Read-me-first note for the release of the SMOS Level 2 Soil Moisture data products	
Processor version	Level 2 Soil Moisture V620
Release date by ESA	Release 1: 7 May 2015, valid only for the operational data set. Release 2: 2 March 2016, valid for both the operational and reprocessed data set. Release 3: 21 April 2017, valid for operational data: NOAA Interactive Multisensor Snow and Ice Mapping System (IMS) is used to represent the snow cover extent.
Authors	Expert Support Laboratory Level 2 Soil Moisture + Array Systems Computing Inc.
Further information	Details on the processing algorithms can be found in the Algorithm Theoretical Baseline Document (ATBD) ▪ SO-TN-ARR-L2PP-0037 v3.9 available here: https://earth.esa.int/web/guest/-/data-types-levels-formats-7631 Information about the L2 soil moisture products structure can be found in the SMOS Level 2 and Auxiliary Data Products Specifications document: ▪ SO-TN-IDR-GS-0006 v8.1 available here: https://earth.esa.int/web/guest/-/data-types-levels-formats-7631 Information on how to access the SMOS data can be found here: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/smos/content/-/asset_publisher/t5Py/content/how-to-obtain-data-7329 Additional information, including change history of L2SM algorithm and associated processor, documentation can be found on the SMOS L2SM CESBIO website: www.cesbio.ups-tlse.fr/SMOS_blog/ and ARRAY website: www.array.ca/smos
Contact for helpline	For all issues related to data access, formats and read/write, processors, please contact ESA's HelpDesk at 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, rajesh@array.camailto:) or through the BLOG (http://www.cesbio.ups-tlse.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 the version 620 of the Level 2SM Operational Processor (L2SM OP).

The version 620 of the Level 2 Soil Moisture data product is now available for the entire SMOS mission life time with the following file class and version:

File class	File version	From	To
REPR	V620	1 June 2010	14 July 2015
OPER	V620	15 July 2015	present

The data set acquired during the SMOS mission **commissioning phase** (from January 2010 to 31 May 2010) has been acquired during periods when the MIRAS instrument underwent several tests and was operated in different modes causing drifts not fully compensated by the on-ground calibration processing. For that reason, this data set is only available upon request and should not be used for long term data exploitation.

The SMOS data users are invited to use this new data set, which supersedes the previous one generated by the algorithm baseline version 551 and to read this note carefully to ensure optimum exploitation of the version 620 data set. Further information on the quality of the data set can be found in the reprocessing QC report available here: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/smos/content/-/asset_publisher/t5Py/content/data-quality-7059)

This note is organised as follows:

- Main improvements in the L2SM version 620 data set
- L2SM version 620 data set performance and caveats
- Future algorithm evolution

2. Main improvements in the (current) L2SM version 620 data set

The major improvements introduced in the current operational version 620 of the SMOS Level 2 soil moisture products are the following:

- Significant improvements in the quality of the input Level 1 brightness temperature version 620, leading to a larger ratio of successful soil moisture retrievals with improved accuracy.

- Better management in the L2SM processor version 620 of the so-called Current files[1] by separating these Current files between ascending and descending orbits, as

[1] Current files are generated by the L2SM post-processor and provide initialization parameters at the level of the Discrete Global Grid (DGG) point used for the soil moisture retrieval. The parameters provided by the current files are: optical thickness for low and dense vegetation, counters about RFI contamination, surface roughness and flood probability.

vegetation water content, for instance, can be expected to change more between morning and afternoon than from day to day. This has been done for all Current files. This improvement led to several configuration changes as described below:

a. Splitting in ascending (ASC) and descending (DSC) orbits reduces the number of acquisitions available for each DGG point. As a consequence, the ASC- DSC separation had to be associated with an increase in the temporal validity threshold, to allow gathering enough measurements. For low vegetation, the validity threshold was brought to 10 day and was left unchanged for the others: Forest and Roughness.

b. Retrievals are attempted only when the range of available angles spans over more than 20°

- The value in the Current file is only updated if the χ^2 is improved; previously the smallest Data Quality Index was selected, which favoured low opacity constraint and thus low soil moisture values as by construction and all other things equivalent, the DQX is "better" for low soil moisture values (see ATBD https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/smos/content/-/asset_publisher/t5Py/content/data-processors-7632)

