Read-me-first note for the release of the SMOS Level 2 Soil Moisture data products	
Processor version	Level 2 Soil Moisture V700
Release date by ESA	25th May 2021
Authors	Expert Support Laboratory (ESL) Level 2 Soil Moisture + ARGANS
Further information	Information on how to access the SMOS data can be found here: SMOS Science Products - Earth Online (esa.int) (https://earth.esa.int/eogateway/catalog/smos-science-products) Details on the processing algorithms can be found in the Algorithm Theoretical Baseline Document (ATBD) • SO-TN-ARG-L2PP-0037 v4.0 Information about the L2 Soil Moisture (L2SM) products structure can be found in the SMOS Level 2 and Auxiliary Data Products Specifications document: • SO-TN-IDR-GS-0006 v8.6 The documents are available here: SMOS Science Products - Earth Online (esa.int) (https://earth.esa.int/eogateway/catalog/smos-science-products) Information about the SMOS Level 1 data quality can be found in the Monthly Quality reports and in the Mission Reprocessing report available here: MIRAS QUALITY CONTROL REPORTS - Earth Online (esa.int) (https://earth.esa.int/eogateway/instruments/miras/quality-control-reports) Additional information, including change history of L2SM algorithm and associated processor documentation can be found on the SMOS L2SM CESBIO website: https://labo.obs-mip.fr/smos-blog and ARGANS website: https://argans.github.io/smos-l2sm-processor/
Contact for helpline	Please contact ESA's HelpDesk on eohelp@esa.int.
Comments to ESL Level 2 soil moisture team	The Level 2 Soil Moisture team would like to get your feedback on the product, either directly (yann.kerr@cesbio.cnes.fr, marias@argans.co.uk, Klaus.Scipal@esa.int) or through the BLOG (https://labo.obs-mip.fr/smos-blog) where you can also find the latest news!

1. Introduction

This note summarises the quality of the SMOS Level 2 Soil Moisture data products generated by version v700 of the Level 2 Soil Moisture Processor (L2SM).

Version 700 of the Level 2 Soil Moisture data product is now available for the entire SMOS mission lifetime with the following file class and version:

File class	File version	From	To
REPR	v700	1 June 2010	24 May 2021
OPER	v700	25 May 2021	present

Measurements from the **commissioning phase** (12 January 2010 - 31 May 2010) show drifts due to instrument tests taking place during this period. Even though data are available (upon request) it is not advisable to use them.

SMOS data users are invited to read this note carefully to ensure optimum exploitation of the version 700 dataset, which supersedes the previous version 650. Further information on the quality of the dataset can be found in the reprocessing verification report and in the validations report (available from June 2021 onwards) here: https://earth.esa.int/eogateway/instruments/miras/quality-control-reports

This note is organised as follows:

- Main improvements to the L2SM version 700 dataset (Section 2).
- L2SM version 700 dataset performance and caveats (Section 3).
- Future algorithm evolutions (Section 4).

2. Main improvements of the current L2SM version 700 dataset

The main improvements introduced in the current operational version v700 of the SMOS Level 2 Soil Moisture products (as compared to the previous version v650) are related to significant processing algorithm updates, processing parameters configuration and auxiliary files changes, besides the use of the new SMOS Level 1C v724 brightness temperature products.

The first aspect is the following algorithm updates:

- The retrieved Chi2 post-rescaling that was introduced in v650 to mitigate any possible mischaracterization of added observational errors has been replaced by the integration of an estimate of such errors, σ_{IRE}, directly in the cost function normalization along with the observation radiometric resolution. Both approaches aim at making the «Chi2P» (the probability of the Chi2) and the «FL_Chi2_P» flag usable for additional quality control checks. The new approach in v700 allows the model fits to readjust consequently in the retrieval.

- Bircher’s organic soil dielectric constant empirical model is added. Now, the dielectric constant of all wet ground surfaces (FNO, FFO) is considered as a weighted mixture of mineral soil (Mironov’s model) and organic soil (Bircher’s model). See *Bircher S. & al., A new dataset of resonant cavity measurements and model evaluation. Remote Sensing, MDPI, 2016, 8 (12), pp.1-17* or the ATBD for more details.

- The nominal (FNO) and forest (FFO) are packed to a single fraction for the free parameters retrieval, still keeping the specific fixed parameters properties linked to each of these surfaces.

The second aspect concerns the following configuration changes and ADFs changes which are summarized in Table 1:

- The IGBP saline water fraction (FWS) and fresh water fraction (FWP) in the AUX_DFFFRA auxiliary file are replaced by water fraction derived from the ESA CCI-Land Cover (LC) radar water body @ 250m (more specifically, derived from the Copernicus Climate Change Service C3S GLC 2017 dataset). As a consequence, all other fractions increase/decrease accordingly.

- The AUX_LANDCL single scattering albedo (ω) for low vegetation land cover classes are changed from 0 to 0.06 and their associated roughness (HRmax=0.1, HRmin=0.05) changed to (HRmax=0.3, HRmin=0.1).

- The fraction of organic soil (RSOM) field is added in the auxiliary data file (ADF) to the texture map (AUX_DFFSOI) to support the introduction of Bircher's organic soil dielectric constant model. The RSOM fractions DFFG @ 4km are derived from SoilGrids ORCDRC maps @ 250m with SOC > ~102 g/kg.

