Soil Moisture Active Passive (SMAP) Project

Algorithm Theoretical Basis Document

SMAP-Sentinel L2 Radar/Radiometer Soil Moisture (Active/Passive) Data Products: L2_SM_SP

PREPARED BY:

10-31-2019
Narendra N. Das, JPL   Date
SMAP Algorithm Development Team Lead for
L2_SM_SP Data Products

Dara Entekhabi, MIT
SMAP Science Team Lead
SMAP Mission

APPROVED BY:

12-24-2019
Dara Entekhabi, MIT   Date
SMAP Science Team Lead

Yueh, Simon H (3290)   Digitally signed by Yueh, Simon H (3290)
DN: cn=Yueh, Simon H (3290), o, ou,
email=simon.yueh@jpl.nasa.gov, c=US
Date: 2019.12.23 23:56:04 -08'00'

12-23-2019
Simon Yueh, JPL   Date
SMAP Project Scientist

Dec 24, 2019
JPL D-104870

JPL

Jet Propulsion Laboratory
California Institute of Technology

DOCUMENT LOG DATE

REVISION	DATE	SECTIONS CHANGED	REASON FOR CHANGE
Version 1	Sep 1, 2019	N/A	N/A
Version 2	Oct 31, 2019	New Results added	Revision
Version 3	Dec 24, 2019	Edited after review	Reviewed by SMAP Science Team

Soil Moisture Active Passive (SMAP)

Algorithm Theoretical Basis Document
SMAP-Sentinel L2 Radar/Radiometer
(Active/Passive) Soil Moisture Data Products

Release, v.3

Narendra N. Das
Jet Propulsion Laboratory
California Institute of Technology, Pasadena, CA

Dara Entekhabi
Massachusetts Institute of Technology
Cambridge, MA

Scott Dunbar, Andreas Colliander, Mario Chaubell, Simon Yueh
Jet Propulsion Laboratory
California Institute of Technology, Pasadena, CA

Thomas Jagdhuber
Microwaves and Radar Institute
German Aerospace Center
Oberpfaffenhofen, Germany

Peggy E. O'Neill
NASA, Goddard Space Flight Center
Greenbelt, MD

Wade Crow, Fan Chen
HRSL, USDA

Jet Propulsion Laboratory
California Institute of Technology

Algorithm Theoretical Basis Documents (ATBDs) provide the physical and mathematical descriptions of the algorithms used in the generation of science data products. The ATBDs include a description of variance and uncertainty estimates and considerations of calibration and validation, exception control and diagnostics. Internal and external data flows are also described.

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

AMSR	Advanced Microwave Scanning Radiometer
ASRIS	Australian Soil Resource Information System
ATBD	Algorithm Theoretical Basis Document
CLASIC	Cloud and Land Surface Interaction Campaign
CONUS	Continental United States
COSMOS	Community Earth System Models
CRN	Climate Reference Network
DEM	Digital Elevation Model
EASE2	Equal Area Scalable Earth
ECMWF	European Center for Medium-Range Weather Forecasting
FAO	Food and Agricultural Organization
GMAO	Goddard Modeling and Assimilation Office
GPCP	Global Precipitation Climatology Project
GPS	Global Positioning System
GRUMP	Global Rural Urban Mapping Project
GSFC	Goddard Space Flight Center
HWSD	Harmonized World Soil Database
IGBP	International Geosphere-Biosphere Program
JPL	Jet Propulsion Laboratory
LSM	Land Surface Model
LIS	Land Information System
NOAA	National Oceanic and Atmospheric Administration
NSDC	National Soil Database Canada
NSIDC	National Snow and Ice Data Center

PALS	Passive and Active L-band System
PGE	Product Generation Executable
QC	Quality Control
RFI	Radio Frequency Interference
RMSE	Root Mean Square Error
RSS	Root Sum of Squares
RVI	Radar Vegetation Index
SCA	Single Channel Algorithm
SCAN	Soil Climate Analysis Network
SDS	Science Data System
SGP	Southern Great Plain (field campaigns)
SMAP	Soil Moisture Active Passive
SMAPVEX	SMAP Validation EXperiment
SMEX	Soil Moisture Experiment
SMOS	Soil Moisture Ocean Salinity (ESA space mission)
SRTM	Shuttle Radar Topography Mission
STATSGO	State Soil Geographic Database
TBC	To Be Confirmed
TBD	To Be Determined
USDA	United States Department of Agriculture
UTC	Coordinated Universal Time
ubRMSE	Unbiased Root Mean Square Error
VWC	Vegetation Water Content

1   INTRODUCTION

This document is the Algorithm Theoretical Basis Document (ATBD) for the surface soil moisture data product created by combining the SMAP L-band (~1.4 GHz) radiometer and the Copernicus Sentinel-1A/1B C-band (~5.405 GHz) radar observations, respectively.

