or more recently the NSIDC IMS composite products.

3.2.3.2.8   Overriding rules for fractions FM$_0$

The additional classes (high topography, frost and snow) are characterized as supplementary, in the sense that in order to introduce them one has to override the previous complementary classes in such a way that the resulting set becomes complementary. The overriding order and extent must be specified:

- First high topography (excluding FWO)

- Next frost (excluding FWO)

- Finally snow (excluding FWO), **only** for wet or mixed snow (see section 3.1.4.3).

The function **MEAN_WEF** over the DFFG local subset is applied to this analysis of land cover in order to obtain mean fractions FM_0.

Slightly different rules are given below for computing parameters and incidence angle fractions FM and FV.

3.2.3.3 First stage of the decision tree

3.2.3.3.1 Thresholds

Each branch of the tree is determined by testing whether a fraction F** in the FM0 list (Table 16) is larger than a specific value: F** > THVAL * FREF, where

- THVAL is a threshold TH_*** (%)

- FREF is the fraction of reference that depends on a key TH_***_D and is either:

  - equal to 1, when the threshold value is meant for the whole fraction,

  - equal to the land total fraction FLA=1 – FWO, when the threshold value is meant for the land fraction.

- A table in the UPF (TGRD) provides:

- the value of each threshold TH_***

- the key TH_***_N which defines the relevant fraction F** (column A in table 14)

- the key TH_***_D for defining the fraction of reference (whole area or FLA only).

- the rank TH_***_R assigned to each test when operating the tree

Table 17 illustrates the list of thresholds and corresponding FM0 fractions. The actual values are given in the TGRD UPF.

Table 17: Decision tree stage one thresholds

Aggregated FM0 fractions		Acro	Value %		Acro	Value %
Open water		TH_W1	60		TH_W2	90
Topography		TH_TS	5		TH_TM	40
Frost		TH_R1	5		TH_R2	98
Snow (if wet)		TH_S1W	5		TH_S2W	98
Snow (if mixed)		TH_S1M	5		TH_S2M	98
Forest		TH_FF	40		TH_F2	60
Soil		TH_NO	40			
Wetlands					TH_WL	90
Barren					TH_EB	80
Ice sea-ice or permanent snow					TH_TI	80
Urban					TH_UH	80

Note: the low threshold TH_UL   is not considered in the decision tree.

3.2.3.3.2 Decision tree branches

Table 18 shows the selection of tree branches, according to the threshold tests, **in the stipulated order**. This is meant as an illustration; the order for carrying out the tests (1rst column) is provided by TGRD UPF. This order is what matters, even though some numbers are skipped.

Assuming there are 16 thresholds (that is, the maximum value NB_TH_DEC of TH_***_R is 18), there are 17 branches, with the last one corresponding to the case where **no** threshold whatsoever has been exceeded.

Table 18: branches of stage one decision tree[14]

N° / Retrieval case	FWO > TH_W2 *1	FWO > TH_W1 *1	FTS > TH_TS *FLA	FTM > TH_TM *FLA	FSW > TH_S2W *FLA	FSM > TH_S2M *FLA	FSW > TH_S1W *FLA	FSM > TH_S1M *FLA	FRZ > TH_R2 *FLA	FRZ > TH_R1 *FLA	FFO > TH_F2 *1	FNO > TH_NO *1	FWL > TH_WL *FLA	FEB > TH_EB *FLA	FTI > TH_TI *FLA	FEU > TH_UH *FLA
1 All open water	1															
2 Heterogeneous OW	0	1														
3 Strong topo pollution	0	0	1													
4 Soft topo pollution	0	0	0	1												
5 All wet snow	0	0	0	0	1											
6 All mixed snow	0	0	0	0	0	1										
7 Wet snow pollution	0	0	0	0	0	0	1									
8 Mixed snow pollution	0	0	0	0	0	0	0	1								
9 All frost	0	0	0	0	0	0	0	0	1							
10 Frost pollution	0	0	0	0	0	0	0	0	0	1						
11 Forest cover	0	0	0	0	0	0	0	0	0	0	1					
12 Soil cover[15]	0	0	0	0	0	0	0	0	0	0	0	1				
13 All wetlands	0	0	0	0	0	0	0	0	0	0	0	0	1			
14 All barren	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
15 All ice	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
16 All urban	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
17 Heterogeneous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3.2.3.4 Select forward models

3.2.3.4.1 Structure of forward models

Each model used in retrieval includes several "modules". They differ only by the way of modelling the dielectric surface layer. As shown on Table 19, there are presently 4 models.

The water model MW includes two variants MWS, MWP (without/with salinity). The nominal model may include later a variant for **sand**.

The model MS for mixed snow is not validated at the time of writing this version of ATBD.

---

[14]Highlight colours: blue means water body, dark green forest, light green low vegetation and yellow special surfaces
[15] Highlighted in yellow to remind that DT12 is our nominal surface for soil moisture retrieval.

Table 19: Structure of forward models

Forward models	Dielectric constant formulation	Roughness	Snow layer	Vegetation layer	Atmo/Iono/Sky
MD	(A,U) cardioid	identical	none	identical	Identical
MN	L-MEB SM dependent				
MWP, MWS	Water, SST dependent				
MS	MD or MN if transparent		TBD after val see section 3.7.2		

Note: although models are perfectly defined as forward models, they need to be slightly modified when they are used for retrieval in order to insure a good convergence to valid retrieved values. Please read the section 3.5.1 that describes the modification needed to the formulation of the wet soil dielectric constant model (whatever Dobson or Mironov) used in MN.

3.2.3.4.2   Forward models for each decision tree branch

Table 20 illustrates the models selected for each FM fraction, as a function of the branches identified by their ranking order in the first column. This selection is provided by UPF (TGRD)

Table 20: Selected models

N°	Retrieval case	Selected models for fractions FM									
		FWP	**FWS**	**FSN**	**FRZ**	**FFO**	**FNO**	**FWL**	**FEB**	**FTI**	**FEU**
1	All open water	**MWP**	**MWS**	MD0	MD0	MN0	MN0	MWP0	MD0	MD0	MD0
2	Heterogeneous OW	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**
3	Strong topo pollution	MWP0	MWS0	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**
4	Soft topo pollution	MWP0	MWS0	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**
5	All wet snow	MWP0	MWS0	**MD**	MD0	MN0	MN0	MWP0	MD0	MD0	MD0
6	All mixed snow	MWP0	MWS0	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**
7	Wet snow pollution	MWP0	MWS0	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**
8	Mixed snow pollution	MWP0	MWS0	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**
9	All frost	MWP0	MWS0	MD0	**MD**	MN0	MN0	MWP0	MD0	MD0	MD0
10	Frost pollution	MWP0	MWS0	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**
11	Forest cover	MWP0	MWS0	MD0	MD0	**MN**	**MN**	MWP0	MD0	MD0	MD0
12	Soil cover	MWP0	MWS0	MD0	MD0	**MN**	**MN**	MWP0	MD0	MD0	MD0
13	All wetlands	MWP0	MWS0	MD0	MD0	MN0	MN0	**MWP**	MD0	MD0	MD0
14	All barren	MWP0	MWS0	MD0	MD0	MN0	MN0	MWP0	**MD**	MD0	MD0
15	All ice	MWP0	MWS0	MD0	MD0	MN0	MN0	MWP0	MD0	**MD**	MD0
16	All urban	MWP0	MWS0	MD0	MD0	MN0	MN0	MWP0	MD0	MD0	**MD**
17	Heterogeneous	MWP0	MWS0	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**	**MDd**

Notes:

- The forwards models for each fraction FM and decision tree branch are referred to as defined in Table 19. A "0" suffix have been added for **default** contributions.
  - The models used for retrieval are printed in bold and highlighted in yellow in each row except for forest cover (DT11) and Soil cover (DT12) in green to remind that they are the two categories using the nominal model (NM) for the dielectric constant.
  - The aggregate of fractions used for retrieval define an overall aggregated fraction **FRE**.
  - For frozen conditions, an option is to **override** the tests on lines 6 and 7 and to process with MN (see section 3.1.4.2)
- As long as the dedicated **SNOW cover** model **MS** is not defined for wet or mixed snow, it is replaced by the cardioid model. This model is applied to different fractions depending if the snow is characterized as "mixed" or "wet"

- The cardioid model **MD** is used for either some prevailing cover cases, or highly inhomogeneous situations. In the second case it is referred to as **MDd** in the table.

- Inasmuch as the cardioid retrieval is requested systematically, an **additional MD retrieval** must be performed for cases # 1, 11, 12 and 13, applied to the same aggregated fractions.

- In multiple model retrieval, either different models (e.g. case 1) or same models (e.g. case 2), the **free** parameters are the same for all concerned fractions i.e. in case 1, for example, if T_{SURF} is retrieved then T_{SURF} is the same for both FWP and FWS. **Fixed** parameters values, on the other hand, are always fractions model dependent.

- For branches 11 and 12, a retrieving option is introduced to retrieve SM on both the FNO and FFO fraction simultaneously, using the same MN model. For branch 11, every other FNO parameter is fixed; for branch 12, every other FFO parameter is fixed. This option of course only to be considered when a FNO (branch 11) or a FFO (branch 12) is present. The option is to be chosen when the reference TAU_nadir for FFO is **smaller** than a specified threshold **TH_TAU_FN**. When this option is triggered, then the flag FL_DUAL_RETR_FNO_FFO is raised.

3.2.3.4.3 Default versus retrieval models

It must be stressed that formally the direct models may be **slightly different** depending whether they are used for building default contributions or for retrieving floating parameters. This feature has been indicated above for the MD model. This will also have to be specified for the other models concerning surface temperature and some vegetation parameters; see Table 21.

Table 21: Default models vs. retrieval models

	Default	Retrieval
Dielectric constant formulation		
MD	Complex dielectric constant	(A,U) cardioid with U frozen
MN	Check case of sand	Check case of sand
MW		

3.2.3.5 Computing reference values for parameters

Reference values for space varying quantities entering the relevant forward models are obtained or computed over the DFFG relevant area for aggregated fractions FV that depend on the incidence angle. The idea is that with varying view angle the fractions might vary and, in some cases, with non-negligible impact (even with the weighing) due to presence of water for instance.

Fractions FV are very similar to FM or FM_0 fractions (from which the incidence angle dependency has been removed). However:

- Concerning **open water**, **distinct** reference values are requested for sub-models MWS and MWP applied to distinct sub-fractions FWS and FWP of FWO;

- For each DFFG node, the resampled **"supplementary"** covers **override** the complementary ones. However, the **topographic** supplementary covers, which were taken into account when computing the fractions FM_0 for driving the decision tree, are **ignored** when computing reference values.

Atmospheric and sky parameters are computed a **single time** for the DGG node;

Use is next made of WEF to compute average reference parameter values for each aggregated fraction FV.

This summarized description concerns the most complicated occurrences, featuring several fractions, possible NPE occurrences, non-uniform quantities. There are many practical cases where the full computation from DFFG can be avoided.

Table 22 summarizes the categories of necessary reference values to be used in relevant fractions and forward models.

Table 22: Categories of necessary reference values

Cover	Soil	For-est	Wet-lands	Open water	Bar-ren	Ice	Ur-ban	Frost	Snow	
Parameters										
Atmosphere, sky										ALL
Physical temperature	MN	MN	MW	MW	MD	MD	MD	MD	MD*	
Dielectric constant					MD	MD	MD	MD	MD*	
OS			MW	MW						
SM	MN	MN								
Others (dielectric layer)	MN	MN								
Roughness	MN	MN								
Vegetation layer	MN	MN	MW		MD	MD	MD	MD	MD	
MOUNTAIN										None

*   concerning snow, the table corresponds to present choices for wet and mixed snow
**   concerning snow, this is a placeholder waiting for a mixed snow model to become available.

A number of particular cases will probably request clarifications.

- When applying the MD model three possibilities exist (this does not cover the case of the additional retrieval MDa, see 3.2.5.1):

  - MD0: the cardioid model is dedicated to a non-nominal surface as a default contributor. The reference values for A_card and U_card is computed using prescribed dielectric constants for that type of surface provided in TGRD UPF.

  - **MD**: the cardioid model is dedicated to a non-nominal surface as a retrieval model. The prior value for A_card and the reference value for U_card is computed using the dielectric constant for that type of surface provided in UPF.

  - **MDd**: the cardioid model is dedicated to a group of non-nominal surface as a common retrieval model. The prior value for A_card and the reference value for U_card is taken directly from the default values provided in TGRD UPF.

- Optical thickness retrieval: this parameter does not depend on polarization. The initial (reference) value is obtained through summing the standing vegetation component and the average of the (polarization dependent) litter component.

3.2.3.6   Decision tree stage 2 for retrieval conditions

3.2.3.6.1   Minimum, full and maximum retrievals

When filtering out L1c pixels, we defined (section 3.2.2.1.7) an **initial validation index MVAL0**. Now it is possible to account for the overall fraction FRE selected for retrieval and to estimate a more realistic **validation index MVAL**:

$$MVAL = MVAL0 * FRE$$

In order to define retrieval conditions, one must compare MVAL to **thresholds**. Although 2 thresholds will be initially sufficient, it is wiser to define 3 of them: **TH_MMIN1**, **TH_MMIN2** and **TH_MMIN3.** Then 4 cases may occur:

	MVAL < TH_MMIN1	→ L1c pixel is finally **invalidated**
TH_MMIN1 ≤	MVAL < TH_MMIN2	→ "minimum" retrieval
TH_MMIN2 ≤	MVAL < TH_MMIN3	→ "full" retrieval
TH_MMIN3 ≤	MVAL	→ "maximum" retrieval

Initially selected values (provided in the parameter file) were initially TH_MMIN1 = TH_MMIN0 = 10, TH_MMIN2 = 35 and TH_MMIN3=60.

These values are in any case qualitatively related to the length of "dwell lines" along which TB data are available. They will be adjusted during the commissioning phase. Initial values are chosen in such a way that case #4 (maximum retrieval) is **not** met.

The following Table 23 is a draft for selecting retrieval options depending whether the MVAL value leads to cases 2, 3 and 4 above (last digit in the name of the retrieval model on 1st line of array).

Table 23: Selected free parameters for retrieval[16]

Parameter	Unit	TAU_R	MD2	MD3	MD4	MN2	MN3	MN4	MW2	MW3	MW4
Forward model			epsilon + TAU (cardioid)			SM + TAU (L-MEB)			OS + TAU		
A	B	C	D	E	F	G	H	I	J	K	L
D_TSURF	K	[0 TH_23]	nil	nil	2.5	nil	nil	2.5	2.5	2.5	2.5
		[TH_23 TH_34]	nil	nil	2.5	nil	nil	2.5	nil	2.5	2.5
		> TH_34	nil	nil	2.5	nil	nil	2.5	nil	2.5	2.5
D_A_CARD	-	[0 TH_23]	20.0	20.0	20.0						
		[TH_23 TH_34]	20.0	20.0	20.0						
		> TH_34	20.0	20.0	20.0						
D_SM	%	[0 TH_23]				20.0	20.0	20.0			
		[TH_23 TH_34]				20.0	20.0	20.0			
		> TH_34				10.0	10.0	10.0			
D_TAU	-	[0 TH_23]	nil	nil	nil	nil	nil	nil	nil	nil	nil
		[TH_23 TH_34]	0.1	0.5	0.5	0.1	0.5	0.5	0.1	0.5	0.5
		> TH_34	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
D_TTH	-	[0 TH_23]	nil	nil	nil	nil	nil	nil	nil	nil	nil
		[TH_23 TH_34]	nil	nil	1.0	nil	nil	1.0	nil	nil	1.0
		> TH_34	nil	nil	nil	nil	nil	nil	nil	nil	nil
D_RTT	-	[0 TH_23]	nil	nil	nil	nil	nil	nil	nil	nil	nil
		[TH_23 TH_34]	nil	nil	2.0	nil	nil	2.0	nil	nil	2.0
		> TH_34	nil	nil	nil	nil	nil	nil	nil	nil	nil
D_OMH	-	[0 TH_23]	nil	nil	nil	nil	nil	nil	nil	nil	nil
		[TH_23 TH_34]	nil	nil	0.1	nil	nil	0.1	nil	nil	0.1
		> TH_34	nil	nil	0.1	nil	nil	0.1	nil	nil	0.1
D_DIFF_OM	-	[0 TH_23]	nil	nil	nil	nil	nil	nil	nil	nil	nil
		[TH_23 TH_34]	nil	nil	0.1	nil	nil	0.1	nil	nil	0.1
		> TH_34	nil	nil	nil	nil	nil	nil	nil	nil	nil
D_HR	-	[0 TH_23]	nil	0.1	0.1	nil	0.1	0.1			
		[TH_23 TH_34]	nil	nil	nil	nil	nil	nil			
		> TH_34	nil	nil	nil	nil	nil	nil			

The names of parameters considered for retrieval (given in col. A) are self-explanatory: T_SURF stands for surface temperature and corresponds either to soil surface layer temperature or water, snow or ice temperature, TAU stands for τ_{NAD}. Soil moisture is expressed in percentage.

Retrieval choices are defined by the **a priori standard deviations ASTD**, which are shown in the table (col D to L). In many cases (ASTD = "nil"), the parameter is not retrieved. Large values mean the parameter is practically left free.