Table 1- Changes in auxiliary files between v650 and V700 dataset

ADFs	v650	v700
AUX_DFFFRA	FWP, FWS, from IGBP @ 1km resolution Other fractions from IGBP	FWP, FWS, from ESA CCI-LC radar water body map @ 250m resolution. Other fractions from IGBP
AUX_LANDCL	FNO associated landcover ω=0 HRmax=0.1 HRmin=0.05	FNO associated landcover ω=0.06 HRmax=0.3 HRmin=0.1
AUX_DFFSOI	-	RSOM, the ratio of organic soil to mineral soil, is added

The water fractions (FWO) derived from the ESA CCI-LC (C3S GLC 2017) radar water body at 250m resolution map usually increase when compared to the IGBP ones at 1km resolution. The 250m resolution captures more details of the river networks and small lakes particularly at Northern latitudes as it can be seen in the four plots of Figure 1. These plots show the difference of FWO v700 – v650 in percentage for four different ranges of values from ±2.5% to ±25%. This enables to appreciate the magnitudes of the changes and their locations. The main consequence by having more water fraction is to make the contribution of cold brightness temperature more significant in the retrievals modelling and thus to reduce retrieved soil moisture values for those areas where the v650 was found to be too wet.

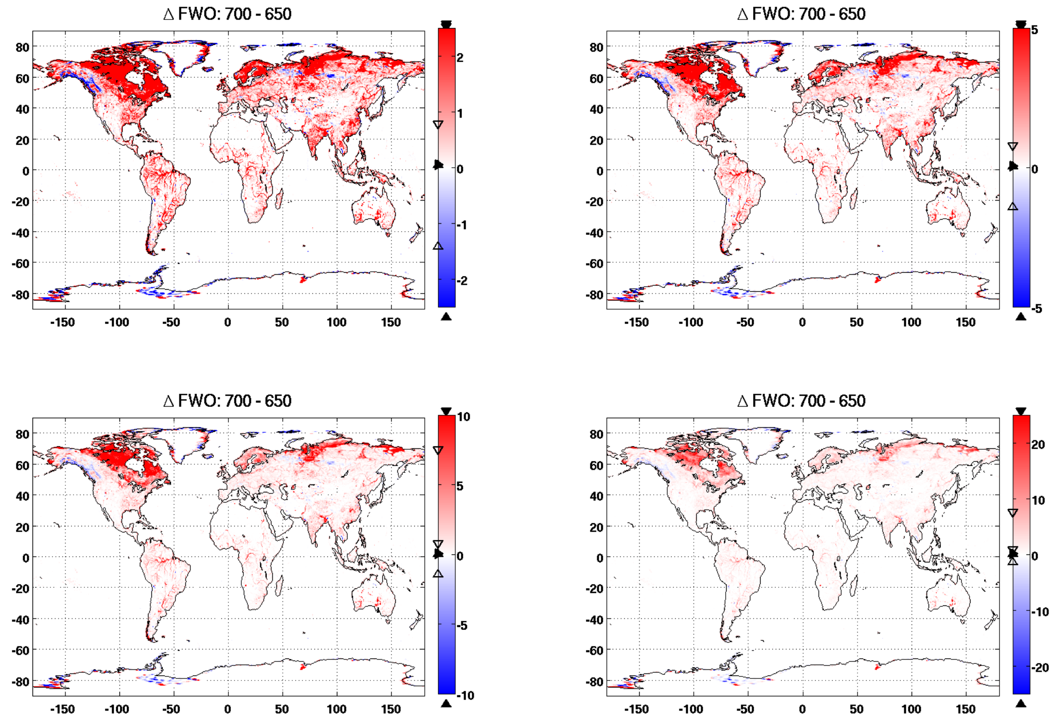


Figure 1: FWO differences v700-v650 in % for different colour-bar ranges.

The IGBP land-cover classes being the same between the two versions, only marginal relative changes are observed in the model selection (decision tree) due the FWO change. On the static map below (Fig. 2), the two green colours represent the maximum extent (not considering snow, frozen soil events) where the soil moisture and opacity retrievals can be carried on. For the other categories the ground dielectric constant and the opacity are retrieved.