SMAP Requirement: Baseline Science Mission

a) *The baseline science mission shall provide estimates of soil moisture in the top 5 cm of soil with an error of no greater than 0.04 [m_3 m_{-3}] volumetric (one sigma) at 10 km spatial resolution and 3-day average intervals over the global land area excluding regions of snow and ice, frozen ground, mountainous topography, open water, urban areas, and vegetation with water content greater than 5 [kg m_{-2}] (averaged over the spatial resolution scale);*

The SMAP-Sentinel active-passive soil moisture product (L2_SM_SP) is not mandated to meet the SMAP Baseline Science Mission requirement. However, a best effort will be made for the L2_SM_SP product to meet the accuracy requirement at 9 km spatial resolution. This document provides the theoretical basis and technical description of the L2_SM_SP product.

1.1   Overview and Background

The important role of surface soil moisture as a terrestrial hydrology state variable is well recognized. Various applications like weather forecasting, climate change prediction, agricultural production, water resources management, drought prediction, flood area mapping, and ecosystem health monitoring require information on surface soil moisture for skillful modeling and forecasting. The outcomes from these applications have direct impact on human society and the management of our environment. Therefore, mapping surface soil moisture with sufficient accuracy over the required ranges of spatial and temporal scales is imperative to fulfill the needs of these applications.

Surface soil moisture can be measured over a range of scales from point scale (*in situ*) to coarse scale at various temporal resolutions. At point scale soil moisture measurements are conducted using *in situ* measurement networks (e.g., SCAN sites and Oklahoma Mesonet in the continental United States) that can have high accuracy but are spatially very sparse. Coarse scale (> 40 km) soil moisture measurements are obtained from satellite-based footprints using L-, C- and X-band radiometers (e.g., SMOS, AMSR-E, and WindSat) [1-2]. The satellite-based C- and X-band radiometers have shallow sensing depth (< 2 cm) and also have significantly reduced sensitivity to soil moisture for even small amounts of vegetation, leading to high retrieval errors in soil moisture estimates over vegetated regions [1]. Satellite-based C-band radars such as the ERS scatterometer also have coarse resolution (~50 km) and have been used to retrieve surface soil moisture over sparsely vegetated regions with moderate accuracy. The European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) mission, launched in November, 2009, is the first wide-swath L-band soil moisture mission, and has retrieved soil moisture over a much higher range of vegetation conditions at a spatial resolution of ~40 km with a sensing depth of ~5 cm [2], an improvement over the C-band radiometers of AMSR-E and WindSat.

All these measurement technologies only partially satisfy the required criteria of high spatial and temporal resolution, wide spatial coverage, optimal sensing depth and desired accuracy in retrieved soil moisture over moderate vegetation conditions. Surface soil moisture retrieved using

these approaches is not suitable for hydrometeorology, ecology, water resource management, and agronomy because these applications require high spatial (< 10 km) and temporal (< 3 days) resolution soil moisture information. Recognizing the importance of fine spatial and temporal resolution surface soil moisture measurements with global coverage, the Soil Moisture Active Passive (SMAP) mission focused on producing a high spatial resolution (<=9 km) soil moisture using the Active-Passive (radar + radiometer) approach/algorithm that benefits the fields of hydrology, meteorology and ecology for potential geophysical applications and scientific advances.