- Significant improvements in the detection algorithm for Radio Frequency Interface (RFI) for both ascending and descending orbits. The new algorithm tolerance is stricter, so more RFI sources with medium level intensity are flagged (below the geophysical maxima of the scene) as illustrated in Figure 1. Further information on the RFI flagging algorithm used in the L2SM processor is present in the ATBD available here: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/smos/content/-/asset_publisher/t5Py/content/data-processors-7632

- The Data Quality IndeX (DQX) of the retrieval, instead of the radiometric accuracy of measurements, is increased when RFIs are present. As a consequence, the χ^2 increases over contaminated areas. The number of outliers detected is thus increased.

- The retrievals underneath forest have been improved by a new parameterisation and by increasing the number of iteration (now raised to 30) before stopping the retrieval loop.

- The thresholds for soil moisture and opacity acceptance have been moved from DQX SM = 0.1 to 0.2 for soil moisture and from DQX tau =0.2 to 0.4 for opacity. It enables to have more valid retrievals under dense vegetation (forests).

- It was found that the land use map overestimated water bodies in some tropical areas (Congo basin, notably) leading to "no retrievals" for the soil moisture. We corrected the land use map using the latest version of ESA's GlobCover V2.3 and updating classes corresponding to water bodies.

- An improvement to the algorithm, leading to regularisation of the behaviour for soil moisture near 0, was implemented. It consisted in a normal regularisation to avoid asymmetry and non-continuous behaviour for the first and second derivative at the origin.

- The content of the L2 soil moisture user product (L2SM UDP) was slightly modified with a new field: X-Swath (abscissa of the dwell line) was made available and the RFI probability is now computed over a 12-day moving window by using two AUX_DGGRFI files in input to the L2SM processor.

Several auxiliary data files were updated or additions were made:

- The soil texture map is now at the 4km grid scale instead of 10km and is now based on the same source used by JPL's SMAP mission.
- The LAI and maximum LAI phenology maps have been updated in association with the new forest parameterisation.
- The galaxy map (deep sky downward radiation) has been updated to correct some anomalies in the HI line with marginal impact on soil moisture retrievals.
- Since 19 December 2016, auxiliary dataset from the NOAA Interactive Multisensor Snow and Ice Mapping System (IMS) is used by the operational SMOS soil moisture level 2 processor to represent the snow cover extent. The NOAA IMS, based on data acquired by various sensors, provides a better representation of the snow cover compared to previously snow density information from ECMWF forecast. Globally the impact in terms of soil moisture is not significant.