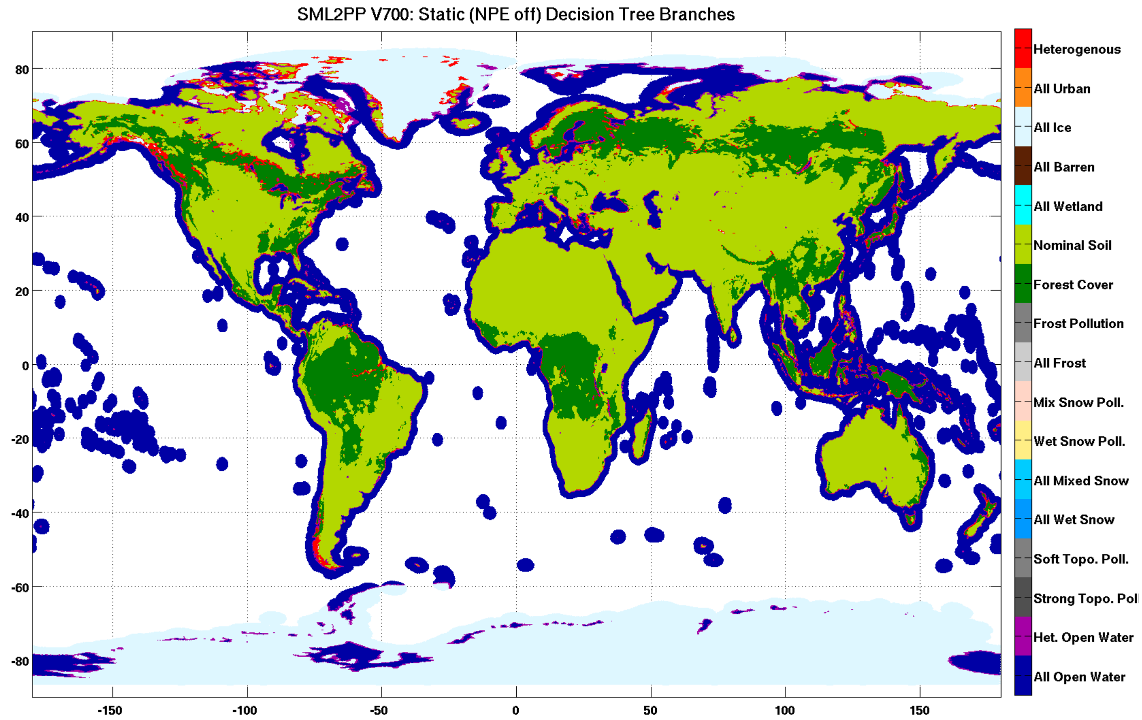


Figure 2: Decision tree branch cases for v700 retrieved models.

3. L2SM version v700 dataset - performances and caveats[1]

3.1 Performances

The comparison of the new v700 with respect to the v650 is performed against a collection of calibration-validation sites time-series of 12 networks of *in-situ* measurements, namely SCAN (157 stations), OZNET (17 stations), AMMA (2 stations), HOBE (6 stations), SMOSMANIA (21 stations), REMEDHUS (8 stations), VAS (1 Elbara retrieval), WATERSHEDS (4 averaged stations), SODANKYLA (1 averaged station), MESONET (103 stations), ARGENTINA (19 stations) and IGARKA (2 stations).

The v700 (2) and v650 (1) retrievals performances are shown by the two following Taylor plots diagrams and numerical metrics; they were made on the concatenation of all-time series to provide an overall performance as a single time series of filtered (Chi2 probability and RFI) and co-located SMOS/in-situ data representing more than 70000 samples in each single co-located time series (see Figure 3 and 4).

The v700 retrieved soil moisture correlation (R) with *in-situ* soil moisture is significantly improved, by 0.14 (ascending orbits) and by 0.10 (descending orbits), as well as the standard deviation of soil

---

[1] The products released are prone to be improved. Feedback to Yann Kerr (yann.kerr@cesbio.cnes.fr) and Manuel Arias (MArias@argans.co.uk) on any issues you may identify while using or validating the data is most welcome.

moisture difference (STDD) SMOS-*insitu* decreased by 18% (ascending orbits) and by 10% (descending orbits). Both are also consistently improved across all networks/stations compared to v650.

Concatenated debiased allsites CalVal Sites Insitu Time Series: 308 kept / total 383 - Ascending orbits
CalVal Series Rejection Correlation Threshold if all scenarios corr < 0.500 or it exits R pvalues > 0.05 or #data < 20

Δ statistics: SMOS - Insitu

Series	R	Bias	STDD	RMSD	#kept
v650	0.53	-0.022	0.084	0.087	77139
v700	0.67	-0.052	0.069	0.087	93041

SMOS/Insitu series statistics

Series	$Mean_{SMOS}$	$Mean_{Insitu}$	STD_{SMOS}	STD_{Insitu}
v650	0.190	0.213	0.093	0.078
v700	0.164	0.217	0.089	0.077

Filter Stats

Series	v650 All	v650 Inter	v700 All	v700 Inter
Initial	207764	171444	207852	171519
Retrieved	76.16%	75.57%	76.17%	75.58%
Successful	95.62%	95.66%	97.29%	97.25%
PSchi2	60.79%	61.40%	74.70%	74.17%
PRFI	59.19%	59.67%	72.57%	71.88%
RRFI	59.07%	59.54%	72.46%	71.77%

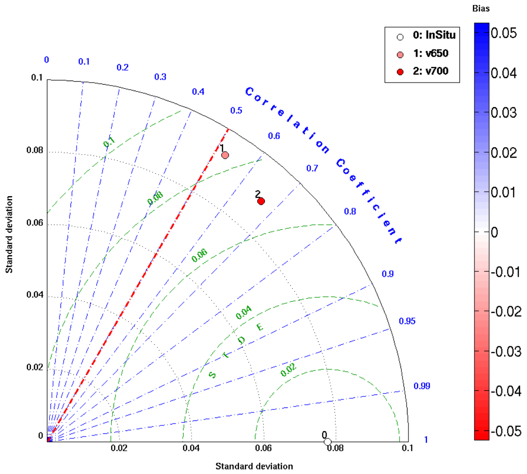


Figure 2: Statistics computed over the sites after concatenating all the data to have reliable metrics; ascending orbits.

Concatenated debiased allsites CalVal Sites Insitu Time Series: 309 kept / total 385 - Descending orbits
CalVal Series Rejection Correlation Threshold if all scenarios corr < 0.500 or it exits R pvalues > 0.05 or #data < 20

Δ statistics: SMOS - Insitu

Series	R	Bias	STDD	RMSD	#kept
v650	0.56	-0.025	0.081	0.085	72243
v700	0.66	-0.042	0.073	0.084	86124

SMOS/Insitu series statistics

Series	$Mean_{SMOS}$	$Mean_{Insitu}$	STD_{SMOS}	STD_{Insitu}
v650	0.190	0.215	0.093	0.079
v700	0.177	0.220	0.095	0.078

Filter Stats

Series	v650 All	v650 Inter	v700 All	v700 Inter
Initial	211724	171241	211713	171243
Retrieved	76.42%	76.18%	76.45%	76.21%
Successful	95.92%	95.85%	97.51%	97.44%
PSchi2	56.80%	56.36%	68.72%	67.43%
PRFI	55.93%	55.45%	67.39%	66.07%
RRFI	55.86%	55.38%	67.30%	65.99%