1.2   The Soil Moisture Active Passive (SMAP) Mission

1.2.1   Science Objectives

The National Research Council's (NRC) Decadal Survey, *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, was released in 2007 after a two year study commissioned by NASA, NOAA, and USGS to provide them with prioritized recommendations for space-based Earth observation programs [3]. Factors including scientific value, societal benefit and technical maturity of mission concepts were considered as criteria. The NRC recommended SMAP data products that have high science value and provide data towards improving many natural hazards applications. Furthermore, SMAP draws on the significant design and risk-reduction heritage of the Hydrosphere State (Hydros) mission [4]. For these reasons, the NRC report placed SMAP in the first tier of missions in its survey. In 2008 NASA announced the formation of the SMAP project as a joint effort of NASA's Jet Propulsion Laboratory (JPL) and Goddard Space Flight Center (GSFC), with project management responsibilities at JPL. The target launch date was late 2014 [5]. SMAP actually launched on January 31, 2015.

The SMAP science and applications objectives are to:

- •   Understand processes that link the terrestrial water, energy and carbon cycles;
- •   Estimate global water and energy fluxes at the land surface;
- •   Quantify net carbon flux in boreal landscapes;
- •   Enhance weather and climate forecast skill;
- •   Develop improved flood prediction and drought monitoring capability;
- •   Improve the agricultural applications.

1.2.2   Measurement Approach

Table 1 is a summary of the SMAP instrument functional requirements derived from its science measurement needs. The goal is to combine the attributes of the radar and radiometer observations (in terms of their spatial resolution and sensitivity to soil moisture, surface roughness, and vegetation) to estimate soil moisture at a resolution of 10 km, and freeze-thaw state at a resolution of 1-3 km. An active-passive (radar + radiometer) approach/algorithm [5] was considered to provide high spatiotemporal resolution soil moisture product to meet the SMAP science and applications objectives.

The SMAP instrument originally incorporated an L-band radar and an L-band radiometer that shared a single feedhorn and parabolic mesh reflector. As shown in Figure 1, the reflector is offset from nadir and rotates about the nadir axis at 14.6 rpm (nominal), providing a conically scanning antenna beam with a surface incidence angle of approximately 40°. The provision of constant incidence angle across the swath simplifies the data processing and enables accurate repeat-pass estimation of soil moisture and freeze/thaw change. The reflector has a diameter of 6 m, providing a

‚Oct 31, 2019

radiometer 3 dB antenna footprint of 40 km (root-ellipsoidal-area). The real-aperture radar footprint was 30 km, defined by the two-way antenna beamwidth. SMAP real-aperture radar and radiometer data were planned to be collected globally during both ascending and descending passes.

To obtain the desired high spatial resolution the radar employs range and Doppler discrimination. The radar data can be processed to yield resolution enhancement to 1-3 km spatial resolution over the 70% outer parts of the 1000 km swath. Data volume prohibits the downlink of the entire radar data acquisition. Radar measurements that allow high-resolution processing will be collected during the morning overpass over all land regions and extending one swath width over the surrounding oceans. During the evening overpass data poleward of 45° N will be collected and processed as well to support robust detection of landscape freeze/thaw transitions.

The baseline orbit parameters are:
- Orbit Altitude: 685 km (2-3 days average revisit and 8-days exact repeat)
- Inclination: 98 degrees, sun-synchronous
- Local Time of Ascending Node: 6 pm

Table 1. SMAP Mission Requirements

Scientific Measurement Requirements	Instrument Functional Requirements
Soil Moisture: ~±0.04 m_3m_{-3} volumetric accuracy(1-sigma) in the top 5 cm for vegetation water content ≤ 5 kg m_{-2}; Hydrometeorology at ~10 km resolution; Hydroclimatology at ~40 km resolution	L-Band Radiometer (1.41 GHz): Polarization: V, H, T_3 and T_4 Resolution: 40 km Radiometric Uncertainty*: 1.3 K L-Band Radar (1.26 and 1.29 GHz): Polarization: VV, HH, HV (or VH) Resolution: 10 km Relative accuracy*: 0.5 dB (VV and HH) Constant incidence angle** between 35° and 50°
Freeze/Thaw State: Capture freeze/thaw state transitions in integrated vegetation-soil continuum with two-day precision, at the spatial scale of land-scape variability (~3 km).	L-Band Radar (1.26 GHz and 1.29 GHz): Polarization: HH Resolution: 3 km Relative accuracy*: 0.7 dB (1 dB per channel if 2 channels are used) Constant incidence angle** between 35° and 50°
Sample diurnal cycle at consistent time of day (6am/6pm Equator crossing); Global, ~3 day (or better) revisit; Boreal, ~2 day (or better) revisit	Swath Width: ~1000 km Minimize Faraday rotation (degradation factor at L-band)
Observation over minimum of three annual cycles	Baseline three-year mission life
* Includes precision and calibration stability ** Defined without regard to local topographic variation	

The SMAP radiometer measures the four Stokes parameters, V, H and T_3, and T_4 at 1.41 GHz. The T_3-channel measurement can be used to correct for possible Faraday rotation caused by the ionosphere, although such Faraday rotation is minimized by the selection of the 6am/6pm sun-synchronous SMAP orbit.