While the above table has been filled for **illustration**, numerical ASTD values are specified in TGRD UPF. The number of free parameters NP can be obtained from this data.

3.2.3.6.2   Condition on a priori optical thickness

For each parameter, 3 cases are defined, depending on the initial value **TAU_R** of the optical thickness for the retrieved fraction (col. C), as compared to thresholds **TH_23 & TH_34** supplied by the TGRD UPF.

---

[16] Highlight colors: green low opacity range, blue medium opacity range and yellow high opacity range

Values smaller than TH_23 mainly correspond to ice or barren soil, or open water excepting wetlands, i.e. cases where the optical thickness is known to be negligible. Values larger than TH_34 should mainly correspond to forests, where optical thickness is known to be high. The value of TAU_R then helps to select the retrieved parameters.

3.2.3.6.3 Using current maps for TAU and HR

Using the current maps depends on following conditions:

- The global switches controlling the use of the current maps grant it.

- The data exists in the LUT for the DGG node.

- The delay since updating the current map is smaller than limits TH_CUR_TAU_NAD_LV_VAL_PERIOD and TH_CUR_HR_VAL_PERIOD.

- There is no delay check for forests; if present, the data is always used.

Then, for the CURRENT map orbit pass matching the L1C under processing ASCENDING_FLAG:

- The CURRENT value of the parameter overrides the default value obtained from auxiliary data

- The CURRENT value of the DQX overrides the prior standard deviation obtained from TGRD UPF

3.2.4 Iterative solution

3.2.4.1 Formulation of the retrieval problem

Basically, the retrieval algorithm consists in minimizing a quadratic cost function. It is more complex than a standard maximum likelihood estimation, because:

- constraints (a priori estimates and a priori standard deviations) are introduced for some among the retrieved parameters.

- when dealing with **surface** brightness temperatures, strong correlations between the data must be accounted for, in such a way that the cost function becomes a **quadratic** form rather than a sum of weighted squared differences.

- In this ATBD version, since the retrieval is carried out at antenna level, the remark just above becomes no longer relevant.

3.2.4.2 Cost Function to be minimized

3.2.4.2.1 Simulated TB

Let $TBF(\theta, p_i\ldots)$ be the direct (forward) model for a brightness temperature at the antenna level; TBF depends on the incidence angle θ and on physical parameters p_i. We consider here only the NP parameters to be retrieved. For mixed pixels, TBF includes contributions simulated using default models.

The TBF are built according to equation **Eq 9**a for simulating TOA TBs for horizontal and vertical polarizations and then using **Eq 76** (including antenna polarization selection) for obtaining simulated TBs at the antenna level. The fractions FV_i are computed through applying the θ-dependent WEF function; aggregation rules are those (FM) for reference values (section 3.2.3.5) rather than those (FM_0) for driving the decision tree.

3.2.4.2.2 Cost function

Then, the retrieved p_i values are those which minimize the cost function COST:

$$COST = \left(TBM_m - TBF(\theta_m, p_i..)\right)^t \left[COV_T\right]^{-1} \left(TBM_m - TBF(\theta_m, p_i..)\right) + \sum_i \frac{[p_i - p_{i0}]^2}{\sigma_{i0}^2} \qquad \textbf{Eq 80}$$

where the TBM_m are **NT** measured values, the "t" superscript stands for transposition, $[COV_T]$ is the variance matrix for the observed TBM, and the p_{i0} are prior estimates of the **NP** free physical parameters, with prior variances σ_{i0}^2.

Note COST can also be written:

$$COST = \left(TBM_m - TBF(\theta_m, p_i.)\right)^t \left[COV_T\right]^{-1} \left(TBM_m - TBF(\theta_m, p_i.)\right) +$$
$$\left(p_i - p_{i0}\right)^t \left[COV_{Prior}\right]^{-1} \left(p_i - p_{i0}\right) \qquad \textbf{Eq 81}$$

where [COV_{Prior}] is a diagonal matrix with the terms σ_{i0}^2:

$$COV_{Prior} = \begin{bmatrix} \sigma_{10}^2 & 0 & L & 0 \\ 0 & \sigma_{20}^2 & 0 & 0 \\ M & 0 & O & 0 \\ 0 & L & 0 & \sigma_{NP0}^2 \end{bmatrix}$$

Eq 82

And finally:

$$COST = (DIFF)^t [COV_Z]^{-1} (DIFF)$$

Eq 83

where the square (NT+NP) ranked [COV_Z] matrix is built by aligning along the main diagonal the matrix [COV_T] and [COV_{Prior}]; the vector **(DIFF)** has a (NT+NP) length and consists of NT terms ($TBM_m - TBF(\theta_m, p_i)$) followed by NP terms equal to ($p_i - p_{i0}$).

The NT number is equal to the number M_AVA of validated L1c observed views. As indicated above, the TBF values are modelled assuming the 3rd and 4th Stokes parameters are zero at Earth's surface level.

3.2.4.3 Building matrixes for L1C pixels

The uncertainty on the observation system is defined from one part containing the antenna radiometric uncertainties, for a L1C pixel, which are provided as standard deviation for each view of this node and a second part accounting for other sources of errors coming from the image reconstruction. The first part is formed by the vector, [DTB_a] having M_AVA components and is provided by L1C files and since algorithms uses variances, the terms must be squared. The second part is provided as the extra variance σ_{IR}^2 which is configurable and provided in the TGRD UPF.

The total contribution is given by a diagonal matrix written as follows:

$$COV_T = \begin{bmatrix} (DTB_{a1})^2 & 0 & \Lambda & 0 \\ 0 & (DTB_{a2})^2 & 0 & M \\ M & 0 & O & 0 \\ 0 & \Lambda & 0 & (DTB_{aNT})^2 \end{bmatrix} + \sigma_{IR}^2 \times I_{NT}$$

Eq 84

Where subscript a refers to antenna and I_{NT} is the identity NTxNT matrix.

Introducing Bayesian a priori constraints on retrieved parameters results in writing the overall block diagonal matrix of uncertainties [COV_Z]:

$$COV_Z = \begin{bmatrix} [COV_T] & 0 \\ 0 & [COV_{Prior}] \end{bmatrix}$$

Eq 85

The overall matrix [COV_Z], holding brightness temperatures related uncertainties, DTB_{ak} and σ_{IR}^2 buidling COV_T and prior parameters values uncertainties, σ_{i0}^2, building COV_{Prior}, is introduced in the cost function written above. The vectors (DIFF) in the cost function must be written in the order consistent with the order chosen to build the COVz matrix elements.

During the retrieval process, use is made of the covariance matrix for the computation of retrieved parameters posterior [COV_{Post}] (see section 3.2.4.4). In [COV_{Post}], the (DIFF) vector is replaced (section 3.2.4.4.2) by a the Jacobian matrix of partial derivatives [**DRV_T**], next a [DRV_Z] matrix. This matrix must be written in the order consistent with the order chosen for the COV matrix.

3.2.4.4 Implementation and convergence criterions

3.2.4.4.1 Implementation

A possible way to implement the iterative retrieval consists in using the principle of the **Levenberg-Marquardt algorithm (L-M)** [125].

The Levenberg-Marquardt algorithm is simply a starting point to build the technical algorithm allowing minimizing the cost function. Its common published presentation (see also "numerical recipes" and common libraries of routines) doesn't include neither the out-of-diagonal terms of the quadratic form used in the cost function, nor the Bayesian constraints in the cost function given in equation Eq 85 above.

No particular difficulty was found in implementing these developments. MATLAB and FORTRAN routines have been written and tested by ESL to this purpose; the former one has been communicated to ARRAY. It is not known whether an extended L-M algorithm is described in the open literature. Anyway, using the L-M algorithm is by no means mandatory: the task is to apply a robust and efficient method in order to minimize the cost function.

3.2.4.4.2 Theoretical retrieval variances

If there were no constraints on the retrieved parameters, the square NP-ranked [COV$_{Post}$] covariance matrix for the retrieved parameters would be written:

$$[COV_{Post}] = \left[[DRV_T]^t [COV_T]^{-1} [DRV_T] \right]^1 \qquad \textbf{Eq 86}$$

where [DRV$_T$] is a (NT, NP) matrix, the terms of which are derivatives $\partial(TBM_m)/\partial p_i$.

Allowing now for a priori constraints on the parameters, the square NP-ranked [COV$_{Post}$] covariance matrix for the retrieved parameters can be written:

$$[COV_{Post}] = \left[[DRV_Z]^t [COV_Z]^{-1} [DRV_Z] \right]^1 \qquad \textbf{Eq 87}$$

where [COV$_Z$] is defined above in subsections 3.2.4.2 and 3.2.4.3; [DRV$_Z$] is a (NT+NP, NP) sized matrix which first consists of the [DRV$_T$] matrix, extended by a square NP-ranked unity matrix.

The vectors including theoretical retrieval a posteriori variances and standard deviations on parameters P are obtained from the diagonal of [COV$_{Post}$]:

$$(\textbf{RSTD}) = \text{sqrt} \left(\text{diag}([COV_{Post}]) \right) \qquad \textbf{Eq 88}$$

The RSTD for a particular retrieved parameter P is denoted either RSTD(P) or RSTD$_P$ e.g RSTD(SM), RSTD$_{SM}$.

3.2.4.4.3 Enhanced theoretical retrieval variances

The sections 3.2.4.2.2, 3.2.4.3 and 3.2.4.4.2 are a particular case of error propagation using the linear tangent model of our modelling at the retrieved solution and considering only the retrieved parameters.

Indeed the (RSTD) provide the post retrieval standard deviation for the retrieved parameters considering the model sensitivity given by the Jacobian matrix DRV$_Z$ (linear tangent model) limited to the retrieved set of parameters as the driver converting the radiometric uncertainty and prior uncertainty included in COV$_T$ to retrieved parameter space uncertainty at the current iteration of the L-M and is used as part of the convergence criteria.

We consider also an enhanced form (RSTD$_E$) computed after the retrieval of the optimal values p$^*_i$ and accounting for the uncertainties of any set of fixed parameters participating to the modelling but not retrieved. Let NFP be the number of these fixed parameters pf$_k$ having fixed values pf$_{k0}$ and prior uncertainty variance σf_k^2.

The Jacobian matrix $\left[DRV_T \right]$ is extended to form the matrix $\left[DRV_T^E \right]$ by adding NFP columns of direct model partial derivatives $\partial(TBM_m)/\partial pf_i$ evaluated at (θ_m, p$^*_i$, pf$_{k0}$, ...) for each parameters pf$_k$ and for the NT measured values. $\left[DRV_T^E \right]$ becomes a (NT,NP+NFP) matrix. Similarly to section 3.2.4.4.2, $\left[DRV_T^E \right]$ is extended by the square NP+NFP-ranked unity matrix to form the final $\left[DRV_Z^E \right]$ (NT+NP+NFP,NP+NFP) sized matrix.

The COV_Z^E uncertainty matrix becomes the block diagonal (NT+NP+NFP, NT+NP+NFP) matrix:

$$\mathrm{COV}_Z^E = \begin{bmatrix} [\mathrm{COV_T}] & 0 & 0 \\ 0 & [\mathrm{COV_{Prior}}] & 0 \\ 0 & 0 & [\mathrm{COV_{Fixed}}] \end{bmatrix} \text{ with } \mathrm{COV_{Fixed}} = \begin{bmatrix} \sigma_{f1}^2 & 0 & \Lambda & 0 \\ 0 & \sigma_{f2}^2 & 0 & 0 \\ M & 0 & O & 0 \\ 0 & \Lambda & 0 & \sigma_{fNFP}^2 \end{bmatrix}$$

Finally, the (RSTDE) is computed using Eq 85 and Eq 86, replacing DRVZ and COVZ by the extended version above. Only the first NP associated to the retrieved parameters is kept and defines (RSTD$_E$) which includes the prior uncertainty and sensitivity of the added fixed parameters.

Although this formulation can be applied to any parameter participating in the modeling (e.g the clay fraction) we will limit the set of fixed parameters of the potentially retrieved ones as defined in Table 23 and only for soil moisture based retrieved model (DT11 or DT12).

The fixed parameters pf_k and their associated uncertainty variance σf_k^2 are configurable through the UPF including the particular case where NFP is equal to zero (no enhancement). In this latter case, the (RSTD$_E$) becomes equal to the (RSTD) obtained at the end of the retrieval.

3.2.4.4.4 Data Quality Index

The RSTD defines the standard deviations of the uncertainty on retrieved parameters we should consider explaining the standard deviation of the noise on the brightness temperatures (DTBa). It is obtained intrinsically through the sensitivity of the model around the retrieved solution (the Jacobian of the modelling at the solution) as shown in the previous section.

The Data Quality Index, DQX, can be seen as an extension. It is meant to represent the overall uncertainty on retrieved values or on derived values. It can embed all the other sources of errors we are able to characterize on top of the RSTD that are worth to consider the retrieved values more uncertain than the RSTD itself which can be seen as a lower bound of the DQX.

For that reason, instead of using the (RSTD), the DQX is based on the (RSTD$_E$) that offers the possibility to propagate the uncertainty of some fixed parameters.

At this moment, a second addition concerns also the pixels that are regularly contaminated by RFI as given by $C_{RFI}(R_{RFI})$ (see section 3.1.6.1) for which we want to convey the user the low confidence we have in the retrieved values by enlarging their RSTD$_E$ by a factor of C_{RFI}:

$$\textbf{(DQX)} = (RSTD_E) * C_{RFI} \qquad\qquad\qquad \textbf{Eq 88b}$$

When no RFI are detected in the short past history then $DQX = RSTD_E$ ($R_{RFI}=0 \Rightarrow C_{RFI}=1$). When 100% of RFI were detected then $DQX = 4 \times RSTD_E$ ($P_{RFI}=1 \Rightarrow C_{RFI}=4$), although in this latter extreme probably no retrieval could have been attempted.

In future release of the algorithms other error aspects to consider are foreseen:

- Intrinsic model error: it is the error model structure due to its formulation and/or the associated simplifications with respect to the unknown true process. It is always tough to obtain, but potentially assessable statistically. Perhaps a better place would be in the cost function with the two others, the background error and the observation errors.

- Extrinsic model error: it comes from the uncertainty of all the fixed parameters that propagates through the modelling. Theoretically, if known (assessed) they can be accounted using a similar approach to RSTD computation. The use of RSTD$_E$ is a first attempt toward this direction, helping potentially to deal with the issue raised at the following point. As per this ATBD revision, the RSTD$_E$ computation is an implemented experimental feature for extended testing and optimisation not yet meant to be used for operational processing. The appropriate set of fixed parameters to consider as well as the magnitude of their associated uncertainty to be propagated have to be carefully chosen to obtain the best uncertainty on retrieved parameters consistently with others algorithm aspects such as the current files which make use of past, potentially enhanced, DQX.

- Lower error bound: the modelling sensitivity can become mathematically very high, for instance when SM retrieved values are very close to 0. As result the RSTD can become ridiculously small with no physical meaning.

3.2.4.4.5 Control parameters and convergence criteria

The following operating parameters must then be specified (see Table 24):

Table 24 : Parameters for iterative retrieval

Parameter	Meaning	Suggested value
NITM	Maximum number of iterations	10
KDIA	Initial value of the diagonal increment	0.1
KDIA_MAX	Maximum value of the diagonal increment	1000
FDIA	Multiplying factor for KDIA between successful iterations	0.1
FCV1	Convergence test on increment of retrieved parameters	0.005
And/or FCV2	Convergence test on decrease of the cost function	
FCOND	Test for matrix conditioning	1e10
(DPD)	Parameter increments for computing derivatives (vector)	See User's data file

In this table:

- KDIA and FDIA are parameters characteristic of the L-M descent method.

- NITM is self-explanatory. The current iteration number is to be incremented **systematically**, including steps which are simply repeated with a larger Marquardt diagonal increment.

- FCVI and FCV2 are convergence tests. It is considered that a **single** test (FCV1) will be adequate. The simplest convergence test is to consider the increments on retrieved quantities from one iteration to the next. Then verify that they are much smaller (by a factor FCV1) than the estimated retrieval standard deviations **RSTD**, computed for the current iteration. If however this test is performed for high values of the current value of the Marquardt diagonal increment, it becomes meaningless. Therefore it must be checked that the test is performed between "**true**" iterative steps (where free parameters are updated), as opposite to steps which are simply repeated with a larger Marquardt diagonal increment), and for a Marquardt increment which does not exceed a maximum value KDIA_MAX.

The KDIA_MAX parameter should also be used as a failure criterion.

- FCOND should be a feature of any matrix inversion sub-algorithm. In MATLAB the argument is the ratio of the largest to smallest singular values. No way to set the numerical value has been found other than trial and error. Value currently used in the breadboard MATLAB routine is 10^{10}.

- In the breadboard version, derivatives are computed numerically. Then, it is necessary to supply a **table of parameter increments** on which these derivatives are computed (line "**DPD**" of table). DPD values are not critical; an array of values DP_SM, DP_A_card, DP_TAU_nad, DP_T_SURF, DP_TTH, DP_RTT, DP_OMH, DP_DIFF_OM will be suggested in TGRD UPF for each floating parameter.

Failure of convergence is detected by:

- NITM reached before the convergence criterion is met.

- KDIA_MAX exceeded.

- Poorly conditioned matrix.

These failures give rise to **L2 generated flags**.