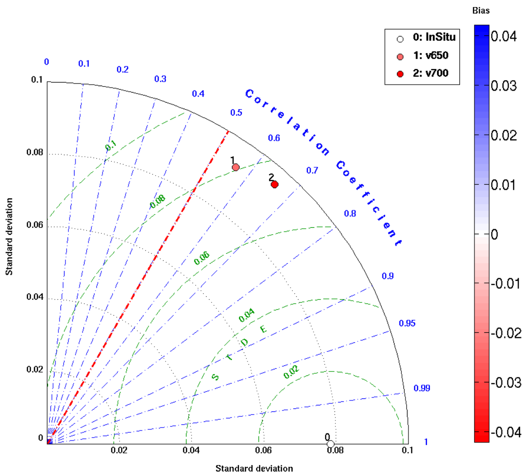


Figure 3: Statistics computed over the sites after concatenating all the data to have reliable metrics; descending orbits.

As shown in Table 2, at global scale, the v700 provides more successful retrievals for soil moisture and opacity, about 5% globally with a strong spatial heterogeneity ranging from 1.3% to 12.8%, depending on areas.

The negative bias (SMOS – *in-situ*) increased by -0.03 m3/m3 (ascending orbits) and by -0.02 m3/m3 (descending orbits). Unlike R and STDD, there is more variability across networks/stations. For the sites where v650 retrieved soil moisture was wetter (e.g., WATERSHEDS, AMMA, OZNET) the v700 lower retrieved values almost cancel the positive biases. For the sites where it was the opposite (e.g., MESONET, SMOSMANIA, ARGENTINA) then the negative bias increases. For other sites like SCAN it is balanced from wetter than in-situ for v650 becoming drier for v700 but with equivalent absolute magnitude biases.

Finally, it is worth noting than the RMSD is similar between the two versions meaning that the above changes are mostly linked to redistribution between systematic (bias) and dynamics/random (R & STDD) errors, improving the latter potentially at the expense of the former.

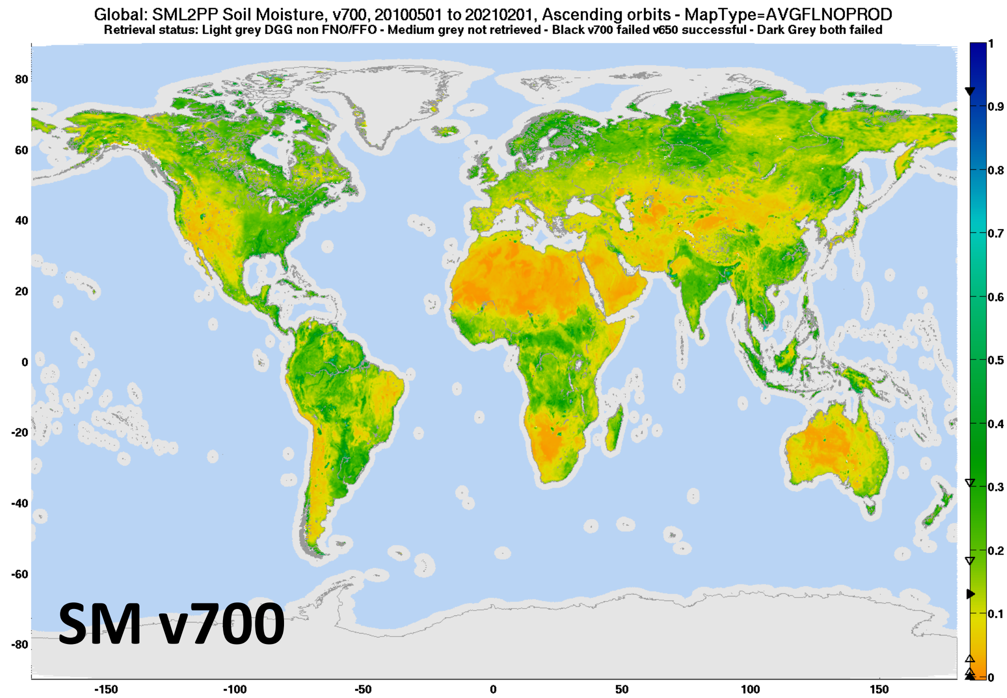
Table 2: retrievals performances V700 versus v650 over the whole 10+ years reprocessing campaign.

		Whole reprocessing (20100501 - 20210131) – Nominal (DT12) & Forest (DT11)							
		North America	South America	West Europe	Africa	East Europe	Asia	Australia	Global
ASC	v700	96.2 %	97.0 %	86.3 %	89.3 %	89.6 %	72.7 %	96.8 %	89.6 %
	v650	90.2 %	92.6 %	77.2 %	86.7 %	82.6 %	65.3%	96.1 %	84.2 %
	Δ7-6	+6.0 %	+4.4%	+9.1 %	+2.6 %	+7.0 %	+7.4 %	+0.7 %	+5.4 %
DSC	v700	96.7 %	94.8 %	74.0 %	92.1 %	75.2 %	73.3 %	96.1 %	87.8 %
	v650	91.6 %	91.1 %	61.2 %	90.8 %	66.5 %	65.9 %	95.5 %	82.6 %
	Δ7-6	+5.1 %	+3.7 %	+12.8 %	+1.3 %	+8.7 %	+7.4 %	+0.6 %	+5.2 %

On global maps, v700 retrieved soil moisture is lower than v650 close to water body as expected by the increased water fraction and for certain range of low vegetation due to the increased single scattering albedo (ω=0→0.06). For some places (mostly at North latitudes) it is balanced by the strong presence of organic soil which is now modelled in the v700 which often leads to wetter retrieved soil moisture than v650. Figure 5 to Figure 6 illustrate these changes with respect to soil moisture.