At L-band anthropogenic Radio Frequency Interference (RFI), principally from ground-based surveillance radars, can contaminate both radar and radiometer measurements. Early measurements and results from the SMOS mission indicate that in some regions RFI is present and detectable. The SMAP radar and radiometer electronics and algorithms have been designed to include features to mitigate the effects of RFI. To combat this, the SMAP radar utilizes selective filters and an adjustable carrier frequency in order to tune to pre-determined RFI-free portions of the spectrum while on orbit. The SMAP radiometer will implement a combination of time and frequency diversity, kurtosis detection, and use of T_4 thresholds to detect and where possible mitigate RFI.

The SMAP current data products are listed in Table 2. Level 1B and 1C data products are calibrated and geolocated instrument measurements of surface radar backscatter cross-section and brightness temperatures derived from antenna temperatures. Level 2 products are geophysical retrievals of soil moisture on a fixed Earth grid based on Level 1 products and ancillary information; the Level 2 products are output on half-orbit basis. Level 3 products are daily composites of Level 2 surface soil moisture and freeze/thaw state data. Level 4 products are model-derived value-added data products that support key SMAP applications and more directly address the driving science questions.
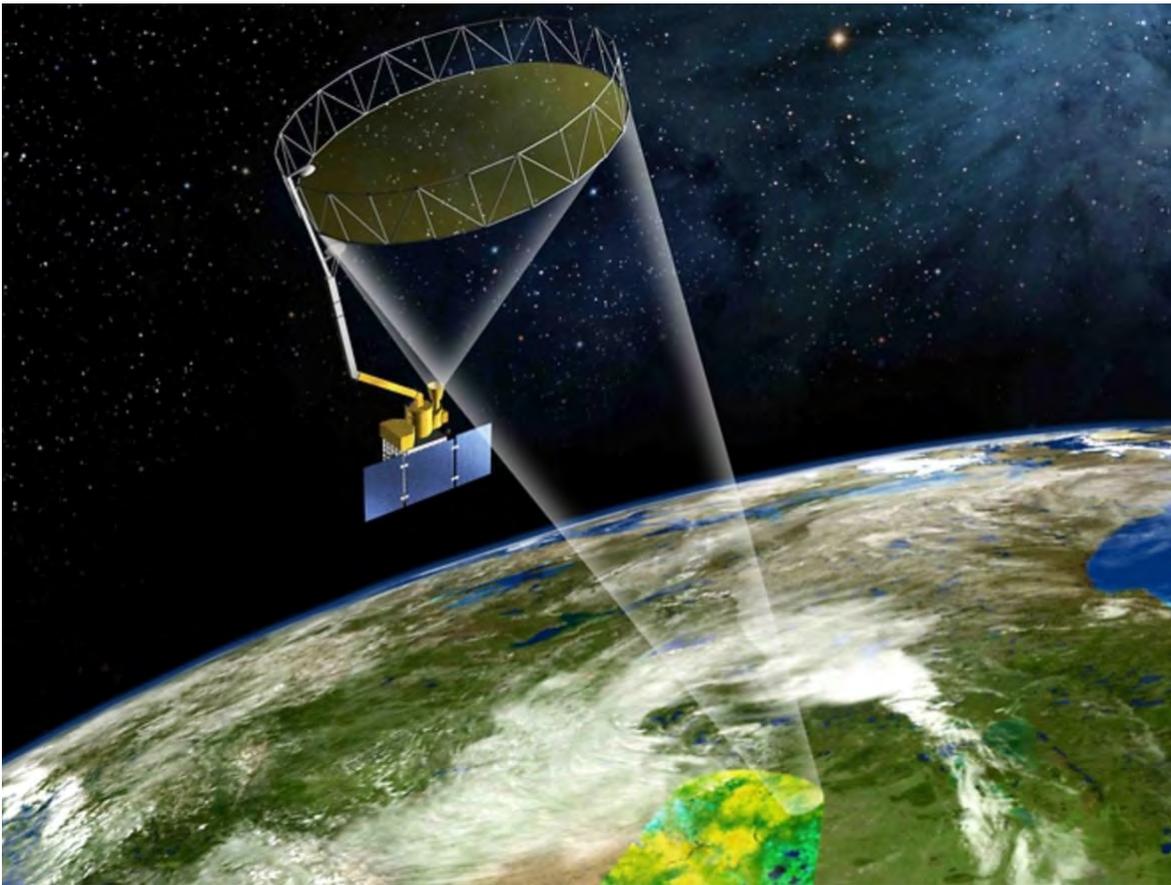