3.2.4.5 Polarization modes

The polarization options have been mentioned whenever relevant in various 3.2.4 subsections. Table 25 summarizes differences in the processing.

Table 25: comparison between polarization modes

Items	Dual polarization	Full polarization
Number of data M	M_AVA	M_AVA
Direct model call	(TH(θ), TV(θ)) x M_AVA	(TH(θ), TV(θ), 0, 0) x M_AVA
Single view transformation matrix	[MR2]	[MR4]
Size of data covariance matrix [COV$_T$]	(M_AVA , M_AVA)	(M_AVA , M_AVA)

3.2.5 Post processing

3.2.5.1 Post retrieval analysis; repeated retrievals and retrieval flags

The goal is to analyze the results after the parameter retrieval has been attempted. It consists in computing posterior error variance on the retrieved parameters, and in checking consistency with other products or expected ranges. This will help to assess whether the retrieval needs to be redone. And in the case the retrieval is acceptable whether it is so with caveats. Depending on the situation, specific flags are raised and outputs to UDP are built following specific rules.

The following table describes the sequence proposed following the initial retrieval attempt according to conditions stipulated in the decision tree 3.2.3.6 subsection.

In this table, the theoretical retrieval standard deviations (**RSTD**) come into use in column H while the quality index (**DQX**) is used in column C to check the maximum accepted uncertainty for retrieved parameters.

Some among the retrieval tests drive **repeated** retrieval attempts; others simply give rise to retrieval **flags**.

Table 26: Retrieval analysis conditions, options and actions[17]

	Converged?	**DQX within range?**	**Acceptable retrieved values ?**	**Out-liers?**	**Action**	**Remaining outliers?**	**Goodness of fit?**
(a)	see § 3.2.4.4	(b)	(e)	(c)		(c')	(d)
	B	C	H	D	E	F	G
R4	NO				Try R3		
R4	YES	NO			Try R3		
R4	YES	YES	NO		Try R3		
R4	YES	YES	YES	YES	Filter, try again R4, build output (f)	If YES, flag	if NO, flag
R4	YES	YES	YES	NO	Build output		if NO, flag
R3	NO				Try R2		
R3	YES	NO			Try R2		
R3	YES	YES	NO		Try R2		
R3	YES	YES	YES	YES	Filter, try again R3, build output (f)	If YES, flag	if NO, flag
R3	YES	YES	YES	NO	Build output		if NO, flag
R2	NO				Build output (g)		
R2	YES	NO			Build output (g)		
R2	YES	YES	NO		Build output (g)		
R2	YES	YES	YES	YES	Filter, try again R2, build output (f)	If YES, flag	if NO, flag
R2	YES	YES	YES	NO	Build output		if NO, flag
R2	YES	YES	NO	Build output		if NO, flag	if NO, flag

(a)   Retrieval status: R2, 3 or 4 means the retrieval under analysis corresponds to the option 2, 3 or 4 in Table 23

(b)   **DQX** are acceptable if all smaller than thresholds TH_DQX_**. (TH_DQX_SM; same for A_card, T_SURF,

---

[17] Highlighted in yellow failed case requiring a second retrieval, in green full nominal case.

HR, TT_H, RTT, OMH, DIFF_OM) are supplied in TGRD UPF. Note that this test will be active only for values smaller than the a priori ASTD also supplied by TGRD UPF.

(c)   Outliers are detected through comparing abs(TBM - TBF)/DTB to TH_FIT. TH_FIT is a provided constant; suggested value is 4.

If abs(TBM - TBF)/DTB > TH_FIT, delete data, count and report the number of removed outliers in N_CLEANED. **Note**: outliers are removed **altogether**. After the 2nd retrieval, the outliers are also checked to update flags but no more retrieval is tried. RFI counters are updated from N_CLEANED.

(c')   If wild data are still present, flag, count and report the number of remaining outliers in N_WILD.

Note: RFI counters are not updated from N_WILD but that option may be considered in the future.

(d)   Fit is good if CHI2 within given range. Chi2 formula and range (upper threshold TH_CVAL) to be provided.

(e)   Acceptable range for free parameter, P, if it belongs to its extended validity range, $[P_{min} – C\_R_P*RSTD_P, P_{max} + C\_R_P*RSTD_P]$, with P being any of: SM, A_card, T_SURF, HR, TT_H, RTT, OMH, DIFF_OM. The P_{min}, P_{max}, C_R_P values are supplied in TGRD UPF.

(f)   It is assumed that removing outliers will not impact columns B & C.

(g)   No product, FL_NOPROD is raised.

The following output **descriptors or flags** may in summary characterize the retrieval when successful: outside stipulated ranges, (FL_RANGE), outside acceptable retrieval DQX (FL_DQX), poor fit quality (see below)-.

Concerning the **systematic MD retrieval**, when it is carried out in **addition (MDa)** to any other (MN or MW) retrieval, the following rules are proposed:

If the initial (MN or MW) retrieval fails, no MDa retrieval is attempted;

The additional MD retrieval **MDa** is only attempted **once**, i.e. for the highest level (4, 3, 2) for which the former retrieval (with MN or MW) was successful;

If then this additional retrieval fails, no further attempt with MDa for poorer retrievals is carried out.

This seems to be the only case where recording a retrieval failure in the product cannot be avoided.

3.2.5.2   Updating current parameter maps

Even though not part of the ATBD, it must be clear that this section applies to a post-processor rather than to the processor itself. Still, it needs to be documented.

However, it is the responsibility of the processor to inform the post-processor that an action is to be taken for given nodes by raising the appropriate flags FL_CURRENT_***.

The RFI, Tau, HR and Flood maps will be updated offline after the data processing and in view of being used as inputs for the next processing round. In consequence, the processor outputs products. Then another processor (a post-processor) reads these products, checks the quality and eventually updates the maps according to a procedure described below. The updated auxiliary data will be used in input at the next run of the processor.

Current parameters maps values are updated separately per ascending and descending orbits and they are used in a similar way as inputs of the algorithms. A processed orbit being ascending (resp. descending), the ascending (resp. descending) current parameters maps are used in inputs of the processor and the resulting processor products will serve to update the ascending (resp. descending) current parameters maps by the post-processor.

In the following four subsections **this distinction is implicit**.

3.2.5.2.1   Updating current RFI map

RFI map designates here a DGG map named DGG_CURRENT_RFI. No angular dependence is considered in the baseline. Provisions to take angular information into account might prove useful. The map includes both the total number of times views have observed the DGG node: N_{SNAP}, and the total number of cases (N_{RFI-p}) where an anomalously high TB has been unambiguously detected for polarization p (X and Y), as the results of any of 3 tests: TB range testing, L2 RFI test (see section 3.2.2.1.5) and post retrieval outliers detection.

After one of several orbits have been processed by the level 2 processor, typically one day of processed orbits, the current RFI map is updated by the post-processor based on the UDPs content from these processed orbits. The newly created current RFI map will be used as input to the algorithms to process the future orbits. A series of DGG_CURRENT_RFI will be thus generated. At a given date a DGG_CURRENT_RFI map contains all the accumulated past story of RFI events detected by the algorithms since the beginning of the SMOS mission.

The update process is as follow: for each DGG node of the DGG_CURRENT_RFI map, the counter N_{SNAP} is incremented by the UDP' M_AVA0 values while the N_{RFI-X}, N_{RFI-Y} counters are incremented from the UDP' N_RFI_X, N_RFI_Y when RFI are detected for the X and/or Y polarized views among the M_AVA0 checked.

The **polarization** information may be more relevant at the Earth surface as RFI signal may be polarized and thus affect only one polarization while, at antenna level, RFI may be distributed on both X and Y components. However, though the information is used as if they were one ($N_{RFI}=N_{RFI-X}+N_{RFI-Y}$), the statistics are established for each through individual tracking.

As explained above, the DGG_CURRENT_RFI maps are also an input to the algorithms and provide the RFI local (in space) states of any DGGs. Actually, we use the integral characteristics of the DGG_CURRENT_RFI maps which contain the accumulated sums of the number of detected RFI and the number of observations, by differentiating two DGG_CURRENT_RFI counters ΔN_{RFI}, ΔN_{SNAP}. These counters are obtained from the latest updated DGG_CURRENT_RFI and the DGG_CURRENT_RFI from N days before, which also provides more localized-in-time information. Indeed, RFI is dynamic in time; new sources can appear, temporarily or not, but also known sources can disappear when for instance authorities succeed in switching them off.

Over any node, if the DGG_CURRENT_RFI map indicates potential RFI while the pre-processing tests are negative, the retrieved parameters uncertainties are degraded by a factor depending upon the ratio $R_{RFI} = \Delta N_{RFI} / \Delta N_{SNAP}$ (see Eq 70 in 3.1.6.1). Some care must be taken to verify that ΔN_{SNAP} is not equal to 0 even though with N > 4 days (12 days is targeted) it should not occur since all DGG would be revisited at least once. The RFI time window size defined by the N days is configurable through the AUX_DGGRFI_Window_Size UPF parameter.

This algorithm is activated or deactivated by the global switch UPDATE_RFI defined in TGRD UPF**.** The use of RFI map to flag the input TB is activated or deactivated by the global switch USE_CURRENT_RFI defined in TGRD UPF.

The map would thus contain for each node N_{SNAP}, N_{RFIp} for each polarization requiring as N_{SNAP} will grow quickly 4 bytes.

3.2.5.2.2   Updating current optical thickness TAU_NADIR LV and FO map

This external post-processing action still requires confirmation and detailed explanations.

The basic concept is when the retrieval of Tau_Nadir parameters is possible, successful and made on good retrieval conditions then the retrieved opacities update (replace) the evolving map or helps improving the current values used for Tau Nadir.

The current TAU_NADIR maps will then be used in next swath processes rather than the Tau nadir parameter derived from the surface auxiliary data (LAI and LAI max).

The idea is that the retrieved opacities TAU nadir is thought to be more accurate and/or more up to date than the LAI derived ones and must remain valid for at least 3 days to 6 days. 3 days is maximum revisit time of a DGG at the equator when both ascending and descending orbits are considered simultaneously. When orbits are separated in and ascending batch and descending batch then the revisit time at the equator becomes 6 days.

We kept the baseline 3-6 days as a reasonable maximum limit validity (but configurable) for past retrieved opacities as the vegetation is often evolving slowly even at the onset or senescence phase. However, with the added control on Tau retrieval quality through a mid-swath restriction (see below) when updating the current maps, this time validity limit must be relaxed toward longer period. To be useful, the current maps update strategy must grant at least one revisit for each DGG or for a maximum of DGGs before the end of the prescribed maximum validity period. But in any case, an opacity in this evolving map is older than the prescribed period TH_CUR_TAU_NAD_LV_VAL_PERIOD (in days) with respect to a processed orbit then the processor reverts to the LAI proxy.

It is also considered that for forests we will only update once for fine tuning as the relevant variable is more linked to the branch volume than leaves and thus should remain fairly constant. Any variation should be linked to changes in other variables (interception, litter water content…). However, to stay consistent with the low vegetation and to avoid possible dead locks a similar time strategy is considered but with a longer period (configurable) of validity of one month.

We need some care as the retrieved opacities may be affected by spurious events such as heavy rainfalls, RFI or obtained under not favorable conditions for instance at the swath edges with limited number of BTs and limited range of incidence angles

compared to the existing ones in the evolving maps. This requires some quality controls before 1) asking a replacement with new values and 2) replacing a potentially better and valid existing ones.

No update is allowed in the following cases (force to unset FL_CURRENT_TAU_NAD_LV/FO)

- The retrieved opacities are successful but made on marginal or non-optimal information content. Vegetation acts as a depolarization compared to bare soil with a strong angular dependency. Such an effect can be only well captured by SMOS when the range of incidence angles, $\Delta\theta$, used for the retrieval is large enough but also that the observations are kept and not discarded e.g. due to RFI. Several criteria can be of use and have to be verified which is (are) the best for this purpose:

  $\Rightarrow$ constraint on |X_Swath| > threshold

  $\Rightarrow$ constraint on MAVA < threshold

  $\Rightarrow$ **constraint on MVAL0 or MVAL < threshold**

  $\Rightarrow$ constraint on $\Delta\theta$ < threshold

  For the time being the constraint on MVAL0 has been chosen as it includes several useful information in one value: content of information through radiometric accuracy, linked to the swath size and thus to the incidence angle range.

- Risk of interception.

- Presence of snow cover, or freezing conditions (vegetation is almost transparent and tau nadir is abnormally low).

In case an update is granted then no replacement of existing valid values is allowed in the following case:

- The Chi2 of the existing values are better than those for the new ones. Ideally, the probability of Chi2, Chi2_P, should be used but as per today the Chi2 is off the expectation the Chi2_P cannot be used but will be considered in future versions.

In case an update is granted, and existing valid values exist but **STree1** has changed, denoting a strong surface modification, then the update is forced.

If no TAU nadir is retrieved, or update not considered as valid, the date of the last acquisition is checked. If the lag is larger than the allowed TH_CUR_TAU_NAD_LV_VAL_PERIOD, a default or missing value is put.

The new TAU_NADIR is for a whole node but actually has a meaning only for the class for which the retrieval process was carried out (e.g., low vegetation or forest).

Therefore, when the retrieval of optical thickness is possible, successful and valuable enough, the TAU map class to be updated is the fraction for which TAU_nadir is retrieved, as provided by UPF (TGRD).

The CURRENT_TAU_NADIR_LV is only updated following LV class retrievals and if FL_CURRENT_TAU_NAD_LV is raised:

- Current TAU_NAD is updated with the retrieved TAU_NAD value.

  Current DQX TAU_NAD is updated with a **specific** computation of DQX TAU_NAD which is obtained as indicated in sections 3.2.4.4.2, 3.2.4.4.4 but with a matrix COV$_z$_CURRENT computed assuming that TAU_NAD is completely free (that is, practically, a very high ASTD). Even using this formulation, the DQX TAU_NAD can be very small and lock the future retrievals of Tau to the candidate TAU_NAD. As safeguard measure, a minimum floor cut-off value is considered.

  The UDP DQX_TAU_CUR is set to max(TH_Curr_Min_DQXTLV, DQX TAU_NAD).

- Current Chi2 is updated with the retrieved Chi2.

- Current STree1 is updated with the retrieval STree1

- Current date is updated with swath acquisition date.

The CURRENT_TAU_NADIR_FO is only updated following FO class retrievals and if FL_CURRENT_TAU_NAD_FO is raised:

- Current TAU_NAD is updated with the retrieved value.

- Current DQX TAU_NAD is updated with a **specific** computation of DQX TAU_NAD which is obtained as indicated sections 3.2.4.4.2, 3.2.4.4.4 but with a matrix COV$_z$_CURRENT computed assuming that TAU_NAD is completely free (that is, practically, a very high ASTD). Even using this formulation, the DQX TAU_NAD can be very small and lock the future retrievals of Tau to the candidate TAU_NAD. As safeguard measure, a minimum floor cut-off value is considered.

  The UDP DQX_TAU_CUR is set to max(TH_Curr_Min_DQXTFO, DQX TAU_NAD).

- Current Chi2 is updated with the retrieved Chi2.

- Current STree1 is updated with the retrieval STree1

- Current date is updated with swath acquisition date.

The update of the current maps is activated or deactivated by the global switches UPDATE_TAU_NADIR_LV and UPDATE_TAU_NADIR_FO defined in TGRD UPF**.**

The use of DGG_CURRENT_TAU_NADIR_LV maps and DGG_CURRENT_TAU_NADIR_FO maps in the retrieval is activated or deactivated by the global switches USE_CURRENT_TAU_NAD_LV and USE_CURRENT_TAU_NAD_FO defined in TGRD UPF.

The minimum floor values TH_Curr_Min_DQXTLV and TH_Curr_Min_DQXTFO are configurable and defined in TGRD UPF.

3.2.5.2.3   Updating current roughness HR map

The basic concept is when the retrieval of HR parameter is possible, successful and made on good retrieval conditions then the retrieved HR updates the evolving HR map or helps improving the current value used for HR.

The current HR map will then be used in next swath processes rather than the HR parameter derived from the surface auxiliary data.

The idea is that the retrieved HR are thought to be more accurate and/or more up to date than the HR derived ones and must remain valid for at least 3 days (DGG in swath maximum revisit time at the equator). The soil roughness is assumed to evolve very slowly with a time frame larger than 3 days. However, similarly to the forest opacity a validity time (configurable) of one month is considered to avoid dead locks.

No update is allowed in the following cases (force to unset FL_CURRENT_HR)

- The retrieved HR is successful but made on marginal or non-optimal information content. Roughness effects have a strong angular signature. Such an effect can be only well captured by SMOS when the range of incidence angles, $\Delta\theta$, used for the retrieval is large enough but also that the observations are kept and not discarded e.g. due to RFI. Several criteria can be of use and have to be verified which is (are) the best for this purpose:

  $\Rightarrow$ constraint on |X_Swath| > threshold

  $\Rightarrow$ constraint on MAVA < threshold

  $\Rightarrow$ **constraint on MVAL0 < threshold**

  $\Rightarrow$ constraint on $\Delta\theta$ < threshold

  For the time being the constraint on MVAL0 has been chosen as it includes several useful information in one value: content of information through radiometric accuracy, linked to the swath size and thus to the incidence angle range.