The most noticeable difference on retrieved opacity, though visible also on retrieved soil moisture maps though fainter, is the smoothing effect introduced by the joint retrieval over the nominal (FNO) and forest (FFO) fractions. There is also an increase of the retrieved opacity at high latitude probably in connection of soil modelling where the new organic soil dielectric constant model is playing an active role. Figures 7 and 8 refer to the vegetation opacity.

The following maps were obtained by using almost the entire 11 years of the v700 reprocessed campaign data and the associated v650.



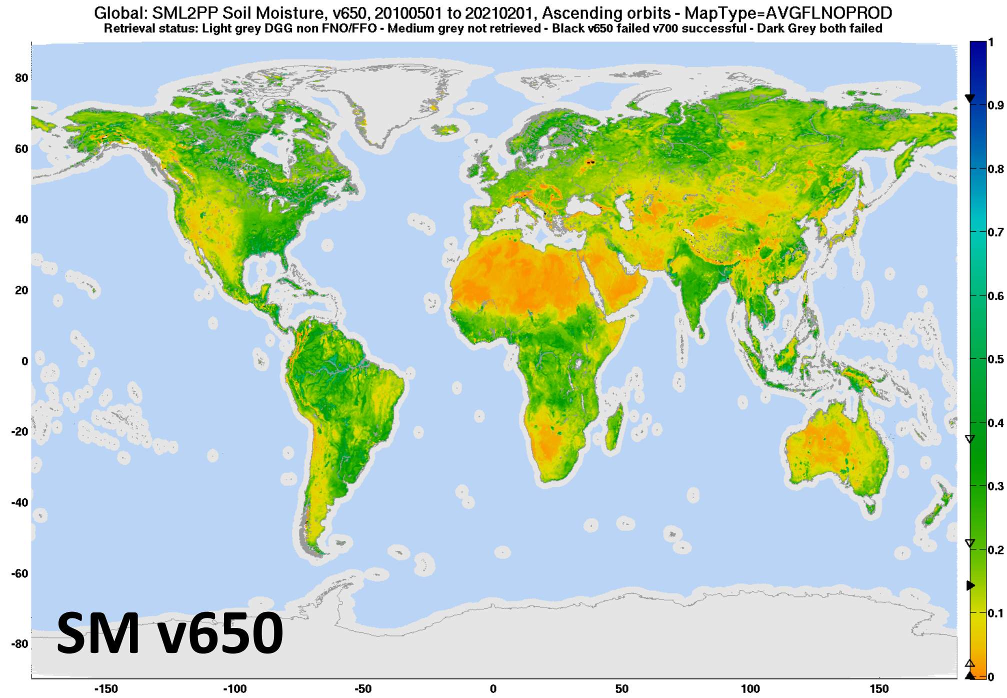


Figure 4: Averaged soil moisture over almost 11 years from 2010501 to 20210201; top v700, bottom v650.

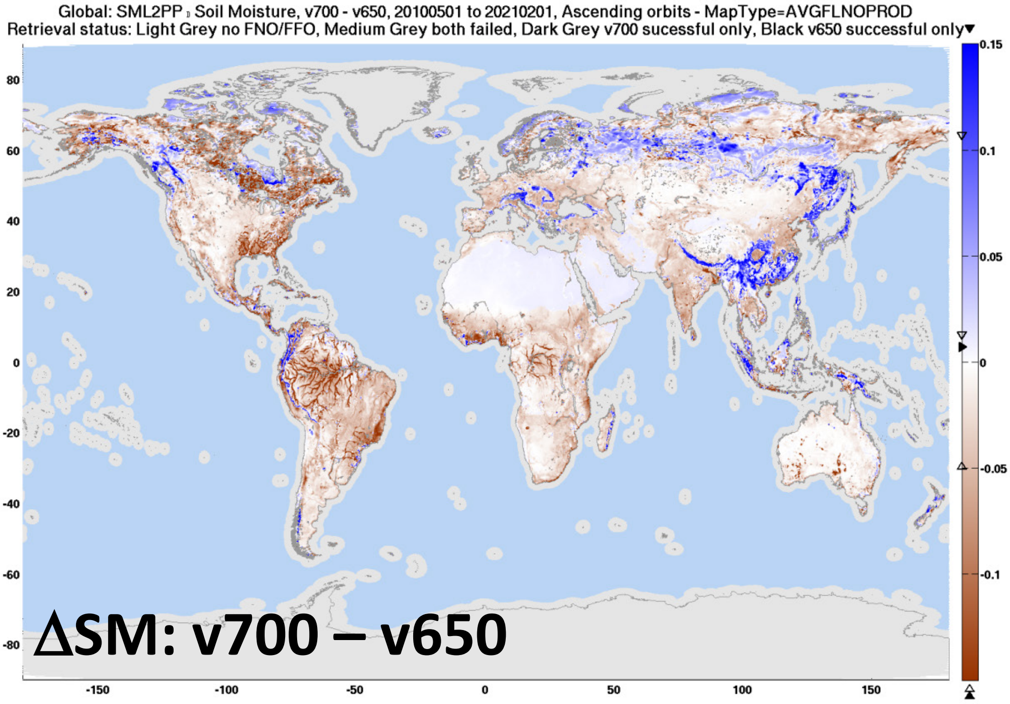
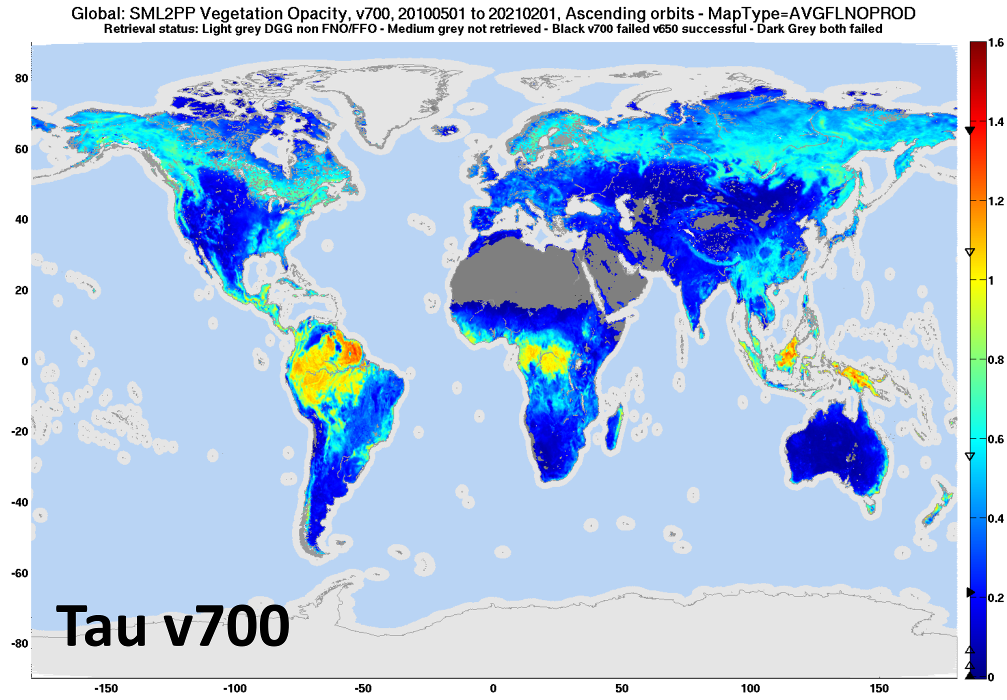