Figure 1: *The SMAP observatory is a dedicated spacecraft with a rotating 6-m light-weight deployable mesh reflector. The radar and radiometer share a common feed.*

Table 2. SMAP Data Products Table.

Product	Description	Gridding (Resolution)	Latency**	Comments	
L1A_Radiometer	Radiometer Data in Time-Order	-	12 hrs		Instrument Data
L1A_Radar	Radar Data in Time-Order	-	12 hrs		
L1B_TB	Radiometer T_B in Time-Order	(36x47 km)	12 hrs		
L1B_S0_LoRes*	Low-Resolution Radar σ_o in Time-Order	(5x30 km)	12 hrs	Discontinued	
L1C_S0_HiRes*	High-Resolution Radar σ_o in Half-Orbits	1 km (1-3 km)*	12 hrs	Discontinued	
L1C_TB	Radiometer T_B in Half-Orbits	36 km	12 hrs		
L2_SM_A*	Soil Moisture (Radar)	3 km	24 hrs		Science Data (Half-Orbit)
L2_SM_P	Soil Moisture (Radiometer)	36 km	24 hrs		
L2_SM_AP*	Soil Moisture (Radar + Radiometer)	9 km	24 hrs	Discontinued	
L3_FT_A*	Freeze/Thaw State (Radar)	3 km	50 hrs	Discontinued	Science Data (Daily Composite)
L3_SM_A*	Soil Moisture (Radar)	3 km	50 hrs	Discontinued	
L3_SM_P	Soil Moisture (Radiometer)	36 km	50 hrs		
L3_SM_AP*	Soil Moisture (Radar + Radiometer)	9 km	50 hrs	Discontinued	
L4_SM	Soil Moisture (Surface and Root Zone)	9 km	7 days		Science Value-Added
L4_C	Carbon Net Ecosystem Exchange (NEE)	9 km	14 days		
L2_SM_SP	SMAP-Sentinel Soil Moisture (Radar + Radiometer)	3 km and 1 km	3-7days	Enhanced Product	Science Value-Added
L2_SM_P_E	Soil Moisture (Radiometer)	9 km	24 hrs	Enhanced Product	
L3_FT_P	Freeze/Thaw State (Radiometer)	9 km	24 hrs	Enhanced Product	

*The SMAP project discontinued the product due to the radar failure.
**The SMAP project makes the best effort to reduce the stated latency.