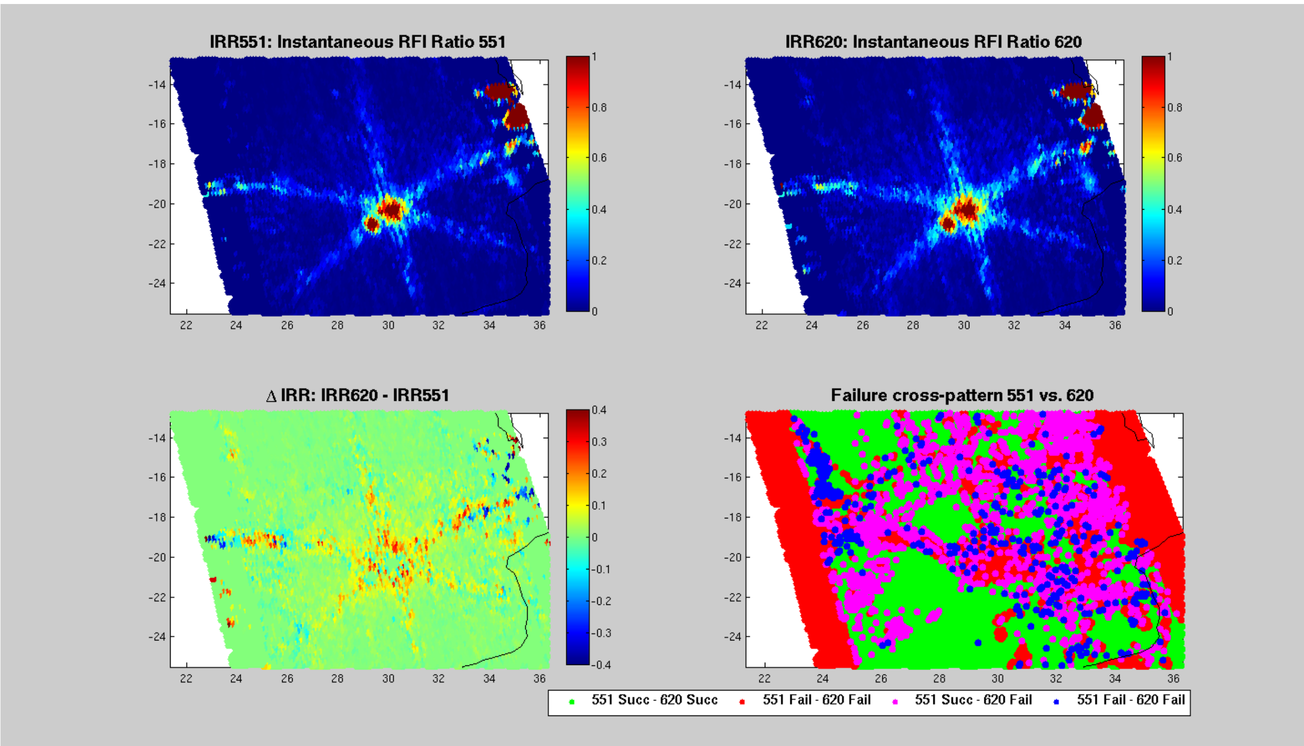
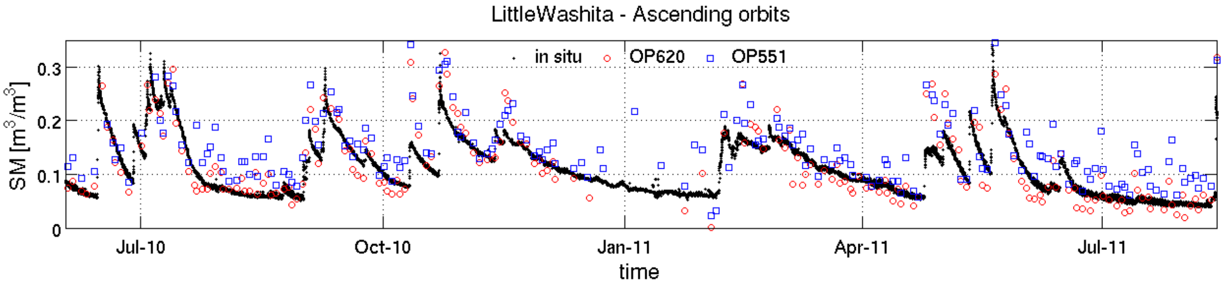


Figure 1: Differences in RFI Flagging and soil moisture retrieval between L2SM processor version 551 (previous) and version 620 (current). Upper panels show the Instantaneous RFI Ratio (IRR) for the previous version 551 (left) and the current version 620 (right) of the L2 soil moisture processor. The IRR indicates the probability of RFI contamination over the DGG point. Lower left panel shows the differences between the two IRR. Positive difference indicates improved RFI detection for the version current 620. Lower right panel shows the difference in successful soil moisture retrieval between version 551 (previous) and the current version 620. The large number of DGGs coloured in violet shows the area where the previous processor version 551 was providing degraded

measurements while the current processor version 620 does not retrieve soil moisture due to strong presences of RFI (P. Richaume - Cesbio).

3. L2SM version 620 data set performances and caveats[2]

The Figures below show the performance assessments by comparing version 551 and version 620 of the L2SM products over the USDA Watershed (T. J. Jackson and R. Bindlish – USDA) for a case with the best improvement in the retrieved soil moisture (Figure 2) and for a case with the least improvement (Figure 3).