In case an update is granted then no replacement of existing valid values is allowed in the following cases:

- The Chi2 of the existing values are better than those for the new ones. Ideally, the probability of Chi2, Chi2_P, should be used but as per today the Chi2 is off the expectation the Chi2_P cannot be used but considered later.

In case an update is granted, and existing valid values exist but **STree1** has changed, denoting a strong surface modification, then the update is forced.

The CURRENT_HR is only updated/used following the same class than the associated retrievals:

- Current HR is updated with the retrieved value.
- Current DQX HR is updated with a **specific** computation of DQX HR which is obtained as indicated sections 3.2.4.4.2, 3.2.4.4.4 but with a matrix COV$_Z$_CURRENT computed assuming that HR is completely free (that is, practically, a very high ASTD). Even using this formulation, the DQX HR can be very small and lock the future retrievals of HR to the candidate HR. As safeguard measure, a minimum floor cut-off value is considered.

  The UDP DQX_HR_CUR is set to max(TH_Curr_Min_DQXROU, DQX HR). TH_Curr_Min_DQXROU is configurable and defined in TGRD UPF.

- Current Chi2 is updated with the retrieved Chi2.
- Current STree1 is updated with the retrieval STree1.
- Current date is updated with swath acquisition date.

This algorithm is activated or deactivated by the global switch UPDATE_HR defined in UPF (TGRD section 3.3.7)**.** The use of HR map in the retrieval process is activated or deactivated by the global switch USE_CURRENT_HR defined in UPF (TGRD section 3.3.7**)**

3.2.5.2.4 Updating current flood map

This must wait future developments (see section 3.7).

In the future, a CURRENT_FLOOD_MAP will contain some accumulated SMOS retrieved information associated with external risks information to provide a more accurate flag on flooding risk.

It represents a possible implementation, compliant with the operational time constraint, of the paradigms presented in sections 3.1.3.3, 3.2.3.2.4.

3.2.5.3 Computing a surface brightness temperature field

It was understood that the L2 processor was to deliver brightness temperature fields at a given angle, θ_B (that of the browse product *a priori*).

Two types of brightness temperatures will be provided. One is defined above the surface level (ASL) in the Earth reference frame for polarization (H/V), and the other is defined at the antenna level (TOA) in the satellite reference frame for polarization (X/Y).

In both cases the recomputed and/or corrected surface temperature are simulated with the direct model, $TB_p(.)$ using the retrieved and the fixed parameters/variables and evaluated at the incidence angle θ_B :

- In the first case the modelled TB values are meant to be directly comparable with those available in L1C browse products and thus it requires propagating the current modelled $TB_H^{TOA}(\theta_B)$, $TB_V^{TOA}(\theta_B)$ to the antenna level $TB_{XX}^{TOA}(\theta_B)$, $TB_{YY}^{TOA}(\theta_B)$. Being also included in L1C browse product observation, the sky and atmosphere contribution should be included in the modelling.

  To operate this transformation, it is first necessary to obtain a proper c angle at θ_B for computing MR2(a) or MR4(a) matrix. This can only be performed cleanly through an interpolation of the two observed geometries. For this purpose, two adjacent views having their incidence angles, θ_1 and θ_2, bracketing θ_B (i.e $\theta_1 \leq \theta_B \leq \theta_2$) are to be found so as we can compute their associated a_1 and a_2. Then, two options can be considered:

  - We can compute two pairs of antenna simulated TBs,:

    - $\left(TB_{XX}^{TOA}(\theta_1, a_1), TB_{YY}^{TOA}(\theta_1, a_1)\right)^T = MR2(a_1) \cdot \left(TB_H^{TOA}(\theta_1), TB_H^{TOA}(\theta_1)\right)^T$ and

    - $\left(TB_{XX}^{TOA}(\theta_2, a_2), TB_{YY}^{TOA}(\theta_2, a_2)\right)^T = MR2(a_2) \cdot \left(TB_H^{TOA}(\theta_2), TB_H^{TOA}(\theta_2)\right)^T$

      and do a linear interpolation in the brightness temperature domain at the incidence angle θ_B.

  - Or we can do the interpolation in the geometric angle domain to obtain a_θ and then compute the simulated antenna TBs,

    $$\left(TB_{XX}^{TOA}(\theta_B, a_\theta), TB_{YY}^{TOA}(\theta_B, a_\theta)\right)^T = MR2(a_\theta) \cdot \left(TB_H^{TOA}(\theta_B), TB_H^{TOA}(\theta_B)\right)^T$$

  The first option is the preferred one since TBs interpolation is smoother than angular interpolation though it costs a little more (two calls to direct models, and two interpolations).

  The DQX computation must follow the same choice, either evaluated at (θ_1, a_1) and (θ_2, a_2) and linearly interpolated at θ_B or evaluated at $(\theta_B, a_{\theta B})$.

In the second case the modelled TB values are meant to be comparable with those coming from a tower radiometer or a low-level airborne one. Thus, the polarization shall be in the Earth reference frame H and V. These TBs may be very useful during the commissioning phase to be compared to fields campaigns experiment. The simulated brightness temperature above the surface level is given by $TB_{p\,\theta_B}^{ASL} = TB_p\left(\theta_B\,;\,P^*\,;\,TB_{SKY}, TB_{atm}, L\right)$, for horizontal and vertical polarizations. It includes the downward fluxes and attenuation (Sky, atmosphere and vegetation) but, in a strict sense it should not include the direct upward atmosphere radiation and attenuation. But this latter effect has a very small contribution over land and can be neglected leading to direct reuse of the general formulation of **Eq 9**.

In any case, we consider producing **two** pairs of TBs per L1c node one for ASL and the other for TOA. The TBs are computed for the same incidence angle as the L1c browse products: $\theta_B = 42.5°$. Note that this incidence angle value is nevertheless configurable.

The Data Quality IndeX of those simulated brightness temperature is somewhat complex as the resulting brightness temperature is then computed with a complex approximate and non-linear model. The DQX should then express the impact of radiometric uncertainties, the uncertainties of the fixed parameters, the model error, the mis-registration (aux and TB), the aggregation induced errors etc.

So, in a first approximation, we could propagate the DQX of each retrieved parameters using a linear tangent approximation of the forward model for H/V polarization to obtain the DQX_TBs for ASL, and then just multiply by MR2(**a**) (MR4(**a**) in full polarization mode) to obtain the DQX_TBs for antenna TOA.

$$DQX\_TB_{p\,\theta_B}^{ASL} = \sqrt{\sum_{Par \in freepar}\left(DQX_{Par}\cdot\left.\frac{\partial TB_{par}()}{\partial Par}\right|_{\left(\theta_B;Par^*;L\right)}\right)^2} \quad \text{p being H or V} \qquad \textbf{Eq 89}$$

$$\left(DQX\_TB_{XX_{\theta_B}}^{TOA}, DQX\_TB_{YY_{\theta_B}}^{TOA}\right)^T = MR2\left(\boldsymbol{a}_{\theta_B}\right)\cdot\left(DQX\_TB_{H_{\theta_B}}^{ASL}, DQX\_TB_{V_{\theta_B}}^{ASL}\right)^T \qquad \textbf{Eq 90}$$

The vector where the partial derivatives is evaluated is set accordingly with the previous definitions for the TB models.

3.2.5.4 Further post processing operations

This is a placeholder for computing some quantities in case it is decided to include them in output files. Possible examples include:

- Estimated components of the optical depth: standing vegetation, litter
- Cardioid components of the dielectric constant

3.2.5.5 Computing elements for User Data Product

The UDP essentially reports the results of retrieval.

In addition, 3 specific parts are defined:

- PCD (**Product Confidence Descriptor**) includes indications about the global quality of the product. It contains both confidence value and flags.
- PPD (**Product Processing Descriptor**) includes indications about data product interpretation and process status.
- PSF (**Product Science Flags**) includes information about geophysical features

Every information in UDP refers to L1c pixels for which at least one retrieval attempt has been successful. Failures are addressed in the DAP.

3.2.5.5.1 Preparing UDP retrieval information

Retrieved parameters fields

Depending on the processing mode (i.e., ESL mode or standard user mode (default)), the retrieved values will either be checked and forced or not before building the outputs[18].

In standard user mode (operational mode):

---

[18] Only the ESL mode is implemented in current DPM; standard user mode needs to be done.

In addition to the validity range [P_{min} P_{max}] for each parameter P, we define an extended validity range [P_{min} − C_R_P*RSTD$_P$, P_{max} + C_R_P*RSTD$_P$]; the C_R_P coefficients are stored in TGRD UPF.

If the value of any retrieved parameter, P, is outside its extended validity range then the retrieval is considered to have failed and all retrieved values and their DQX are set to missing value (-999) in UDP.

FL_NOPROD is raised

FL_RANGE is raised to explain the reason of FL_NOPROD

If any value is between the validity and the extended validity range, the values stored in UDP for each retrieved parameter, P, are forced to min(max(P,P_{min}),P_{max}). This rule ensures that the parameters values stored in UDP belong to their TGRD UPF validity range even if we accepted a slightly extended range. (e.g. we consider the SM retrieval successful even if SM belongs to [-RSTD$_{SM}$ 0] or [0.5 0.5+RSTD$_{SM}$] but the value reported in UDP is 0 for the former case or 0.5 for the latter case).

In ESL mode (analysis mode) :

Flags FL_NOPROD and FL_RANGE are raised same as in standard user mode, but values are reported as is in UDP with no control or limitation.

- **Fixed parameters fields**
  - Surface BT field: see section 3.2.5.3 above
  - For dielectric constant derived from retrieval models **other** than MD: $(\varepsilon'_D, \varepsilon''_D)$ values are to be computed from retrieved physical model parameters (see section 3.1.2.2). The relevant physical DQX parameter values are to be **propagated** to ε_D. The same approach as in Eq 89 is applied but to the function that links physical parameters to ε_D.
  - For dielectric constant derived from **MD retrieval**: $(\varepsilon', \varepsilon'')$ values are to be computed from the retrieved D parameter A_card and the fixed U_card (see section 3.1.4.7). In addition, the A_card DQX has to be **propagated** to ε.

3.2.5.5.2 Preparing PCD elements for UDP

3.2.5.5.2.1 Separate elements

PCD for UDP only refers to pixels for which at least one retrieval attempt has been successful.

- Goodness of fit
  - **CHI_2 value**:

    it is the COST function reached at the end of the retrieval, **normalized** by the number NFD of degrees of freedom: NFD = NT – NP.

    At the present time the Chi2 values are on average too high, far from the theoretical expectation even selecting carefully unbiased retrievals. The observation uncertainty considered in the Chi2 is limited to the radiometric accuracy which is a good estimate of the noise magnitude at the BT level but is unable to capture BT profile anomalies resulting from image reconstruction scenes dependent bias that add extra oscillations along the BT profiles. The overall observations uncertainty is underestimated and generates too high Chi2 values that are unusable to compute meaningful Chi2 probabilities. To cope with this issue and represents this extra sources of errors the σ_{IR}^2 variance is introduced as and additive term at the cost function level in **Error! Reference source not found.**. As a complement, the final COST is corrected by: RCOST=COST*Chi_2_Rescale_factor+Chi_2_Rescale_offset to capture any remaining anomalies.

    Like σ_{IR}^2, Chi_2_Rescale_factor and Chi_2_Rescale_offset is configurable and defined in TGRD UPF.

    The CHI_2 value is then computed and reported in the UDP as RCOST/NFD.

    It is a temporary measure pending a proper accounting of image reconstruction scene dependent bias in the cost function formulation.
  - **CHI_2_P**, main goodness of fit indicator; is the χ^2 high-end acceptability probability. This is the probability that **no anomaly** occurred about the fit. This figure is given by:

CHI_2_P = GAMM$_Q$(NFD/2, RCOST/2)

Where GAMM$_Q$ is the upper tail of the incomplete gamma function (see [126], chapters 6, 7 & 26).

With NFD in the range of several tens, CHI_2_P should go down to a small number of percent when CHI_2 exceeds a figure around 1.3 to 1.4. Note that very high values of CHI_2_P (when CHI_2 is "too small") are suspicious also, as they raise the possibility of correlated noise which would be unduly fitted by the direct model.

The table below illustrates the values of CHI2_P against normalized CHI2 values depending on the number NFD of degrees of freedom.

Table 27: CHI2_P as a function of normalized CHI2 and number of degrees of freedom

FCH / NFD	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20	1.25	1.30	1.35	1.40	1.45	1.50
8	.857	.819	.779	.736	.692	.647	.603	.558	.515	.473	.433	.395	.359	.326	.294	.265	.238	.213	.191	.170	.151
10	.891	.855	.815	.772	.725	.678	.629	.580	.532	.485	.440	.398	.358	.320	.285	.253	.224	.197	.173	.151	.132
12	.916	.883	.844	.801	.753	.703	.651	.598	.546	.495	.446	.399	.355	.314	.276	.241	.210	.182	.157	.135	.116
15	.942	.913	.878	.835	.787	.735	.679	.622	.564	.507	.451	.399	.350	.304	.263	.225	.192	.163	.137	.115	.095
20	.968	.946	.916	.877	.830	.776	.717	.653	.587	.522	.458	.397	.341	.289	.242	.201	.166	.135	.109	.088	.070
25	.982	.966	.941	.907	.863	.809	.747	.679	.607	.534	.462	.394	.331	.275	.224	.181	.144	.113	.088	.068	.052
30	.990	.978	.959	.929	.888	.835	.772	.700	.623	.544	.466	.391	.323	.261	.208	.163	.126	.096	.072	.053	.039
35	.994	.986	.970	.945	.908	.857	.794	.719	.638	.553	.468	.388	.314	.249	.193	.147	.110	.081	.058	.041	.029
40	.997	.991	.979	.957	.923	.875	.812	.736	.651	.561	.470	.384	.306	.238	.180	.134	.097	.069	.048	.033	.022
45	.998	.994	.985	.967	.936	.891	.829	.752	.663	.568	.472	.381	.298	.227	.168	.121	.085	.059	.039	.026	.017
50	.999	.996	.989	.974	.947	.904	.843	.765	.674	.574	.473	.377	.291	.217	.157	.110	.075	.050	.032	.020	.013
60	1.00	.998	.994	.984	.963	.925	.868	.790	.693	.586	.476	.371	.277	.199	.138	.092	.059	.037	.022	.013	.007
70	1.00	.999	.997	.990	.973	.941	.888	.810	.711	.596	.478	.364	.265	.183	.121	.077	.047	.027	.015	.008	.004
80	1.00	1.00	.998	.994	.981	.954	.904	.828	.726	.606	.479	.358	.253	.169	.107	.065	.037	.020	.011	.005	.003
90	1.00	1.00	.999	.996	.986	.963	.918	.844	.740	.615	.480	.352	.242	.156	.095	.054	.029	.015	.007	.003	.002
100	1.00	1.00	.999	.997	.990	.971	.930	.858	.753	.623	.481	.346	.232	.145	.084	.046	.024	.011	.005	.002	.001
110	1.00	1.00	1.00	.998	.993	.977	.939	.870	.765	.630	.482	.341	.223	.134	.075	.039	.019	.009	.004	.001	.001
120	1.00	1.00	1.00	.999	.995	.981	.948	.881	.776	.637	.483	.336	.214	.125	.067	.033	.015	.006	.003	.001	.000

- **AFP**: The equivalent disk surface radius (in km) of the mean antenna footprint surface. The mean antenna footprint surface is the weighted mean surface of the 3dB ellipses associated with the M_AVA views weighted by their inverse radiometric variances.

$$AFP = \sqrt{\sum_{k}^{M\_AVA} \frac{Footprint\ Axis1_k \times Footprint\ Axis2_k}{COVa_{k,k}} \Big/ tr\left(\frac{1}{COVa}\right)} \qquad \textbf{Eq 91}$$

- COV_a is already computed and defined section 3.2.4.3; tr(M) is the trace of the matrix M (the sum of the diagonal terms).

- VIEWS:
    - several counters report (detail in DAP, § 3.2.5.6.1) **views** conditions based on L1 information. It should include indications that radiometric uncertainties have been **enhanced** for some views (3.2.2.1.5).
    - FL_RFI_PRONE flag set to 1 when RRFI is above the threshold TH_CURRENT_RFI. A suggested value for TH_CURRENT_RFI is 0.1.
- **RETRIEVAL flags**. They include:
    - FL_RANGE, raised as soon as **any** retrieved parameter exceeds their allowed extended range (see TGRD UPF); The highest should be the one selected to go into TGRD UPF, unless we add one and have one reserved for soil moisture.
    - FL_DQX, raised as soon as **any** retrieved parameter exceeds their allowed range for DQX (see TGRD UPF); The highest should be the one selected to go into TGRD UPF, unless we add one and have one reserved for soil moisture.
    - FL_CHI2_P, provided for easier reading by users, is a flag that is raised whenever CHI_2_P does not belong to the interval of probabilities [TH_CHI2P_MIN, TH_CHI2P_MAX]. If the CHI_2_P probability of good fit is lower than TH_CHI2P_MIN then that means the fit is not good enough. If the CHI_2_P probability of good fit is greater than TH_CHI2P_MAX then that means the fit is too good to be true, suggesting possibly fitted noise.

3.2.5.5.2.2   Global quality figure

General considerations

The UDP includes several information concerning the confidence of products. In terms of user's needs however, it seems appropriate to build a global quality figure for retrieved SM.