Currently, SMAP has a non-functioning L-band radar, although the L-band radiometer continues to operate nominally producing high-quality global data. On July 7th, 2015, the SMAP radar malfunctioned and ceased transmitting. Consequently, the production of high-resolution active-passive algorithm-based soil moisture products (L2_SM_AP and L3_SM_AP) became impossible, and only ~2.5 months (April 15th, 2015 to July 7th, 2015) of SMAP radar data are available [6]. Therefore, during the SMAP post-radar phase, many ways were examined to restart the high-resolution soil moisture product generation of the SMAP mission. One of the feasible approaches was to substitute - other available SAR data for the SMAP radar. The Copernicus Sentinel-1A and Sentinel-1B SAR data were found most suitable for combining with the SMAP radiometer data because of their similar orbit configuration that allows overlapping of the SMAP/Sentinel swaths with a minimal time difference, a key feature/requirement for the SMAP active-passive algorithm. The Sentinel-1A/1B interferometric wide swath (IW) mode acquisition also provides the co-polarized and cross-polarized observations required for the SMAP active-passive algorithm. However, some differences do exist between the SMAP and Sentinel-1A/1B SAR data. They are mainly: 1) Sentinel has a C-band SAR whereas SMAP operates at L-band; 2) Sentinel has multiple incidence angles within its swath, and SMAP has one single incidence angle; and 3) Sentinel-1A/1B Interferometric Wide (IW) swath width is ~250 km as compared to SMAP with 1000 km swath width. On any given day, the narrow swath width of the Sentinel-1A/1B observations significantly reduces the overlap spatial coverage between SMAP and Sentinel as compared to the original SMAP radar and radiometer swath coverage. Hence, the temporal resolution (revisit interval) suffers due to narrow overlapped swath width and degrades from 3 days to 12 days. One advantage of using very high-resolution resolution Sentinel-1A/1B data in the SMAP active-passive algorithm is the potential of obtaining the disaggregated brightness temperature and thus soil moisture at a much finer spatial resolution of 3 km and 1 km at global extent. Assessment of the high-resolution product at 3 km and 1 km using SMAP's soil moisture calibration and validation sites shows reasonable accuracy of ~0.05 m3/m3 volumetric soil moisture. The SMAP-Sentinel active-passive high-resolution product is now available

to the public (new version released in October 2018) through NSIDC (NASA DAAC). The duration of this product is from April 2015 to current date.

1.3   Product Objectives

SMAP radiometer measurements in the L-band frequency range are sensitive to surface (~0-5 cm) soil moisture, but with the SMAP 6 m reflector antenna, the effective ground resolution is about 40 km. The Copernicus Sentinel-1A/1B C-band SAR provides higher resolution (~30 m) backscatter measurements. The high-resolution advantage of radar is diminished for soil moisture sensing by the higher sensitivity of radar to surface roughness and vegetation scattering.

Soil moisture variations lead to a dynamic range in brightness temperature that can be tens of degrees Kelvin [7, 8, 9]. The SMAP radiometer can map the surface to within a few degrees of Kelvin accuracy. Soil moisture variations cause only about 5 to 10 dB range in the Sentinel-1A/1B radar backscatter cross-section. With a typical instrument sensitivity of about 1 dB and a distributed target ambiguity of -22 dB, this leads to relatively less soil moisture sensitivity for the instrument. Furthermore, the presence of a vegetation canopy reduces the dynamic range of radar backscatter cross-section faster than radiometer brightness temperature.

For the above reasons, neither the SMAP radiometer nor the Sentinel-1A/1B radar can individually meet the SMAP Level 1 requirements for soil moisture spatial resolution (10 km) and accuracy (0.04 cm$_3$/cm$_3$). This ATBD proposes a high-resolution soil moisture by merging the active (Sentinel-1A/1B radar) and passive (SMAP radiometer) measurements to derive 1 km and 3 km products (L2_SM_SP) that meets the L1 spatial resolution requirements of the SMAP mission; however, meeting the SMAP L1 temporal resolution requirement is not possible through this product because of the ~6-12 day revisit interval of the Sentinel-1A/1B spacecrafts.

Relative to one another, the SMAP radiometer brightness temperature measurements are coarser resolution but with higher sensitivity to soil moisture, while the Sentinel-1A/1B radar backscatter cross-section measurements are higher-resolution but with reduced sensitivity to soil moisture especially over densely vegetated land surfaces. The L2_SM_SP soil moisture product merges the two measurements to produce soil moisture retrievals with 3 km and 1 km resolution.

1.4   Historical Perspective

In the past, numerous studies [11, 12, 13, 14] have attempted to obtain high-resolution soil moisture by downscaling coarse resolution (~50 km) soil moisture products from satellite-based microwave radiometers. These studies used fine scale ancillary geophysical information like topography, vegetation, soil type, and precipitation that exert physical control over the evolution of soil moisture. High-resolution thermal infrared data from MODIS and soil parameters were utilized in a deterministic approach to disaggregate SMOS ~40 km soil moisture product to ~10 km soil moisture estimate [15]. From all these approaches, one common aspect is the use of static and dynamic ancillary geophysical data in the downscaling/disaggregation process. The ancillary geophysical data come from different sources with inherent systematic and random errors, as well as registration mismatches that affect the accuracy of the downscaled soil moisture estimates. Also, the physics of interactions between soil moisture and some geophysical parameters at different scales is not well understood.