Figure 5: Difference (v700-v650) of averaged soil moisture over almost 11 years from 2010501 to 20210201.



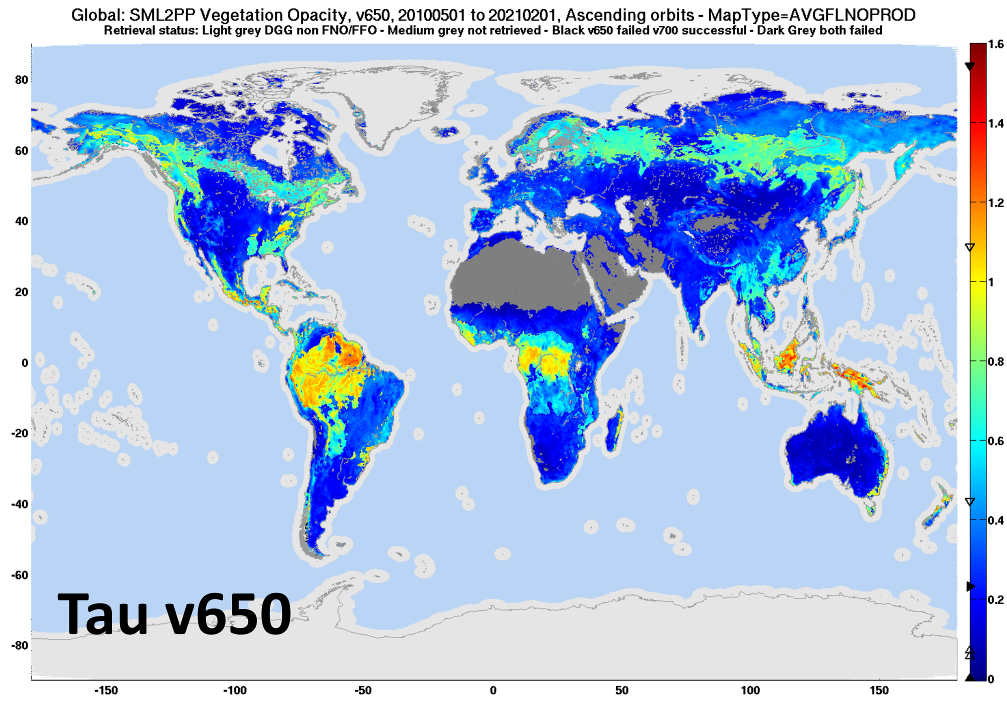


Figure 6: Averaged vegetation opacity over almost 11 years from 2010501 to 20210201; top v700, bottom v650