There are a high number of error sources in the SMOS measurements. Some of them are and will remain poorly known, in such a way that some aggregation of error contributions seems allowed. Considered error categories are:

- Effect of radiometric noise and a priori uncertainties

- Instrument, calibration, reconstruction

- External sources (possible RFI, Sun, etc.)

- Forward models (including auxiliary data)

It is assumed that the shape which would be most convenient for users is a mark ranging from 1 to 20, with integer values, the best note being 1.

GQX should assess the confidence on retrieved SM. Then, it makes sense that the worst mark should correspond to SM uncertainty reaching 20% and therefore the retrieval providing practically no information.

Then, GQX should simply be approximately an estimate of the SM uncertainty. A mark equal to 1 would correspond to an SM uncertainty less than 1%. This is not within SMOS capabilities: the actual best mark would probably be 2.

As shown below, it will be possible to modify this scaling of the GQX easily.

Consistency between uncertainty contributions

The GQX is obtained through combining several potential error contributions. Every contribution includes a scaling coefficient that shall be defined in the **user's parameter file**.

Concerning radiometric noise and a priori uncertainties, this component GQX_1 is built using the retrieval uncertainty $RSTD_{SM}$.

For other categories, the errors appear more in terms of TB error than in term of soil moisture. Therefore, a function defining the sensitivity of SM to TB errors must be introduced.

From simulated retrievals, it appears that this sensitivity CCX varies mainly with the SM itself and the optical thickness. It is expressed below as a quadratic fit on both variables.

The sensitivity function CCX (% K^{-1}) allowing to combine TB errors in terms of SM error is proposed as follows: CCX=CCX0*(CCX1+CCX2*SM+CCX3*TAU+CCX4*SM^2+CCX5*TAU^2+CCX6*SM*TAU)

SM = retrieved soil moisture (%); TAU = retrieved nadir optical thickness; CCXn coefficients (n=1 to 6) =

+1.89962e-001 +1.22461e-002 +4.82144e-001 -4.91387e-005 +5.12738e-002 +5.48654e-003

The map shows the variation of CCX, which ranges between about 0.3 and 1.5 for the considered ranges of SM and TAU. While it is known that typically this figure is roughly of the order of 0.5 (i.e. about 50% SM for a 100 K range) on bare soil, approximating it by a single value would clearly not be satisfactory.

The fit coefficients CCX are stored in the user's file.

The suggested value for CCX0 is CCX0=1, for the case where the GQX is numerically equal to SM uncertainty. If the scaling of GQX is changed, CCX0 will be modified to maintain consistency with the retrieval uncertainty.

Figure 15: CCX variation map

Components of the GQX

The table below shows the contributions considered in building the GQX. It stipulates from left to right the error category EC, the origin of uncertainty, the "driver" D of the contribution, the way the contribution itself is computed using this driver D and coefficients GQXij, suggested GQXij values for the commissioning phase, units, and an assessment somewhat arbitrary of the nature of error (random or bias), with indication of the sign of the bias impact whenever possible.

Concerning the contribution of default fractions, at the onset of the commissioning phase, a large value will be considered for the relevant coefficient, as long as the quality of the retrieval is not well proven. Later on, this value might be reduced. This is also true for other error terms which are poorly known.

Table 28: Components of GQX

EC	Uncertainty origin	Driver D	Contribution GQXi	Init GQX value	GQX unit	Rand/ bias	Comments
1	Radiom ΔTB & prior	$D = RSTD_{SM}$	GQX11*D	1	[]	R	
2	Instrument	None	GQX21*CCX	1	K	B?	
2	Instrument	D = abs(X_SWATH)	GQX22*D*CCX	0	K Km^{-1}	B?	
2	Calibration	none	GQX23*CCX	1	K	B?	
2	Reconstruction	None	GQX24*CCX	0.2	K	B?	
2	Reconstruction	D=0/1 if FL_COAST =F/T	GQX25*D*CCX	0.1	K	B?	For scene dep. bias
2	Reconstruction	D=L1c T25 field #22	GQX26*D*CCX	0	[]	B?	
3	Goodness of fit	D=abs(CHI_2 - 1)	GQX31*D*CCX	0.05	K	R	1 K for 5% prob of no anomaly
3	Outliers	D=N_WILD / M_AVA	GQX32*D*CCX	10	K	R	1 K for 10%
3	RFI	CRFI				/	**For future version**
3	Various L1C flags	None				/	In DQX
3	SUN in front	D=SUN_FOV_C/M_AVA	GQX33*D*CCX	0.5	K	B-	
3	Rain	D=0/1 if FL_RAIN =F/T	GQX34*D*CCX	0.2	K	B?	
3	TEC	D=0/1 if FL_TEC =F/T	GQX35*D*CCX	0.2	K	B?	
3	Sky	D=N_SKY	GQX36*D*CCX	0.2	K	R	

4	Default fractions	D=**FDE** (non water defaults)	GQX41*D*CCX	20	K	R	10 K for FDE=0.5
4	FNO reference values	D=DLCC	GQX42*D*CCX	5	K	R	
4	DEW	None				R	
4	LITTER	D=0/1 if FL_LITTER =F/T	GQX43*D*CCX	0	K	B?	Not activated
4	Interception	D=0/1 if FL_PR =F/T; D=D*TAU	GQX44*D*CCX	1	K		1 K for TAU=1
4	Interception (aux)	D=0/1 if FL_INTERCEPT =F/T; D=D*TAU	GQX45*D*CCX	0.5	K		1 K for TAU=1
4	FLOOD probability	D=0/1 if FL_FLOOD_PROB =F/T; D=D*FWP*SM	GQX46*D*CCX	25	K/%	B+	1 K for FWP=0.1 & SM=0.4
4	Moderate topography	FTM	GQX47*D*CCX	6	K	B	1 K for 0.30
4	Strong topography	FTS	GQX48*D*CCX	20	K	B	1 K for 0.05
4	Evening orbit	D=L1c SPH field #09 D=0/1 if D =A/D	GQX49*D*CCX	1	K		1 K for evening orbit

Notes

- **Goodness of fit contribution**: While ideally this term ought to be driven by CHI_2_P using table 27, introducing temporarily CHI_2 rather than CHI_2_P is more realistic.

- **Default fractions: FDE (internal variable) is t**he sum of non-nominal fractions in the FM classes excepting open water (for which the reflectivity is known with little error).

- **FNO reference values**: an additional parameter DLCC is to be included in the LAND_COVER_CLASSES. The initial GQX42 value is selected assuming that the DLCC refers to the HR coefficient.

Final GQX formulation

Since there are many bias sources and in general their sign is undefined, it seems sensible to consider them as random contribution.

Therefore, the GQX will be the quadratic sum of every contribution in the table (all rows given column 4).

$$GQX = \sqrt{\sum GQX_i^2}$$ **Eq 92**

Other retrievals cases

They can be built in principle. However, the normalization between effect of radiometric uncertainty and additional biases no longer exists when considering the dielectric constant for example. Pending further work, the GQX will be left void in such cases.

Finally, it might be useful for some users to have a quality figure specific of the individual measurements (leaving out the impact of the retrieval). Unfortunately, there is no reliable way to assign a single radiometric error to every view, as this error varies widely over the field of view. A specific process combining L1c data with some pre-processing L2 steps will be needed.

3.2.5.5.3 Preparing PSF elements for UDP

SCENE flags**: a number of flags specifying scene features (see**

Table 32).

- **FL_BARREN**: this flag is set when FEB is above the threshold TH_SCENE_FEB.

- **FL_TOPO_S**: this flag is set when FTS is above the threshold TH_SCENE_FTS.

- **FL_TOPO_M**: this flag is set when FTM is above the threshold TH_SCENE_FTM.

- **FL_OW**: this flag is set when FOW is above the threshold TH_SCENE_FOW.

- **FL_SNOW_MIX**: this flag is set when FSM is above the threshold TH_SCENE_FSN.

- **FL_SNOW_WET**: this flag is set when FSW>TH_SCENE_FSW

- **FL_SNOW_DRY**: this flag is set when FSD>TH_SCENE_FSD

- **FL_FOREST**: this flag is set when FFO is above the threshold TH_SCENE_FFO.

- **FL_TAU_FO:** flag set when FFO > 0 and τ_{F_NAD} is above the threshold TH_SCENE_TAU_FO.

- **FL_NOMINAL**: this flag is set when FNO is above the threshold TH_SCENE_FNO.

- **FL_FROST**: this flag is set when FRZ is above the threshold TH_SCENE_FRZ.

- **FL_WETLANDS**: this flag is set when FWL is above the threshold TH_SCENE_FWL.

- **FL_URB_LOW**: this flag is set when FEU is above the threshold TH_SCENE_FUL.

- **FL_URBAN_HIGH**: this flag is set when FEU is above the threshold TH_SCENE_FUH.

- **FL_ICE**: this flag is set when FTI is above the threshold TH_SCENE_FTI.

- **FL_SEA_ICE:** this flag is set when FSI is above the threshold TH_SEA_ICE

- **FL_COAST:** the tidal flat flag FL_COAST is raised when the FWL fraction in the DFFG_INFO LUT is > 0 and the reference code is 242 (for the time being). Another field will have to be selected in the spares in due time to avoid confusion.

- **FL_FlOOD_PROB:** this flag is computed based on high retrieved soil moisture anomaly. Surface soil moisture is not expected to reach unrealistically high values. This being the case, it might be the consequence of flooding generating open water ponds unknown from the system. This flag is raised when the retrieved SM > TH_FLOOD + SM_DQX.

- The aggregated flags **FL_NON_NOM** is set if any of the above list of flags are set (excluding **FL_NOMINAL**).

- **FL_FLOOD_PRONE**: will be kept **unset** for now and is linked to the use of the previous flag and the DGG_CURRENT_FLOOD map.

- **FL_SAND**: flag raised when sand fraction (S) is above a given threshold TH_SAND.

OCCUR flags: a number of flags reporting some events/conditions are suspected to have occurred on the Earth's surface.

- **FL_DEW**: unset for now, will be addressed later.

- **FL_LITTER**: this flag is raised when τ_{LP} > TH_TAU_LITTER.

- **FL_PR**: this flag is set when PR_INDEX (see section 3.2.5.6.1) < threshold TH_PR. Suggested value; TH_PR = 0.026.

- **FL_INTERCEP**: this flag is raised when the ECMWF SRC (Skin Reservoir Content) is above a given threshold TH_SRC. Suggested value: TH_INTERCEP = 0.02 m.

EXTERNAL flags: some flags and one counter to report condition on external contribution.

- **FL_RAIN**: flag raised when TP = LSC+CP is above the TH_RAIN threshold for at least one of the nine ECMWF cells at the centre of the working area is with (LS+CP)>TH_RAIN.

- **FL_TEC**: flag raised when TEC content is above a given threshold TH_TEC.

- **N_SKY**: count the strong galactic sources detected; N_SKY is the number of views such as TB_{SKY} > TH_STRONG_TBSKY.

The aggregated **FL_SCENE**, **FL_OCCUR** and **FL_EXTERNAL** flags are raised when **any** of several flags belonging to a thematic family is raised.

Note that the FL_SCENE flag is expected to be always raised. If not it will be a symptom of something wrong is happening either in the overall logic on scene decomposition leading to an unhandled surface configuration.

The thresholds TH_SCENE_*** may or may not be equal to the decision tree thresholds; they have to be defined as separate entries in TGRD.

3.2.5.5.4   Preparing PPD elements for UDP

PPD for UDP only refers to pixels for which at least one retrieval attempt has been successful.

The following FL_CURRENT_*** flags are intended to the external updater mechanism for DGG_CURRENT_*** maps. When a flag is set, it informs the updater that something is to be done, update the map for the DGG or not depending on quality control and check.

- FL_CURRENT_TAU_NAD_LV is set when a TAU_NADIR for low vegetation has been retrieved, effectively stored in UDP and the retrieved conditions where optimal to update the current map (see section 3.2.5.2.2).

- FL_CURRENT_TAU_NAD_FO is set when a TAU_NADIR for forest has been retrieved, effectively stored in UDP and the retrieved conditions where optimal to update the current map (see section 3.2.5.2.2).

- FL_CURRENT_HR is set when a HR has been retrieved, effectively stored in UDP and the retrieved conditions where optimal to update the current map (see section 3.2.5.2.3).

- FL_CURRENT_RFI is set when either NRFI-X or NRFI-Y are not null (RFI has been detected online on L1c data or as outliers in the retrieval). Note that DGG_CURRENT_RFI will always be updated after each processed swath at least to update current N_{SNAP} with the M_AVA0 views.

3.2.5.6 Preparing Data Analysis Product (DAP) elements

Most of the information included in DAP are not subject to specific computation but are rather records of temporary or intermediate values used all along the process. The definitions in the DAP (section 3.4.4.2) are normally self-explanatory and will not be described, except for those of the following subsections.

3.2.5.6.1 Indices and numerical values

- **Interception using the** PR_INDEX[19]: As defined section 3.1.2.7, the usual definition of this polarization ratio is given for observed brightness temperatures with polarizations at the Earth reference frame (polarization H and V) for the incidence angle PR_INCI; low PR_INDEX values indicate risk of rain interception being present.

  Since the SMOS observation system provides brightness temperatures at the antenna reference frame (polarization X and Y) and for consecutive incidences angles (i.e. no exact same values are both equal to PR_INCI) the equivalent following formulation will be used but approximated for the geometry of the consecutive views ($TB_{p=X}(\theta_n, \alpha_n) / TB_{q=Y}(\theta_{n+1}, \alpha_{n+1})$ or $TB_{p=Y}(\theta_n, \alpha_n) / TB_{q=X}(\theta_{n+1}, \alpha_{n+1})$) such that their incidences angles satisfy $\theta_n \leq PR\_INCI \leq \theta_{n+1}$

$$\_INDEX = [TB_p(\theta_n, \alpha_n) - TB_q(\theta_{n+1}, \alpha_{n+1})] / [TB_p(\theta_n, \alpha_n) + TB_q(\theta_{n+1}, \alpha_{n+1})] \qquad \textbf{Eq 93a}$$

When:

$$(PR\_INDEX) < TH\_PR * abs(2 * (\sin(0.5*(\alpha_n + \alpha_{n+1})))^2 - 1) \qquad \textbf{Eq 93b}$$

then an interception event is suspected and the flag FL_PR is raised (suggested value: TH_PR = 0.026).

PR_INDEX is included in the DAP file.

- **Swath abscissa X_SWATH**: from L1c information, compute the abscissa X_SWATH of the dwell line corresponding to the L1c node. This quantity is equal to the distance of the L1c node to the ground track.

  Functionally it might be convenient to compute X_SWATH within the pre-processing steps.

  Note: X_SWATH can be directly obtained from a CFI call to the function **xp_target_extra_aux** output number #2. Alternatively, X_SWATH can be also computed from Figure 6 by using $-\left\|\overrightarrow{OP}\right\| \cos(\phi_g)$ for a given view; all the relevant information is available in L2 auxiliary data and L1c header to compute both the vector norm and the angle.

  Note: X_SWATH is an algebraic quantity with a sign given by $-\cos(\phi_g)$

- **RATIO_AVA** value: this value contains the ratio in % between the number of valid SMOS pixels that remains after the prior L1c filtering, M_AVA, (sections 3.2.2.1.4 to 3.2.2.1.7) and the initial available number of SMOS pixel M_AVA0:

$$RATIO\_AVA = M\_AVA / M\_AVA0 \qquad \textbf{Eq 94}$$

3.2.5.6.2 Status L1c counters

DAP status counters are computed from the L1c flags list (section 3.2.2.1.5). For each L1c flag, we count the number of views that raised the flags.

---

[19] Should this index prove to have a value as a geophysical information, It should be transferred to the UDP

For each *FLG*_C in {BORDER_FOV_C, SUN_FOV_C, SUN_POINT_C, SUN_TAILS_C, SUN_GLINT_FOV_C, SUN_GLINT_AREA_C}, compute $\text{FLG\_C} = \sum_{m=1}^{M\_AVA} \text{FLG}_m$

Specifically, for EAF_FOV, compute $\text{EAF\_FOV\_C} = \sum_{m=1}^{M\_AVA} \overline{\text{EAF\_FOV}_m}$

With these definitions, counters report the number of occurrences of problems. When counter indicates 0, no problem occurred.

3.3 Error Budget Estimates

From the soil moisture retrieval study, the following error budget table, Table 29, was built:

Table 29 : Tentative error budget origin and estimates

Error origin	Error level	Error impact on SM (%)		Comments
		random	biases	
Instrument and Reconstruction	0.2 K	0.1	Small	Along FOV averaging plus using redundancies
Sky and ionosphere	0.2 K	0.1	Small	Possible sun effects?
Atmosphere	0.1 K	0.05	Small	
Radiometric noise & a priori uncertainties on retrieved parameters		If 4 is allowed		- strict criterion ≈ 72% of the swath - mean criterion close to 100%
		If 3 is allowed		- strict criterion ≈ 60% of the swath - mean criterion ≈ 95%
Texture parameters			up to 2	
Soil roughness			up to 2	?
Internal non uniformities		Less than 0.4		For homogeneous nominal target part
Registration errors (wrt. aux data)				
Heterogeneity within pixels		0.4 (std 1.3)		Within 15% fraction limits

With respect to this table, it may be necessary to reassess the contribution to instrument errors, accounting for recent changes in the visibility equation. However, this is not likely to change the main features:

- The main contribution to error budget is the **radiometric uncertainty.**

- The retrieval error depends on the part of the instrument field of view where the data were collected. For this reason, any representative overall error budget can only be given as an **average** over the SMOS swath.