Few studies have been conducted to merge high-resolution radar and coarse resolution radiometer measurements in order to obtain an intermediate resolution product. Change detection techniques have demonstrated potential to monitor temporal evolution of soil moisture by taking advantage of the approximately linear dependence of radar backscatter and brightness temperature change on soil moisture change. The feasibility of a change detection approach using the Passive and Active L-band System airborne sensor (PALS) radar and radiometer data obtained during the SGP99 campaign is presented in [16]; a similar approach is used to downscale PALS data using AIRSAR data from the SMEX02 campaign. The limitation of this technique is the estimation of soil moisture relative change and not the absolute value of soil moisture; errors also accumulate over time. A totally different approach is presented in [17], where a Bayesian method is used to downscale radiometer observations using radar measurements. Kim et al. [18] developed a time-series algorithm based on a linear model of backscatter and soil moisture. For estimating soil moisture at intermediate resolution (9 km), they determine the two unknowns of the linear model for each pixel within the coarse radiometer pixel. Piles et al. [19] presented a change detection scheme within a SMAP-like context that uses the approximately linear dependence of change in radar backscatter to soil moisture change at radiometer resolution, temporal change in backscatter at radar resolution and previous day soil moisture data to estimate soil moisture at ~9 km. Like [19], this approach also suffers from accumulation of errors over time. A spatial variability technique developed by [20] to blend SMAP radar measurement and radiometer-based soil moisture data also takes advantage of the approximately linear dependence of backscatter change to soil moisture change at the radiometer resolution which constraints the relative backscatter difference within the coarse radiometer footprint to estimate soil moisture at ~9 km resolution. Unlike [16] and [18], the spatial variability technique used in [20] does not require the previous satellite overpass observations to estimate the current soil moisture value. A new active-passive algorithm is developed by [21] that draws from all the above algorithms and techniques ([16], [18], [19], and [20]). The new active-passive algorithm [21] downscales the coarse-scale radiometer-based gridded brightness temperature using the fine resolution radar backscatter, and then near surface soil moisture is retrieved from the downscaled brightness temperature. The algorithm presented by [21] is the baseline active-passive algorithm to produce the SMAP-only active-passive product (L2_SM_AP) for the ~2.5 months during which the SMAP radar was operational. An article describing the L2_SM_AP algorithm and product was also published by Das et al., [22]. With the failure of SMAP L-band radar, the L2_SM_AP product was discontinued. Therefore, during the SMAP post-radar phase, many ways were examined to restart the high-resolution soil moisture product generation of the SMAP mission. One of the feasible approaches was to substitute the SMAP radar with other available SAR data. The Copernicus Sentinel-1A/1B SAR data were found most suitable for combining with the SMAP radiometer data because of the similar Sentinel orbit configuration that allows overlapping with SMAP swaths with a minimal time difference, a key feature/requirement for the algorithm. The SMAP active-passive algorithm presented in [21] and [22] was further modified to accommodate Sentinel-1A/1B C-band SAR data to produce new enhanced high-resolution (3 km and 1 km) soil moisture data at global extent [23]. The modified SMAP active-passive algorithm is now described in detail in Section 3.2.

1.5   Product Characteristics

The L2_SM_SP product is based on the merger of the SMAP radiometer and processed Sentinel-1A/1B SAR observations at two discrete grid resolutions, i.e., ~33 km and ~1 km, respectively. The Equal-Area-Scalable-Earth 2.0 (EASE2) grid cells of the SMAP radiometer product and processed Sentinel-1A/1B SAR product nest perfectly. The grid definition used in the algorithms is illustrated in Fig. 2 (Section 3.1). The modified SMAP active-passive algorithm disaggregates the

coarse resolution radiometer brightness temperature product based on the spatial variation in high-resolution Sentinel-1A/1B SAR backscatter. In addition, the algorithms require static and dynamic ancillary data. These ancillary data are resampled to the same EASE2 grid prior to ingestion in the L2_SM_SP processing. The dynamic ancillary data used to retrieve soil moisture for particular 1 km and 3 km EASE2 grid cells at a specific point in time are listed in the SMAP L2_SM_SP output files for the benefit of end users.