Series	ρ	μ	σ	RMSE	#kept
OP620	0.90	-0.008	0.035	0.036	200
OP551	0.81	-0.040	0.042	0.058	206

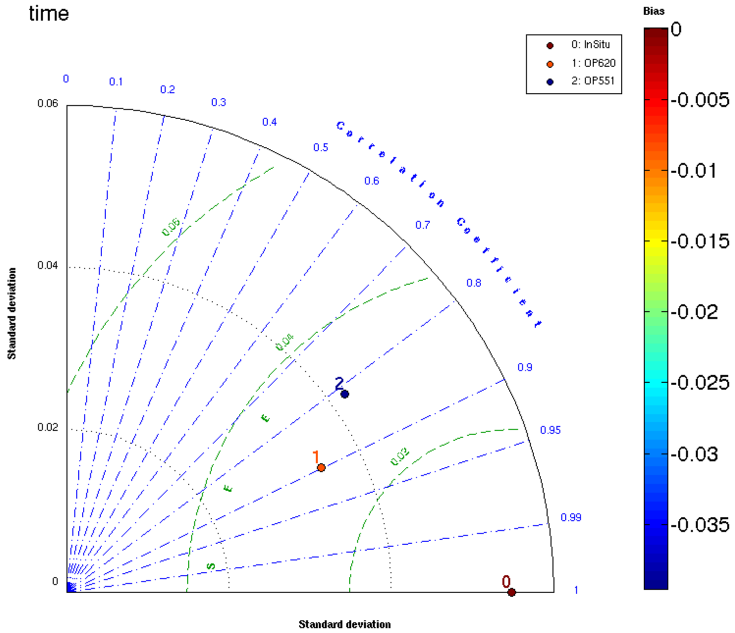
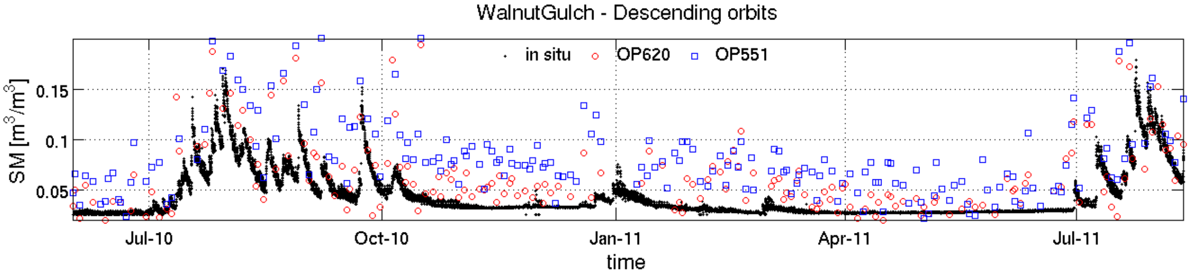


Figure 2 Evolution of retrievals between version 551 and version 620 - best case

---

[2] The products released are prone to be improved. Feedback to yann.kerr@cesbio.cnes.fr and rajesh@array.ca on any issues you may identify while using or validating the data is most welcome.



Series	ρ	μ	σ	RMSE	#kept
OP620	0.73	-0.024	0.038	0.045	199
OP551	0.77	-0.046	0.040	0.061	209

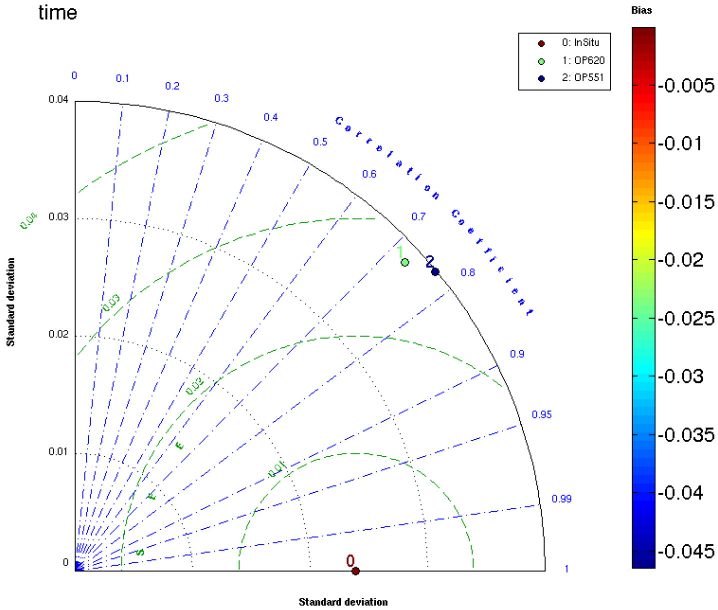


Figure 3: Evolution of retrievals between version 551 and version 620 - worst case (only case of degradation between versions 551 and 620)

The caveats identified in the data set version V620 are the following:

- The quality figures DQX (theoretical retrieval uncertainty) and GQX (overall quality) can only be considered as indicative so far. These quality figures do not reflect the poor quality of the fit (see possible explanations for this below) resulting in large χ^2 when compared to the radiometric uncertainty of individual brightness temperature observations. Uncertainties vary over the whole angular range of observed brightness temperatures which is difficult to quantify as a whole.