- The retrieval error depends on the **parameters themselves**: the retrieval error increases as either soil moisture or vegetation thickness increases. Therefore, any representative error budget can only be given for a **representative** land cover.

- Note this budget does integrate (last line) the impact of erroneous default contributions.

- Conversely, it does not account for **overall calibration bias**, the impact of which can be estimated using the information given in the GQX section.

3.4 Practical considerations

3.4.1 Calibration and Validation

Although calibration and validation methods are strictly speaking not relevant to this document and are covered by the SMOS Cal/Val AO answers at ESA. It is necessary to cover the algorithm validation approach (see [AD 11]). The approach is outlined below.

3.4.1.1 Calibration

The SMOS payload error budget accounts for the following 4 items:

- The basic calibration operation consists in calibrating the reference on-board radiometers, which are 3 noise injection radiometers (in short **NIR**). This operation implies the use of 2 (hot and cold) sources. The hot source is an on-board noise diode; the cold source is the sky. Hot calibration is performed at regular intervals during the flight, with a typical periodicity of one to 10 sequences within each orbit. Cold calibration, which implies a manoeuvre, is performed along one orbit every 4 weeks.

- Concerning the basic calibration operation, the influence of thermal driven, quasi-deterministic variations along each orbit due to changing illumination conditions must be accounted for. This is done either by increasing the frequency of hot calibrations, or by using deterministic relationships.

- Concerning the interferometer, the **on-board calibration system** supplies most of necessary information for subsystems included in the on-board calibration loop. Calibration sequences are performed at regular intervals during the flight (see above for NIR). Again, the influence of thermal variations along the orbit deserves particular consideration.

- As the antennas are not included in the calibration loop, variations of the antenna gains are obtained through looking at the sky (see above for NIR) in a region, near the galactic pole, where the sky radiation over the elementary antenna pattern width is close to uniform ("flat target").

The outputs of the calibration operations described above are incorporated in level 1 processing.

It is foreseen that additional calibration constraints, mainly on antenna properties, will be obtained from comparing redundant visibilities.

- It is estimated that the ultimate calibration can only be obtained using **vicarious** methods. The most likely candidates for this include one surface point at Dome C in the Antarctic, and mainly a network of drifting floaters over the ocean.

- In addition, statistical vicarious methods assuming stability of major features of the brightness temperature global distribution are considered.

- The geo-location accuracy will be tested through an analysis of SMOS data maps in the vicinity of well-known features such as isolated islands.

- The knowledge of many features of the auxiliary data (static maps of soil structure, thickness of forest cover) will be improved by analysing particular subsets of L2 data.

3.4.1.2 Validation

Algorithm validation plan will be based on using made targets, realistic synthetic scenes and eventually real data collected during campaigns. Validation is described in [AD 11] and will imply:
The core of the validation plan consists of three parts, which differ mainly by the data used as input.

- The first part (part 1) is mainly devoted at assessing the retrieval validity domain on conditions similar to those of previous sensitivity studies, i.e. use fully simulated academic data[20] in view of carrying out a systematic exploration of observing conditions for homogeneous scenes.

- In a second step (part 2) the SMOS performance as assessed through performing the retrieval on a large sample as provided by actual, realistic auxiliary data. The assessment will be carried out through statistical analysis of the results.

Optimization of various operating parameters is foreseen in both parts 1 & 2. The assessment of retrieval efficiency is foreseen in part 2.

---

[20] i.e., very schematic, simplified, not necessarily bearing a relation with real life data but elaborated to assess a given point

- Finally, the third part addresses the specific test sites which will be selected for product verification and prepares this verification.

Parts 2 & 3 are meant as **rehearsals** for the product verification phase. Inasmuch as they are properly designed, they should simply be repeated using actual SMOS data.

On top of these activities, specific retrieval algorithm subsets such as updating the current maps will be verified. This task should cover pending validation issues.

3.4.2   Quality Control and Diagnostics

For easier reading of the ATBD, these topics, which are closely linked to the theoretical algorithmic approach and to the analysis likely to trigger a second retrieval attempt, are addressed in section 3.2.5.1.

3.4.3   Exception Handling

Identified exception cases are twofold:

3.4.3.1   Mandatory data are not available

In the general sense many data are expected to be present and usable in the present document. If for some reasons it is not true, then it may result in a process failure.

However, when processing a DGG node, some events can be handled by an exception mechanism that tries to maintain the health for vital parts of the process.

In these cases, the L2 DAP output has to be defined in order to flag these events as well as to record the appropriate information to provide feedback, and ultimately the reasons to explain why the retrieval was not carried out.

The following is a starting point

- Missing values on auxiliary data: any auxiliary data that contain missing values and for which it exists a fallback (FB) default value in the TGRD/UPF table, then the fall-back value is used instead. This concerns namely the evolving aux data.
- Missing values or out of range values for some vital L1c data related to the views: incidence angle, radiometric accuracy, footprint azimuth direction …, discard the related views from the available list of views.
- We must finish this list and define DAP output so that missing data are identified.

3.4.3.2   Numerical computational exceptions

- Some of them may occur in some specific formulas but we know what to do. All the mathematical expressions will be checked against unsafe computation (mainly numerical inaccuracies and limit cases). Consider $\sqrt{x}$ in a given expression where x can reach the zero value but theoretically not below. Due to numerical inaccuracies x can be slightly negative but we know that, for this specific expression, we can state $\sqrt{0^-} \to 0$. Another classical example is $\sin(x)/x$ that is properly defined at 0 only for continuous topology. The classical approaches are to find all potential risk, and either enclose them in a handling/catch exception mechanism or rewrite expressions in an alternative safe form.
- Some of them may occur abruptly and are not expected. In this case, the priority is to avoid absolutely dropouts of the line. So, a chaining handling mechanism must be set up to catch at last any unhandled numerical exceptions and provide a fallback (flags, product content filled with identified errors values, …)

The specific case identified here concerns poorly conditioned matrices during the retrieval.

3.4.4   Output Products

This section considers:

- a **main data product** intended for end users: the **User Data Product** (**UDP**, section 3.4.4.1).
- an **auxiliary file** intended for specific users such as ESLs: the **Data Analysis Product** (**DAP**, section 3.4.4.2).
- and a summary about UDP/DAP reporting values (section 3.4.4.3).

PCD and PSD defined in 3.2.5.5 are part of both UDP and DAP.

The **size** of elements in Table 31 to Table 36 is to be considered as a first indication.

3.4.4.1 User Data Product

This subsection provides the list of data we expect to be present in **L2 UDP**. The main goal is to provide a description with some details, but only from a conceptual standpoint. Therefore, no numerical accuracy, exact item number, specific format nor logical files assignation (single or multiple) will be provided.

Headers are not considered in detail. They are expected to include:

- Many header fields of the input data. Of particular importance are the ASCENDING_FLAG and the MODE. Information concerning the extent space-time domain will be defined.

- Information concerning retrieval options: possibly part of the UDP file, at least references to the version of the file being used for the retrieval.

The UDP consists in a list of **fixed-size** data records. Each record is associated with a single DGG node number, for which a L2 retrieval processing has been **successfully** carried out.

The UDP will contain an indication on the time of acquisition. This will be the median time between start and end of acquisition (first and last view of the L1C node among the M_AVA really used view)

The data in the file will not be scaled so the fields are 4 byte long: total 128 bytes

Table 30: User Data Product (UDP)

Origin	Name	Description	Units
L1c data	DGG_NODE	DGG node ident, (λ, φ) geographic location and altitude[21]	
L1c Data	AVG_TIME	Mean time of acquisition for all the views of the node	s
L2 retrieval or default	SM	If free, the soil moisture (SM) of the soil layer estimate. If fixed, the median of WEFs fixed values in ESL mode, -999 in normal user mode (OP)	[%]
	DQX_{SM}	If SM is retrieved its Data Quality Index; otherwise = -999	
L2 retrieval or default	τ_{NAD}	If free, the nadir optical thickness estimate. If fixed, the median WEF fixed value for τ_{NAD} in ESL mode, -999 in normal user mode (OP)	[na]
	$DQX\tau_{NAD}$	If τ_{NAD} is retrieved its Data Quality Index; otherwise = -	
L2 retrieval or default	T_{SURF}	If free, the surface temperature estimate. If fixed, the median WEF fixed value for T_{SURF} in ESL mode, -999 in normal user mode (OP)	[K]
	DQX_{TSURF}	If T_{SURF} is retrieved its Data Quality Index; otherwise = -999	
L2 retrieval or default	TTH	If free, the angular correction parameter estimate for $\tau\mathbf{SH}$ If fixed, the median WEF fixed value for TTH in ESL mode, -999 in normal user mode (OP)	
	DQX_{TTH}	If TTH is retrieved its Data Quality Index; otherwise = -999	
L2 retrieval or default	RTT	If free, the ratio of angular correction parameter TT_V/TT_H estimate If fixed, the median WEF fixed value for RTT in ESL mode, -999 in normal user mode (OP)	
	DQX_{RTT}	If RTT is retrieved its Data Quality Index; otherwise = -999	

---

[21] Should be the first record

Origin	Name	Description	Units
L2 retrieval or default	ω_H	If free, the horizontal polarization albedo estimate. If fixed, the median WEF fixed value for ω_H in ESL mode, -999 in normal user mode (OP)	[]
	$DQX\omega_H$	If ω_H is retrieved its Data Quality Index; otherwise = -999	
L2 retrieval or default	$DIFF\omega$	If free, the difference of albedos $\omega_V - \omega_H$ estimate. If fixed, the median WEF fixed value for $DIFF\omega$ in ESL mode, -999 in normal user mode (OP)	[]
	$DQX_{DIFF\omega}$	If $DIFF\omega$ is retrieved its Data Quality Index; otherwise = -999	
L2 retrieval or default	H_R	If free, the roughness parameter estimate. If fixed, the median WEF fixed value for H_R in ESL mode, -999 in normal user mode (OP)	[]
	DQX_{HR}	If H_R is retrieved its Data Quality Index; otherwise = -999	
L2 retrieval	ε'	ε real part, from MD retrieval: ε' values are computed from the retrieved A_card and the fixed U_card (see section 3.1.4.7).	[]
	$DQX_{\varepsilon'}$	If retrieval successful, = DQX propagated to ε' ; otherwise = -999	
L2 retrieval	ε''	ε imaginary part, from MD retrieval: ε'' values are computed from the retrieved A_card and the fixed U_card (see section 3.1.4.7).	[]
	$DQX_{\varepsilon''}$	If retrieval successful, = DQX propagated to ε''; otherwise = -999	
L2 retrieval or default	ε'_D	ε real part, derived from retrieval models other than MD: ε'_D values are computed from retrieved physical parameters (see section 3.1.2.2)	[]
	$DQX_{\varepsilon'_D}$	If retrieval successful, = relevant DQX_{PAR} propagated to ε'_D; otherwise = -999	
L2 retrieval or default	ε''_D	ε imaginary part, derived from retrieval models other than MD: ε''_D values are computed from retrieved physical parameters (see section 3.1.2.2)	[]
	$DQX_{\varepsilon''_D}$	If retrieval successful, = relevant DQX_{PAR} propagated to ε''_D; otherwise = -999	
L2 processing	$TB_{H\theta_B}^{ASL}$	Surface level TB (including sky/atmosphere contribution) computed from forward model at specific incidence angle θ_B (42.5 °), for H polarization.	[K]
	$DQX\ TB_{H\theta_B}^{ASL}$	Data Quality Index of $TB_{H\theta_B}^{ASL}$ (see section 3.2.5.3).	[K]
L2 processing	$TB_{V\theta_B}^{ASL}$	Surface level TB (including sky/atmosphere contribution) computed from forward model at specific incidence angle θ_B (42.5 °), for V polarization.	[K]
	$DQX\ TB_{V\theta_B}^{ASL}$	Data Quality Index of $TB_{V\theta_B}^{ASL}$ (see section 3.2.5.3).	[K]
L2 processing	$TB_{X\theta_B}^{TOA}$	Top of Atmosphere TB (including atmosphere and sky contributions) computed from forward model at specific incidence angle θ_B (42.5 °), for X polarization.	[K]
	$DQX\ TB_{X\theta_B}^{TOA}$	Data Quality Index of $TB_{X\theta_B}^{TOA}$ (see section 3.2.5.3).	[K]

Origin	Name	Description	Units
L2 processing	$\mathrm{TB_Y}{}^{TOA}_{\theta_B}$	Top of Atmosphere TB (including atmosphere and sky contributions) computed from forward model at specific incidence angle θ_B (42.5 °), for Y polarization.	[K]
	$\mathrm{DQX\ TB_Y}{}^{TOA}_{\theta_B}$	Data Quality Index of $\mathrm{TB_Y}{}^{TOA}_{\theta_B}$ (see section 3.2.5.3).	[K]
L2 retrieval, L1c data	PCD for UDP	Confidence descriptor. See Table 31	
L2 retrieval, L1c data	PSF for UDP	Science flags. See Table 32	
L2 processing, L1c data	PPD for UDP	Processing descriptor. See Table 33	

Concerning the **DQX:**

- For retrieved parameters, DQX is the Data Quality Index; at least it represents the theoretical parameter posterior standard deviation RSTD, but can be larger accounting for the knowledge of other sources of uncertainty such as repetitive RFI contamination (see section 3.2.4.4.2 and 3.2.4.4.3).
- The **header** should include information allowing obtaining the a priori standard deviations ASTD, in such a way that very small DQX figures for parameters, which are very tightly constrained, are not subject to misinterpretation. Additional information in cases where current tables are used is stored in the DAP.
- For not retrieved parameters or failed retrieval DQX is set to a missing value.

The **Product Confidence Descriptor** (Table 31) includes both scalars and [0 / 1] flags. Some flags summarize other flags or counters relative to L1c data (views) and retrieval.

Table 31: Product Confidence Descriptor (PCD) for UDP

Name	Origin / meaning	Notes	Size
Confidence flags			bits
FL_RFI_PRONE	R$_{RFI}$ above threshold	increased DTBa	1
FL_RANGE	retrieved values outside extended range	Highest of all selected	1
FL_DQX	high retrieval DQX	Highest of all selected	1
FL_CHI2_P	poor fit quality	Raised if CHI2_P outside of [TH_CHI2P_MIN, TH_CHI2P_MAX]	1
FL_NOPROD	No Product is provided	Many possible reasons, retrieval failed, results out of range	1
		SPARE to 8:	3
Confidence descriptors			bytes
GQX	Overall quality index	The goal is to produce a grade between 1 and 20	1
CHI_2	Retrieval fit quality index		1
CHI_2_P	χ^2 high value acceptability probability		1
N_WILD	wild data still present	count	1
M_AVA0	Initial number of views available in L1c		1
M_AVA	Pre-processing; number of views available for retrieval		1
AFP	The radius of the disk surface equivalent to the mean antenna footprint 3dB ellipse	In km See Eq 91	4

Name	Origin / meaning	Notes	Size
	surface		
PROBA_RFI	R_{RFI} computed from the DGG_CURRENT_RFI maps matching the same orbit pass (header ASCENDING_FLAG)	PROBA_RFI > 0.1 ⇒ care is required for the DGG and FL_RFI_PRONE is raised	1
N_AF_FOV(=FALSE)		Count views	1
N_SUN_TAILS	from L1c or pre-processing	Count views	1
N_SUN_GLINT_AREA		Count views	1
N_SUN_FOV_C		Count views	1
N_Software_Error	From L1C		2
N_Instrument_Error	From L1C		2
N_ADF_Error	From L1C		2
N_Calibration_Error	From L1C		2
N_X_Band	From L1C		2
N_RFI_Mitigations	From L1C		2
N_Strong_RFI	From L1C		2
N_Point_Source_RFI	From L1C		2
N_Tails_Point_Source_RFI	From L1C		2
		TOTAL bytes:	34

Table 32: Product Science Flags (PSF)

Name	Origin / meaning	Notes	bits
	Science flags		
FL_NON_NOM	Presence of **other** than nominal soil		1
FL_SCENE	Summary flag: any of following flags on	Consistency check – always 1	1
FL_BARREN	Presence of rocks		1
FL_TOPO_S	Presence of strong topography		1
FL_TOPO_M	Presence of moderate topography		1
FL_OW	Presence of open water		1
FL_SNOW_MIX	Presence of mixed snow		1
FL_SNOW_WET	Presence of wet snow		1
FL_SNOW_DRY	Presence of significant dry snow		1
FL_FOREST	Presence of forest		1
FL_TAU_FO	Large forest optical thickness		1
FL_NOMINAL	Presence of nominal soil		1
FL_FROST	Presence of frost		1
FL_ICE	Presence of permanent ice/snow		1
FL_WETLANDS	Presence of wetlands		1
FL_URBAN_LOW	Presence of limited urban area		1
FL_URBAN_HIGH	Presence of large urban area		1
FL_SAND	Presence of high sand fraction	No change MN processing	1
FL_SEA_ICE	Presence of sea ice		1
FL_COAST	Presence of large tidal flat		1
FL_DUAL_RETR_FNO_FFO	FNO and FFO fractions fusion for the retrieved parameters has been activated		1
FL_FLOOD_PROB	Probable flooding risk		1
FL_FLOOD_PRONE	Possible flooding risk	Not activated	1
FL_OCCUR	Summary flag: any of following flags on		1

Name	Origin / meaning	Notes	bits
FL_DEW	Dew suspected	place holder	1
FL_LITTER	Litter suspected	place holder	1
FL_PR	Interception suspected (Pol ratio)		1
FL_INTERCEP	ECMWF indicates interception		1
FL_WINTER_FOREST	NPE winter forest rule has been activated		1
FL_EXTERNAL	Summary flag: any of following flags on, or N_SKY counter not null		1
FL_RAIN	heavy rain suspected	Threshold TH_RAIN	1
FL_TEC	High ionospheric content	Threshold	1
FL_FARADAY_ROTATION_ANGLE	Faraday Rotation Angle coming from observation and not from model.		
		SPARE to 32	0
	Science descriptors		bytes
N_SKY	strong galactic source	threshold TH_SKY; count views	1
		Total Bytes:	1

In this table, flags FL_** are triggered when a threshold is exceeded. Threshold values are provided in TGRD/User

The PPD specifies main retrieval options and conditions.