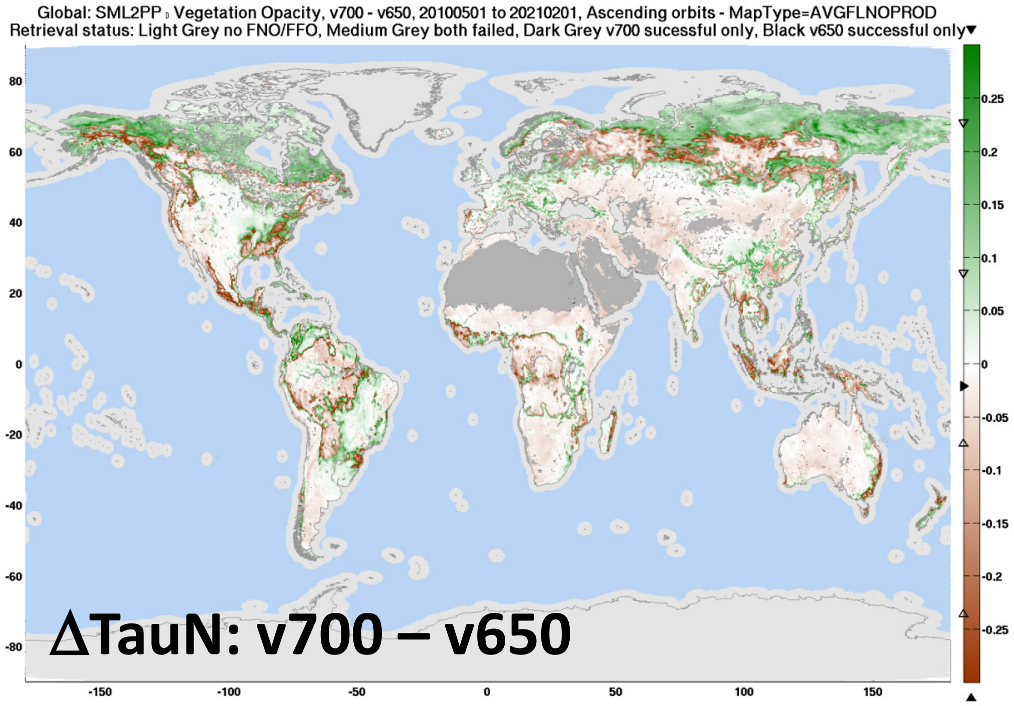


Figure 7; Difference (v700-V650) of averaged vegetation opacity over almost 11 years from 2010501 to 20210201.

3.2 Caveats (please read)

To help the user to understand the soil moisture product content, some basic information about the algorithm decision trees, the used radiative transfer model and the input L1C brightness temperature data are given below. Further algorithm details are illustrated in the soil moisture retrieval Algorithm Theoretical Basis Document (ATBD).

- **Retrieved TauNadir**. The TauNadir product, the vegetation opacity, more often called these recent years as the L-VOD (L-band vegetation optical depth) is a retrieved parameter that became more and more important for its connexion with biomass and thus with carbon stock. It is important to remind that the SMOS retrieved VOD is actually an L-band radiative quantity. Recent studies showed that L-VOD was very temperature-dependent at low temperatures. Indeed, since soil moisture it is essentially sensitive to the dielectric constant of the free (unbound) water content the vegetation contains. The dielectric constant of water itself is mainly dependent of the water temperature and its salinity (ionic conductivity). The main driver still remains the vegetation water content (VWC); the higher the VWC, the higher the L-VOD. However, a totally frozen forest would behave as a dry material with no free water and would have a L-VOD close to 0 while keeping its biomass as it was at warmer season.

  For an appropriate use of SMOS retrieved VOD for biomass/carbon studies such effects have to be taken into consideration. It is particularly important for Boreal forests that can display a strong L-VOD seasonality for areas where the vegetation temperature can drop far below the water freezing point.

Important point to keep in mind: L-VOD is not biomass, but both are related through the capability of a particular biomass to hold (free) water.

- **Retrievals**. In our algorithms we consider the retrievals as successful when several criteria are met: convergences of the optimizer, retrieved values are within an accepted range as well their post-retrieval uncertainties. However, the level of acceptance/rejection of retrievals depends on the application. For instance, people interested in extreme events may want more data accepting more inaccuracies than probably people wanting to assimilate those parameters (who may want less but more reliable data). We chose to set thresholds to maximise the number of retrievals we think are reliable enough.

  There are a few indicators that can be used to select subset with extra quality but at the expense of the number of data.

  - With the inclusion of added observation error in the cost function, as it was the case for the v650 but using a Chi2 post-rescaling strategy, the probability of Chi2 is a reliable information for extra quality selection. The associated flag FL_Chi2_P along with the FL_NO_PROD flag for filtering data is still a good use:

    - FL_NO_PROD = 0 and FL_Chi2_P = 0 (Chi2_P≥5%) – successful retrieval with good fit quality (but less data)

    - FL_NO_PROD = 0 and FL_Chi2_P = 1 (Chi2_P<5%) – successful retrieval with degraded fit quality (but more data)

    - FL_NO_PROD = 1 – rejected retrieval as usual.

  - In complement RFI filtering can be hardened. It is already part of the algorithms at the very beginning of the process by removing clearly contaminated brightness temperatures. However, it is advised to consider the RFI probability (RFI_Prob) computed for the past 12 days as extra criterion to define and optimize appropriate rejection level depending on the final use of the retrieved products.

  The «FL_NOPROD» is the main flag informing on successful retrievals; it can be complemented usefully by using the «FL_Chi2P» flag or by using specific «Chi2_P» values threshold and by using «RFI_Prob» values threshold for extra quality selection.

- **Primary decision tree**. For each node of the SMOS grid, the concerned area (the working area is a 123 km x 123 km box when including the minor contributions) is described as including several fractions: e.g., vegetated soil, ice, forest, open water. Then, the first decision tree (17 branches) selects which of the fraction(s) will be considered for the retrieval. This depends on the respective weights of the various land use types, including also consideration on topography. Let us for example assume that part of the scene consists of vegetated soil while another fraction consists of open sea. Since the sea fraction is not relevant for retrieving soil moisture, the radiometric contribution due to sea is computed using auxiliary data and it is considered as a default contribution. This is only an example; indeed, the forward modelled brightness temperatures used in the retrieval will integrate most of the time default contributions.