- Concerning the radiative nominal model, some deficiencies have been detected. The high sensitivity of the model for very small soil moisture values produces unrealistically good DQX performances in these regions.

- Vegetation opacities are improved but still not fully compatible with what we would expect, as the seasonal variations remain small with respect to high frequency variations. Note that the products cannot be compared directly with optical sensors vegetation index like NDVI.

The above L2 SM product version 620 performances and caveats have to be understood in the context of the L2 retrieval algorithm which is somewhat complicated. This is due to the land surface being very inhomogeneous. Remember also that it is difficult to perform an accurate estimation of soil moisture from modelling or in situ data at the scale of SMOS. Moreover, SMOS senses soil moisture over a top soil layer depth which is not exactly the one considered in modelling or experimental data sets. So careful attention is required when interpreting discrepancies/agreements between SMOS retrievals and other soil moisture data sets (*in situ,* model or satellite).

To help the user in the understanding of the soil moisture product, 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) available here (https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/smos/content/-/asset_publisher/t5Py/content/data-processors-7632).

- **Primary decision tree**. For each node of the SMOS grid, the concerned area (the working area: 123 km x 123 km box when including the minor contributions) is described as including several fractions: vegetated soil, ice, forest, open water and the like. 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.
  a) The nominal model is the standard soil vegetation radiative transfer model and includes soil moisture (L-MEB, L-band Microwave emission of the Biosphere);
  b) The water surface model is used for the sea in coastal pixels, wetland and lakes
  c) 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 as are 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 Stokes or fourth Stokes 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

4. Further algorithm evolution

Two main areas where progress is needed concern the correction of biases and the detection/mitigation of spurious signals:

- **Biases.** Beyond the average biases which are brought down to satisfactory levels through calibration procedures, there are still imperfectly corrected biases which depend on the location within the instantaneous field of view. Such errors may be generated during the reconstruction process in case the detailed instrument properties (e.g. element antenna pattern) are not perfectly known. Several of these problems (across swath spurious trends, pixel biases, so called land-sea contamination) have been detected, identified, and sometimes empirically mitigated by the ocean team; although these errors are less prominent over lands due to the much larger sensitivity of brightness temperatures to physical properties, they are present all the same. A clue to their presence is that the quality of fit is not matching expectations, as illustrated by the fact that the normalized cost function (χ^2-CHI square) is substantially higher than the theoretically expected value of 1.
- **Spurious signals.** On one hand, no significant impact of sun contamination (glint) has been detected, so far, over land but work is in progress. On the other hand, spurious signals due to RFI are much larger and more frequent than expected when knowing that the SMOS bandwidth is - in theory - fully protected. The most efficient mitigation technique consists in requesting trespassers to stop transmitting in the protected 1400 - 1427 MHz band. Indeed this is under way and brings encouraging results. Meanwhile, flagging techniques are explored. Several detection algorithms are being used in the L2 codes. The most robust have been used in version 620 to build maps which depict, for each SMOS grid node, the measured frequency of occurrence of RFI detection over a sizeable SMOS observation period. Studies are still under-way to improve RFI detection, localisation and flagging. But there is still room for improvements, even though version 620 provides a much more accurate RFI flagging.

Some caveats are already identified in the corresponding Section of this note and they are also under investigation.

Version 620 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 products available. The availability of the version 620 reprocessed data set provides a consistent SMOS soil moisture data set spanning almost 6 years. Nevertheless, we believe there is still room for improvements! The data are now yours to be evaluated. Please let us know the 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, Arnaud, Beatriz, Cecilia, Cristina, Delphine, Eric, François

Jean-Claude, Jean Pierre, Jennifer, Marie, Mike, Nemesio, Paolo, Rachid,

Raffaele, Rajesh, Steven, Susanne, Roberto, Tim.