Table 33: Product Process Descriptor (PPD) for UDP

Name	Origin / meaning	Notes	size
	Processing flags		bits
FL_R4	R4 attempted		1
FL_R3	R3 attempted		1
FL_R2	R2 attempted		1
FL_MD_A	MDa failed	part failure	1
	CURRENT flags		bits
FL_CURRENT_TAU_NAD_LV	DGG_CURRENT_TAU_NADIR_LV map update requested	Header ASCENDING_FLAG determines which current map to update: asc. or dsc.	1
FL_CURRENT_TAU_NAD_FO	DGG_CURRENT_TAU_NADIR_FO map update requested	Header ASCENDING_FLAG determines which current map to update: asc. or dsc[22]	1
FL_CURRENT_HR	DGG_CURRENT_HR map update requested	Header ASCENDING_FLAG determines which current map to update: asc. or dsc	1
FL_CURRENT_RFI	DGG_CURRENT_RFI update requested	Header ASCENDING_FLAG determines which current map to update: asc. or dsc	1
FL_CURRENT_FLOOD	DGG_CURRENT_FLOOD map update requested	Not activated for the time being	1
		SPARE to 16	3
		Note: Current using flags have been shifted to DAP and then	

---

[22] asc. stands for ascending orbit pass and dsc. for descending.

Name	Origin / meaning	Notes	size
		shifted back to UDP	
	Processing descriptors		Byte
S_TREE_1	Branches of decision tree stage 1	The winter forest exception case is normal forest branch + FL_WINTER_FOREST set	1
S_TREE_2	Retrieval R2, R3 or R4		1
	CURRENT update values		byte
DQX_TAU_CUR	For external post-process of the current map corresponding to the L2 ASCENDING_FLAG	DQX_TAU for updating CURRENT map assuming all parameters are free	4 (float)
DQX_HR_CUR	For external post-process process of the current map corresponding to the L2 ASCENDING_FLAG	DQX_HR for updating CURRENT map assuming all parameters are free	4 (float)
N_RFI_X	For external post-process process of the current map corresponding to the L2 ASCENDING_FLAG	RFI detected in L2 test X pol; count of deleted views.	1
N_RFI_Y	For external post-process process of the current map corresponding to the L2 ASCENDING_FLAG	RFI detected in L2 test Y pol; count of deleted views.	1
		Total Bytes:	14

Table 34: S_TREE_2 interpretation

Encoding	Reserved		Model (MN, MW, MD)		TAU (min,med,high)		Retrieval Case: Rx	
Bits	7 (MSB)	6	5	4	3	2	1	0 (LSB)
Retrieval Case: Rx								
No Retrieval	xx		Xx		xx		00	
R2	xx		Xx		xx		01	
R3	xx		Xx		xx		10	
R4	xx		Xx		xx		11	
TAU (min,med,high)								
[0 TH_23]	xx		Xx		00		xx	
[TH_23 TH_34]	xx		Xx		01		xx	
> TH_34	xx		Xx		10		xx	
Reserved	xx		Xx		11		xx	
Model (MN, MW, MD)								
MN	xx		00		xx		xx	
MW	xx		01		xx		xx	
MD	xx		10		xx		xx	
Reserved	xx		11		xx		xx	

3.4.4.2 Data Analysis Product

This product is intended for specific users (such as ESLs). This is more than a simple log file for debugging purposes, since the SMOS mission is very innovative and will require extensive analysis after launch.

The DAP exist for every L1c product ingested by the L2 SM processor. They should contain all relevant data for performing algorithm improvements, fine tuning and performing validation and calibration. We could also foresee that auxiliary input data for L1/L2 process could be improved in time with the analysis of this product.

For example, let us consider the case of RFI and its distribution in space and time. With both χ^2 maps and residuals distribution analysis, it is hoped to improve the auxiliary RFI map.

Unlike the UDP, where the content and structure should not depend on what algorithm is used (e.g. iterative inversion of forward models vs. a possible direct statistical inversion), the DAP content and structure depend necessarily on the adopted algorithm. For iterative retrieval, we anticipate the following characteristics and content.

The DAP consists of a list of **variable-size** data records. Each record is associated with unique ISEA node numbers, the ones for which the L2 retrieval processing has been applied, **successfully or not**. Thus, the DAP will contain **exactly the same number of DGG nodes as its L1c product input**.

The content of this product can be divided in two parts:

- A **descriptor (numerical)** data part that records relevant temporary values, used and/or created by the algorithms e.g. vector of residuals $TBM_m - TBF\left(\alpha_m, p_i^* \Lambda\right)$, weighting fraction list... This **variable-size record** is described in Table 35. **Further efforts** have been invested to complement the content of Table 35 with more information that should be relevant to analyse the SML2PP during the commissioning phase.

- A **status (flags)** part, which records all the events that occurred in the processing. This status traces the L2 processing behaviour in generating parameters, as well as in rejecting the L1c nodes. This could be a fixed-size packed bit-field that describes branching conditions, operations, exceptions, flags… This **fixed size record** is described in Table 36.

DAP records the **whole** PSD and PCD outputs, **excluding** the parts recorded in UDP.

Each node present in the L1c input product is associated with the record described in the following table:

Table 35: DAP descriptors[23]

Name	Origin	Description	Bytes
DGG_NODE	L1c Input Data	DGG node ident, (λ, φ) geographic location and altitude	4 * 4
M_AVA0	Pre-process / Post-process	Total number of views available	1
VRES	Iterative retrieval	Vector of Residuals TBM_m-TBF_m. VRES[k] = -999 when a view, k, was not used in retrieval (removed because RFI, Range, Outliers)	M_AVA0 x 4 (floats)
C_FM0	Pre-process	Mean cover fractions for each decision tree class (from MEAN_ WEF and FM$_0$ list))	12 x 2
C_FM	process	Mean cover fractions (from MEAN_ WEF and FM list)	10 x 2
M_ANG	Pre-process / Post-process	Number of unique geometries among the M_AVA kept views. In dual polarization mode M_ANG is always equal to M_AVA. In full polarization mode M_ANG is smaller than M_AVA (M_ANG $\cong$ 2/3 M_AVA) (see section 3.2.1.2).	1
C_FV	process	Cover fractions for each unique view geometry (from WEFs and FV)[24]	M_ANG x 10 x 2
X_SWATH	Pre-process	Abscissa of dwell line (+ - km)	2
N_TB_RANGE		L2 testing TB against range; count of deleted views	1
RATIO_AVA	Pre-process	Ratio of useful views	1
REF_VAL	Process	Initial values for free parameters	9 x 4 (floats)
ASTD	Process	Initial std for free parameters	9 x 4 (floats)
N_CLEANED		wild data removed (count)	1
NIT		Number of iterations to convergence	1

---

[23] Also includes latitude longitude and altitude of course

[24] if strip adaptative option is used, this is a vector since all AFP are the same.

Name	Origin	Description	Bytes
N_Retries	Post-process	Number of "retrieval again" attempted	1
PR_INDEX	Post-process		1
TAU_LV_IN	Pre-process	TAU_LV taken from CURRENT map of the same orbit pass (ASCENDING_FLAG)	4 (float)
DQX_TAU_LV_IN	Pre-process	DQX_TAU_LV taken from CURRENT map	4 (float)
TAU_FO_IN	Pre-process	TAU_FO taken from CURRENT map of the same orbit pass (ASCENDING_FLAG)	4 (float)
DQX_TAU_FO_IN	Pre-process	DQX_TAU_FO taken from CURRENT map of the same orbit pass (ASCENDING_FLAG)	4 (float)
HR_IN	Pre-process	HR taken from CURRENT map of the same orbit pass (ASCENDING_FLAG)	4 (float)
DQX_HR_IN	Pre-process	DQX_HR taken from CURRENT map of the same orbit pass (ASCENDING_FLAG)	4 (float)
TAUL	Process	TauL(SM) if litter effect is activated, 0 otherwise	2
TPHYS[25]	Post-process	Tgc(SM,Tau$_{NAD}$) for the retrieved fraction	2
		TOTAL for M_AVA=60, dual polarization	About 1610
		TOTAL for M_AVA=90, M_ANG=60, full polarization	About 1730

Table 36: DAP flags

Name	Origin / meaning	action	Notes	(c)	bits
SUN_POINT_C	from L1c or pre-processing	delete view			1
SUN_GLINT_FOV_C	from L1c or pre-processing	none			1
FL_DATA_MISS	mandatory data set missing	no retrieval	check fallback options?		1
FL_MVAL0	MVAL0 < TH_MMIN1	no retrieval			1
FL_MVAL	MVAL < TH_MMIN1	no retrieval			1
FL_R4_NITM	R4 attempted, failed NITM				1
FL_R4_KDIA	Failed KDIA				1
FL_R4_COND	R4 attempted, failed COND				1
FL_R4_RANGE	retrieved parameters in range failed				
FL_R4_DQX	DQX in range failed				
FL_R3_NITM	R3 attempted, failed NITM				1
FL_R3_KDIA	Failed KDIA				1
FL_R3_COND	R3 attempted, failed COND				1
FL_R3_RANGE	retrieved parameters in range failed				
FL_R3_DQX	DQX in range failed				
FL_R2_NITM	R2 attempted, failed NITM	failure			1
FL_R2_KDIA	Failed KDIA				1
FL_R2_COND	R2 attempted, failed COND	failure			1
FL_R2_RANGE	retrieved parameters in range failed				
FL_R2_DQX	DQX in range failed				
FL_MD_NITM	MDa failed NITM (a)	part failure			1
FL_MD_KDIA	Failed KDIA				1

---

[25] No more used in modelling but reported as a weighted average value using the median view WEF as weight for analysis purpose.

Name	Origin / meaning	action	Notes	(c)	bits
FL_MD_COND	MDa failed COND (a)	part failure			1
FL_MD_RANGE	retrieved MDa A_card in range failed				
FL_MD_DQX	DQX A MDa card in range failed				
FL_CE	Computational exceptions	failure?	Place holder		
			SPARE to 32		15

(a) MDa = additional MD retrieval

(b) update RFI map

(c) Number is required for summing views

3.4.4.3   UDP/DAP value / DQX reporting summary

The following table summarizes the logic of free and fixed parameter values reported in UDP/DAP as a function of the retrieval status.

Table 37: UDP/DAP parameter value interpretation

CASE	UDP		DAP		Flags raised	Observation
	Value Reported	DQX Reported	Value Reported	DQX Reported		
No Retrieval Tried						
Free Parameters	-999	-999	-999	-999	FL_NOPROD	e.g. AVA too small
Fixed Parameters	-999	-999	-999	-999		
Retrieval Failure						
Free Parameters	Value retrieved[26] or -999	DQX value retrieved or -999	Initial Value	ASTD	FL_NOPROD FL_RANGE FL_DQX	1. Failure Reasons: Value out of bounds, DQX out of bounds, NITmax exceeded.
Fixed Parameters	Ref Value[26] using WEF of median view or -999	-999	Ref Value[26] using WEF of median view or -999	-999		2. FL_NOPROD (0) and NITmax (0 or >0) make a distinction between no retrieval attempted and retrieval failure
Successful Retrieval						
Free Parameters	Value retrieved[27]	DQX value retrieved	Initial Value	ASTD		Value in DAP - for median view, corresponding to the Retrieval Fraction
Fixed Parameters	Ref Value using WEF of median view	-999	Ref Value using WEF of median view	-999		
MDa Retrieval						

---

[26] Only in ESL processing mode with no range constraint, otherwise -999 in standard user mode

[27] In ESL processing mode no range constraint, forced to belong to [P_{min} P_{max}] in standard user mode

Not Tried (MD as main retrieval)	Epsilon MD values, free and fixed parameters reported as for the cases above				Flags set as above	MDa not attempted, as MD model ran in main retrieval
Not Tried (Epsilon MD)	-999	-999	-	-	FL_NOPROD FL_MD_A	MDa is not attempted when MN/MW is in 'No Retrieval Attempted' or 'Retrieval Failure'
Retrieval Failure *(Epsilon MD)*	Value retrieved[26] or -999	DQX value retrieved or -999	-	-	FL_MD_A FL_RANGE FL_DQX	The FL_MD_A flag, part of Product Process Descriptor (PPD) of UDP, flags the MDa failure.
Successful Retrieval *(Epsilon MD)*	Value retrieved	DQX value retrieved	-	-		

3.5 Assumptions and limitations

Refer to section 2.4.

3.5.1 Forward models when used in retrieval system

The retrieval procedure described section 3.2.4 is an unconstrained minimization of a cost function and nothing prevents the retrieved parameters values to become non physically plausible if the cost function is improved in that way. Besides, when the iterative process explores the parameters space before the convergence is obtained even for a meaningful value, these parameters may temporarily have values which are non-physically plausible and used in the numerical computation of models. Apart when obviously wrong optimal solutions can be obtained, direct numerical error exceptions may also occur, even for just temporary illegal values, e.g. taking the logarithm of an unexpected negative value, leading to a retrieval failure.

There are traditionally three ways to take into consideration these aspects:

1. we can use a constrained optimization procedure that guaranty while in the procedure that parameters stay in a valid range. The retrieval procedure become more complex and usually increases the number of iterations and the number of operations, making the retrieval more time consuming. The random error distribution of these parameters is also half-cut at the boundaries creating positive biases at lower limit and negative bias at the upper limit.

2. the valid range constraint can be introduced as a penalty term in the cost function when the parameters values are outbound. Although this approach usually works, it requires to tune properly the weight of the penalty to avoid making the retrieval unstable. Besides it assumes that the forward models can accept forbidden values and do not generate numerical exceptions.

3. models can be modified and regularized to behave properly in these situations by extending the model to drive back the parameter values in the right direction with a mathematical formulation that avoid numerical exception. Depending of the mathematical formulation of the models this task may be easy to difficult or even not possible. Done correctly, it also removes the error biases at the boundaries.

The Dobson and Mironov models, presented section 3.1, are subject to these limitations with respect to the soil moisture parameter. Both models lead to a maximum of emissivity for negative soil moisture value and the Dobson model has a vertical asymptote for negative values that leads to possible numerical exceptions, but more than anything can trap the soil-moisture to high negative value.

For our purpose the option 3) is the most convenient approach while mathematically speaking the option 1) is the cleanest. Basically we will change the $\varepsilon_b(SM,...)$ function to a symmetrized version similar to $\varepsilon_b(|SM|,...)$ but using an approximation C2 class formulation instead of taking the absolute value. The continuity of both the function and its derivative are required the Levenberg-Marquardt minimize. $\varepsilon_b(SM,...)$ is prolongated for the interval of $|SM|<TH_{SM}$ by a parabolic approximation form such that the continuity at the jointing point $SM=TH_{SM}$ of the original function and its prolongation as well as their derivatives

is guaranteed. Besides, we are also looking for a prolongation such that its derivatives at SM=0 is equal to 0 to insure the stabilization of convergence around SM=0.