Note that the retrieved soil moisture corresponds only to the area where the dominant land use is present.

- **Radiative models**. There are basically 3 radiative models in the L2 algorithm, depending on how the dielectric constant of the surface is computed. They can be used either in the retrieval iterative loop, or simply in order to build default contributions.

  The nominal model is the standard soil vegetation radiative transfer model and includes soil moisture (based on L-MEB, L-band Microwave emission of the Biosphere);

  The water surface model is used for the sea in coastal pixels, wetland and lakes;

  The so-called cardioid model is used for retrieval whenever it can only aim at providing information about the dielectric constant itself (e.g., ice, barren surfaces).

  The nominal model will be used over vegetated soil and forest. While these cases are the only ones of direct interest as far as soil moisture is concerned, it may be mentioned that preliminary results using the cardioid model suggest there is indeed physical meaning in them.

  It is important to note that a value of -999 for a geophysical parameter and its associated DQX in the L2SM UDP product implies that either the retrieval for that parameter failed or was not attempted. When all retrieval attempts fail for a node, the «FL_NOPROD» is set to 1, otherwise it is set to 0.

- **Secondary decision tree.** Depending on the content of the working area, one of 3 radiative models is used for retrieval. For each of them, one must then define which parameters are to be retrieved, and what are the constraints assigned to the initial values. The secondary decision tree lists (for each of the 3 models) 3 options depending on the expected vegetation optical thickness and 3 options depending on the "information richness" expected from the data, which is estimated from the incidence angle coverage. As expected, the commissioning phase evidenced that this scheme with 27 options was too complicated and it was meant to be simplified eventually. Currently there are only 2 retrieved parameters for instance.

  Note that the SMOS SM products are not produced in the way of many other similar products, so read carefully the ATBD to see exactly what is being done (as it is not always standard practice) so as to take advantage of SMOS characteristics.

The user should also consider the following remarks on the usage of the Level 1 brightness temperature in the retrieval of the soil moisture parameter:

- **SMOS L1 product (L1C).** The L1C consists of brightness temperatures reconstructed from interferometric data in the reference frame of the SMOS antenna plane. Hence these radiometric data are associated to upwelling Stokes parameters through a transformation combining the Faraday effect and geometrical factors.

SMOS brightness temperatures are NOT TBH, TBV, third or fourth Stokes parameters and CANNOT be compared directly with Earth surface observation or modelling.

- **Soil moisture retrieval efficiency**. The SMOS soil moisture retrieval is based on matching measured and modelled (surface emission) brightness temperatures, with the modelled values varying as a function of the incidence angle and depending on soil moisture as well as other physical parameters. There is definitely, for a very large fraction of nodes of the SMOS grid, a robust ability of the radiative model to match the angular signature of brightness temperature while producing realistic values of retrieved parameters.

- **Polarisation mode**. According to the End-of-Commissioning review decisions, the full polarization acquisition mode has been selected for the operational phase, accounting for the potential information provided by this mode. However, until full understanding of third and fourth Stokes parameters has been achieved, "pseudo dual pol" is used for retrievals over land surfaces; i.e., only the antenna level brightness temperatures corresponding to first and second Stokes parameters will be used in the retrieval. Note that third and fourth Stokes are much improved with respect to initial releases but not yet totally up to our expectations. The Level 2 algorithm does not yet use the cross-polarisation terms in the retrieval process.

  Note that brightness temperature polarisations are always given in the antenna and can only be transformed into ground values (i.e., H and V) through a transformation related to the instantaneous view angles. Only points located on the satellite subtrack are (almost) in H and V. A Matlab tool is provided on the blog (http://www.cesbio.ups-tlse.fr/SMOS_blog/) to perform this task.

- **Time reference.** We remind that the acquisition time all SMOS products is provided in EE CFI Transport Time format and the reference time in all SMOS products is MJD2000. The MJD2000 counts the time elapsed since 2010/01/01 **00:00:00** UTC.

  Take care MJD2000 is not J2000 standard reference time that starts at 2010/01/01 12:00:00 UTC.

4. Planned algorithm evolution

There are different areas where progress is ongoing:

- VOD under focus: studying the two-stream emission model as an alternative to tau-omega for the radiative transfer component (vegetation) of our modelling.

- We will continue improving the algorithm with respect to the parametrisation of the surface. We have significantly gained on SM time dynamics at least observable against in-situ but still with a general bias increase that needs to be addressed.

- Opening to possible new products related to cryosphere. Liquid water in snow and ice, ice cap internal temperature, freeze/thaw, temperature underneath snow cover etc.

- Future generation of L1C will introduce active signal RFI filtering that if working well will help in improving the retrieval quality for certain areas of the Earth.

Some caveats were already identified, as detailed above, and are under investigation.

Version 700 of the soil moisture Level 2 processor, which includes substantial corrections and improvements as a result of the evolution work done since the Commissioning Phase, is now delivering the best soil moisture and L-VOD products available. Nevertheless, we believe there is still room for improvements!

The data is now yours to evaluate. Please let us know of any issues encountered (on scientific points!) so that we can try to fix them for the next – general - release. We are looking forward to collaborating with you on these topics and make SMOS products even better.

Yann, Philippe and Philippe, Ali and Ali,

Ahmad, Anna, Arnaud, Beatriz, Cecilia, Cristina, Emma, Eric, François,

Jean Pierre, Jennifer, Juha, Klaus, Kimmo, Manu, Manuel, Marie, Mike, Nemesio,

Paolo,Rachid, Raffaele, Rajesh, Roberto, Simon, Simone, Steven, Susanne, Tim.