This can be formulated by finding the a(...), b(...), c(...) complex coefficients functions in the following system of equations:

if $|SM| > TH_{SM} \rightarrow \varepsilon_b^S(SM;\Lambda) = \varepsilon_b(|SM|,\Lambda)$ normal dielectric constant function **Eq 95**

if $|SM|\ TH_{SM} \leq TH_{SM} \rightarrow \varepsilon_b^S(SM;\Lambda) = a(\Lambda)\,SM^2 + b(\Lambda)\,SM + c(\Lambda)$ prolongation **Eq 96**

isfying the constraints:

$$\varepsilon_b^S(TH_{SM};\Lambda) = \varepsilon_b(TH_{SM};\Lambda)$$

$$\left.\frac{\partial \varepsilon_b^S}{\partial SM}\right|_{TH_{SM}} = \left.\frac{\partial \varepsilon_b}{\partial SM}\right|_{TH_{SM}}$$

$$\left.\frac{\partial \varepsilon_b^S}{\partial SM}\right|_0 = 0$$

Eq 97

The general solution of the above system is as follow:

$$a(\Lambda) = \frac{1}{2} \times \frac{d\varepsilon_1(\Lambda)}{TH_{SM}}$$

$$b(...) = 0$$

$$c(\Lambda) = \varepsilon_1(\Lambda) - \frac{1}{2} \times d\varepsilon_1(\Lambda) \times TH_{SM}$$

Eq 98

$$\text{with } d\varepsilon_1(\Lambda) = \frac{d\varepsilon_b(TH_{SM};\Lambda)}{dSM} \text{ and } \varepsilon_1(\Lambda) = \varepsilon_b(TH_{SM};\Lambda)$$

Thanks to the very clean and simple mathematical formulation for Mironov model, it allows us to solve analytically the system and provide the expression for a(...), b(...), c(...) coefficients functions. Note that the coefficients functions are complex numbers to address the real part and the imaginary part of the dielectric constant and are functions of the other parameters used in Mironov formulation, C, the clay fraction and, T, the soil temperature, through n_d, k_d, n_b, k_b as they are defined in Eq 19, Eq 20, respectively. The coefficient functions are computed by replacing the following in the above Eq 98.

$$\varepsilon_1(C,T) = (\beta^2 - \alpha^2) - \mathbf{j}(2\alpha\beta)$$
$$d\varepsilon_1(C,T) = (2\beta(n_b - 1) - 2k_b\alpha) - \mathbf{j}(2ab)$$
$$\alpha = k_d + k_b \times TH_{SM}$$
$$\beta = n_d + (n_b - 1) \times TH_{SM}$$

Eq 99

However, it exists a potential caveat as a symmetric function $\varepsilon_b^S(SM;\Lambda)$ creates two identical solutions one for SM and one for –SM and thus two local minima. It is not much expected that the optimizer jumps on the negative domain to find the negative solution, but the risk cannot totally be eliminated especially with a noisy system. In such conditions potentially valuable retrieved solutions will be discarded by the post retrieval analysis (out of domain). To avoid this side effect, it is recommended to forward the absolute value on SM internally to the optimizer when associated with this symmetrization. Symmetrisation of HR(SM;...) and Tg(SM;...) is also required to behave properly for negative SM.

The same approach can be applied to Dobson model but due to its complex formulation, the final solution requires direct functions calls to the model to compute $\varepsilon_l(TH_{SM},C,S,T,\rho_b)$ and, by finite differences, $d\varepsilon_l(TH_{SM},C,S,T,\rho_b)$ in order to use the Eq 98; the computation cost is then largely increased.

3.6 Reprocessing considerations

Even though the processor described in this ATBD is intended to be run, in its operational version, it is also foreseen that after launch and commissioning phase, our knowledge of the whole system will be greatly improved. We should then be in a position to sort out the issues / unknowns detailed in this document, discover mistakes etc. Similarly, L1c data will probably be reprocessed with better calibration, improvements in the L1 processor etc, offering an improved input data. Consequently, reprocessing will take place on regular occasions. It is highly desirable that every reprocessing occasion is used to reprocess

using an improved version of the L2 soil moisture processor but also to take full advantage of the bigger time lag to improve input data sets.

Thus, not detailing all the potential L2 processor improvements, and not addressing the fine-tuning options, we will only suggest what routine improvements could be done to improve reprocessing:

The most obvious is to use current LAI Maps. In the routine processor we can only use available maps, hence maps which correspond to the previous compositing period: at best they will be 15 to 20 day old. For reprocessing the current LAI map could be used: i.e., the compositing period containing the swath acquisition date.

The processor uses ECMWF forecasts. Obviously, it seems better to use instead analysis with the drawback that the time lapse between two is 6 hrs. It seems nevertheless a better option.

In some cases, operational input data favored for the processor could be replaced by better though non-operational products (for instance AMSR/SSMI snow and ice fields, high resolution maps of floods …) but this might require specific work to be implemented.

3.7   Conclusions: further developments

3.7.1   Sand

An expression such as given by Eq 47 could be used in place of the classical approach described for nominal surfaces after validation with actual data (either from campaigns or from actual SMOS data) provided the soil moisture dependency be explicitly described, which is not currently the case. It would then be applicable for surface where the sand content exceeds a very high threshold TH_SAND ($\approx$ 95% or 97 %).

The main source of concern is for large sandy patches (coastal, estuaries) and for arid deserts (Sahara, Arabian peninsula, Kalahari, Australian center) that show very special behaviors (though at higher frequencies). Several potential causes are currently under analysis:

As the penetration depth might be very high (tens of meters), the effective temperature will be hard to assess especially with a surface having a very large diurnal cycle

Similarly, the signal might be relative to deep subsurface features (rocky/laterite beds, fossil water tables, unsuspected deep roughness features (see for instance the Sudanian subsurface river network).

This part is still to be consolidated when we have more substantial evidence. The current understanding is that when the soil map indicates a sand content above some very high fraction, the soil dielectric model would switch from the current to a specific one, yet to be given. We are actually testing other algorithm which could eventually always replace the ones currently used.

3.7.2   Snow

Snow covers about 40% of the Northern hemisphere land mass seasonally but has very different dielectric properties depending on its history. Fresh, dry snow is transparent to microwave radiation, however as snow melts its dielectric constant increases dependent upon snow grain size and liquid water content and may be totally opaque (at Te $\cong$ 273 K) when wet. Grain size and stratification also play a role, as well as the position of snow within the different layers (on ground, on vegetation, on trees, etc.). Consequently, the effects of snow are too complicated to be incorporated into the currently proposed algorithm, and areas with significant snow coverage other than dry snow must be considered as irretrievable, for estimation of only the dielectric constant. The issue will be in identifying and flagging the snow-covered areas.

The location of non-permanent snow areas can be derived directly from the SSM/I data product 'Near Real-Time SSM/I EASE-Grid Daily Global Ice Concentration and Snow Extent' (http://nsidc.org/data/nise1.html)

The Near Real-Time SSM/I EASE-Grid Daily Global Ice Concentration and Snow Extent product (Near real-time Ice and Snow Extent, NISE) provides daily, global near real-time maps of sea ice concentrations and snow extent. The National Snow and Ice Data Center (NSIDC) creates the NISE product using passive microwave data from the Defense Meteorological Satellite Program (DMSP) F13 Special Sensor Microwave/Imager (SSM/I). Sea ice concentration and snow extent maps are provided in two 25 km azimuthal, equal-area projections: The Southern Hemisphere 25 km low resolution (Sl) and Northern

Hemisphere 25 km low resolution (Nl) Equal-Area Scalable Earth-Grids (EASE-Grids). Data are updated daily and are available via ftp and NSIDC's Data Pool for two weeks after initial posting. Snow extent is mapped using an algorithm developed for Scanning Multichannel Microwave Radiometer (SMMR) data and modified for use with SSM/I data [127, 128]. The NISE product is updated daily using the best available data from the past five days.

Dry snow can be considered transparent at L-band microwave [129] and ignored. Snow deep enough to affect L-band microwave is typically found only in areas that will be excluded from soil moisture retrieval because of highly variable topography, and in permanent snow areas, which are excluded from retrieval by their land cover type. The hazardous areas for retrieval are the edges of snow extent, where snowmelt can occur, creating wet snow with a more problematic microwave response. The presence of melting snow can be identified by strong returns in SSM/I data, or in the absence of this data by combining the ECMWF surface temperature field with the SSM/I snow extent data to identify points at the edge of snow areas with a temperature range liable to give rise to snow melt. Where melting snow is probable, a caution flag should be set, and one possible way of improving retrieval might be to consider the melt as equivalent to open water and allow the open water fraction to be retrieved as a free parameter. In reprocessing, it may be possible to model the snow layers in these areas using antecedent precipitation and temperature data, and account for the effect of melting snow on the radiative transfer. When the areas are largely covered with snow, (at high latitudes) the problem will be relatively simple and NSIDC/ NOAA products can be used. At lower latitudes, snow cover and state can be expected to vary quite a lot with space and time. Depending on area and state of the art, it will be necessary to use either NOAA/ EUMETSAT daily maps or model outputs (ECMWF type).

Dealing with snow other than dry is a topic for further research. As a preliminary approach, we suggest defining 3 categories for non-permanent snow cover: dry, wet, and mixed or intermediate. They will be defined through comparing an estimate of the snow temperature T_SNOW to a couple of thresholds. If the snow is dry, it will be assumed transparent and ignored; if it is wet, it will be assumed opaque and subject of a possible retrieval focused on the snow-covered zone, for its dielectric constant properties. In the intermediate case, a default dielectric constant will be retrieved for the whole land area. The basic input will be ECMWF SD until more efficient data such as NSIDC SSM/I, AMSR-E or other future dataset become available. Thus, tests will be conducted by the ESLs with these new data and should they prove more suited and or more accurate data source will be changed accordingly and reported in TGRD update.

In preparation of these tests, a snow map fraction ADF at the DFFG scale is available for the algorithms. The DFFG cells of this map of snow fractions, if not missing values, supersede the ECMWF SD based fractions of snow. The use of this DFFG snow fraction ADF can be activated/deactivated by a general switch of the UPF.

The NSIDC IMS products proved to be a good candidate for such use. Despite the snow fractions are available only for the northern hemisphere, the 4km resolution version of these products is very close to the DFFG specification and is used to feed the DFFG snow ADF (see TGRD for details and references). The southern hemisphere snow fractions fall back to the standard ECMWF SD derived fractions.

The NSIDC IMS based DFFG snow fractions ADF is recommended to be used and activated in operational processing.

3.7.3   Flooded areas

The probability of precipitation-based flooding within a specific area depends upon either high-sustained local precipitation, or high nearby precipitation and a water channel link to the area of precipitation. The precipitation within a given area will be thresholded to determine whether a cell has probably been flooded. To account for flood waters flowing down water channels, where a water channel links an area marked as probably flooded to an area without flood-level precipitation, for a given distance down the river channel, it should be marked as possibly flooded.

Work along those lines has been initiated by Reading University and should be developed.

In past versions of the algorithms the probability of flood flag FL_FLOOD_PROB was to be set when the ECMWF precipitation ([LSCP + CP] or TP) is greater than a rain threshold TH_FLOOD and a surface NPE rule was triggered to convert the associated DFFG cells to the wetland class (FWL). The forecasted rain rate was not spatially accurate enough to be used for such purpose and the NPE flooding rule has been abandoned. Still the FL_RAIN flag reports such events and the FL_FLOOD_PROB is now raised on high retrieved soil moisture anomaly (see section 3.2.5.5.3).

In the future, this simple computation may be improved by using information from an auxiliary DGG_CURRENT_FLOOD map, an introducing an additional flag FL_FLOOD_PRONE.

The possibility of flood FL_FLOOD_PRONE flag should be set when a cell is linked to another cell via a **DFFG**_WATER_CHANNEL_CONNECTION LUT which has a FL_FLOOD_PROB flag set, and closer to it than a TBD threshold *flood_spatial_extent*, which may need to be a function of precipitation. One operational way to implement this flag-setting would be to take a similar approach to distance transformation in image processing. A number of runs equal to the

number of km in *flood_spatial_extent* would be carried out over the FL_FLOOD_PRONE array. In each run, those cells connected via the **DFG**_WATER_CHANNEL_CONNECTION LUT to a cell flagged as FL_FLOOD_PROB or FL_FLOOD_PRONE would be marked with the FL_FLOOD_PRONE flag. Thusly, in each run, the FL_FLOOD_PRONE flags would extend further along river features away from areas with high precipitation.

3.7.4 Other radiative model updates

Model for vegetated soil: remove excessive sensitivity near vanishing SM values
Model for open water: optimize parameterization leaving open the physical temperature

3.7.5 Auxiliary data

Devise a method for updating the soil structure parameters
Devise a method for accounting for small open water surfaces.

Annexes

3.8   *TGRD Cross Reference*

This section provides an index of the TGRD keywords used in the ATBD. Normally all the keywords in the following index should exist in the TGRD and would help to track discrepancies, missing … between the two documents.

For this purpose, two quotation categories, "TGRD UPF" and "TGRD LUT" have been created in this word document. The former is for values declared in the UPF of TGRD, and the latter correspond to other LUTs (DGG o DFFG maps and general purposes tables).

It is now possible to introduce reference keys under those categories for each word in ATBD that is defined in TGRD, and thus to ask word to draw the index table for those categories.

We need to pay attention if a name in these lists is not found in TGRD; when this is not due to synonyms (symbols are acceptable, bad names require to be fixed in ATBD) or typos (requires to be fixed in ATBD), it may signify a missing keyword in TGRD.

Many occurrences of TGRD keywords still need to be quoted to update those two lists.

TGRD LUT

TGRD UPF

3.9 ASSESSMENT OF DECISION TREE

Here are the preliminary results of the decision tree described in depth section 3.2.3 and applied on the global ECOCLIMAP 2004 data base. It is performed over the whole globe and aims at assessing, roughly, how many SMOS pixels will be considered as "nominal".

To perform this analysis, topography, freezing conditions, ice, were ignored. Only permanent snow was considered. Actually, various set of thresholds values have been considered and the following presented Table 38 is a good candidate to consider as a basis.

Table 38: Global model repartition: threshold values

Threshold name	%	Comment
STO	**5**	**not used (topography)**
SS2	98	Snow, used for permanent snow.
SS1	5	
SR2	**98**	**not used (frozen)**
SR1	**5**	
SNO	40	Soil + low veg nominal
SFO	60	soil + forest nominal
SWE	90	Wetlands
SWO	90	Water (only water bodies like lakes, not inland sea)
SEB	80	Barren (here the ECOCLIMAP rock label is used, this is not the best, but low impact)
SEI	**80**	**Not used (ice)**
SEU	80	Urban (for ECOCLIMAP, this threshold too high => no impact)

3.9.1.1.1 ECOCLIMAP 2004 global fractions

The ECOCLIMAP aggregation is built on a lat-lon grid (nlat, nlon) = 540x1080 that corresponds roughly to 40kmx40km cell at the equator, 40kmx20km at 60°of latitude.

Table 39: Global ECOCLIMAP fractions over continent

DT class type	ECOCLIMAP classes	Fraction of continental area (%)
Complementary fractions		
Snow	permanent snow	**31.1**
Soil+low veg	flat bare soil + grassland + tropical grassland+ C3 crops + C4 crops + irrigated crops	**46.5**
Forest	Broadleaf tree + coniferous tree + tropical tree	**20.0**
Wetlands	park marshes	**0.7**
Barren	Rocks	**1.7**
	Total	**100.0**
Supplementary fractions		
Open water	Water	**8.7**
Urban	**Urban**	**0.1**

3.9.1.1.2 Results

The following output was obtained with the decision tree.

Table 40: Global model repartition: surface fractions

Models type	Number of pixels	% of continent area
c2asn: Snow (box model)	61852	29.7
c2bdd: Snow (dielectric)	5299	2.5
c4ann: Nominal soil + low veg (model)	110278	53.1
c4bnn: Nominal soil + forest (model)	27322	13.1
c5ww: Wetland (dielectric)	1	0.0
c6os: Open water (model)	185	0.1
c7add: Barren (dielectric)	434	0.2
c7cdd: Urban (dielectric)	0	0.0
c7ddd: Other (dielectric)	2795	1.3
Total	208166	100.0

A rough-cut test indicates that the percentage of land pixels is 208166/ (1080*540) = fraction of continent = 35.69 % instead of the commonly accepted 29%.

The 6% discrepancy comes from the fraction resolution at 40 km that has a tendency to thicken the coastline that overlaps a little the ocean. Consequently, the surface of continent appears overestimated.

The Urban threshold was never activated. The urban ECOCLIMAP class is slightly different of those of DT, so its threshold needs to be adjusted.

3.9.1.1.3 First analysis

The fraction of nominal behavior, i.e. SM + τ, is: c4ann + c4bnn that is ~ 66%. It seems not very much, but the main limiting factor is permanent snow on poles (~ 30%) that is not really relevant to SM studies. If we do not consider permanent snow areas, we have **95.8 % of nominal pixels** and this is a good result, even though optimistic this time as we will have topography, non-permanent snow, ice, etc.

3.9.1.1.4   What's next:

The above assessment is preliminary and needs further refinement and verifications:

1.   the ECOCLIMAP rocks label is used for decision tree barren. The true meaning of ECOCLIMAP rock label as well as the meaning of barren must be verified.

2.   urban class has no impact. For ECOCLIMAP urban is the fraction of manmade elements within towns. So even considering a big city filling a 40 km x 40 km square, this fraction will not be 100%, but depends on gardens, trees, etc., leading rather to 60-80 %.

3.   wetland, there is a problem 0.0005% is too small (we can think that large north-eastern Europe or Australian region are not capture). So, the wetland label in ECOCLIMAP is to be controlled.

4.   The seasonal characteristics are not accounted for (seasonal snow, ice, frozen surface). This will be done very soon since access to ECMWF data is being to be solved.

5.   c7ddd represents what has not been identified and labelled by the decision tree, that is 1.3 % which seems a little high and needs to be explained.

6.   Topography has still to be accounted for.

7.   Seasonal effects are also to be accounted for.

8.   this analysis is purely geometrical. The WEF has to be accounted for, for completeness.

3.9.1.1.5   Conclusion:

Finally, excluding ocean surfaces, polar ice & snow surfaces that are out of concern for SM, we can remember that the SM + τ model will be used at best in 94 % of the cases. Those 94 % will necessarily decrease because:

- of seasonal aspects, frozen, ice, snow
- of the use of WEF. In the next assessment we will try to tackle the first two aspects.