1.6   Document Outline

This document contains the following sections: **Section 2** describes the basic physics of combining passive microwave and active microwave remote sensing (this Section is similar to the Section in the SMAP-only active-passive ATBD [24]; **Section 3** provides description and formulation of the L2_SM_SP algorithm; **Section 4** presents the test of the L2_SM_SP active-passive algorithm; **Section 5** discusses the practical consideration for implementing the algorithm and generating the L2_SM_SP product; **Section 6** provides description of calibration and validation, quality control, and other practical considerations; **Section 7** the product specifications table; **Section 8** describes possible future work on the algorithm, and; **Section 9** provides a list of references.

2   PHYSICS OF THE PROBLEM

The SMAP active-passive algorithm is essentially focused on the disaggregation of the radiometer brightness temperature using the radar backscatter spatial patterns within the radiometer footprint that are inferred from the radar measurements. The spatial patterns need to account for the different levels of radar backscatter cross-section sensitivity to soil moisture owing to the variability in the density of vegetation cover and surface roughness. For this reason the radar measurements within the radiometer footprint are scaled by parameters that are derived from the temporal fluctuations in the radar and radiometer measurements. Because the radiometer and the radar make measurements over the same location within a given time window, the co-variations over specified periods of time (short relative to plant phenology) are mostly related to surface soil moisture changes rather than contributions of vegetation and surface roughness. The latter two factors change gradually over long time periods such as monthly/seasonally.

The basis for the brightness temperature disaggregation based on radar measurements begins with relating the radiometer measurements with the radar backscatter cross-section measurements in a simple conceptual framework as outlined in this section. This analysis is meant to simply demonstrate the dependencies and it is not directly (i.e., algebraically) part of the SMAP active-passive algorithm formulation.

The brightness temperature at polarization p at L-band and its dependency on surface characteristics may be demonstrated using the $\tau - \omega$ model (refer to the SMAP ATBD: L2 & L3 Radiometer Soil Moisture (Passive) Products. SMAP Project, JPL D-66480, Jet Propulsion Laboratory):

$$T_{B_p} = T_s \cdot e^{-\tau_p/cos\theta} \cdot e_p + T_c \cdot (1 - \omega) \cdot \left(1 - e^{-\tau_p/cos\theta}\right) \cdot \left(1 + r_p \cdot e^{-\tau_p/cos\theta}\right) \qquad (1)$$

where T_s is the soil effective temperature, T_c is the vegetation temperature, τ_p is the nadir vegetation opacity, ω_p is the vegetation single scattering albedo, and r_p is the soil reflectivity [10].

At the morning nodal crossing overpass time, near subsurface-to-surface-canopy thermal conditions are expected to be nearly isothermal so that $T_s \approx T_c \equiv T$. Under low vegetation cover conditions, with the assumption that the single-scattering albedo can be neglected so that $\omega \ll 1$. The polarized emissivity and reflectivity are related by $e_p = 1 - r_p$. Using these assumptions and identities, (1) becomes simply:

$$T_{B_p} = T\big(1 - r_p\big)e^{-2\tau_p/cos\theta} \tag{2}$$

The surface reflectivity can be decomposed into a component for smooth surfaces r_{s_p} corrected for roughness as in $r_p = r_{s_p}e^{-hcos^2\theta}$ where h is a roughness parameter related to the root-mean-square (RMS) deviations of surface roughness. Now (2) becomes:

$$T_{B_p} = T \cdot \left(1 - r_{s_p} \cdot e^{-hcos^2\theta}\right) e^{-2\tau_p/cos\theta} \tag{3}$$

The radar backscatter cross-section for co-polarization pp at (L-band or C-band) is:

$$\sigma_{pp}^t = \sigma_{pp.}^{surf} e^{-2\tau_p/cos\theta} + \sigma_{pp}^{vol} + \sigma_{pp}^{int} \tag{4}$$

The first term is the surface backscatter, σ_{pp}^{surf}, modified by the two-way attenuation through a vegetation layer of nadir opacity τ_p. The second term represents the backscatter from the vegetation volume, σ_{pp}^{vol}. The third term represents interactions between vegetation and the soil surface, σ_{pp}^{int}.

From the empirical models presented in [22] and [23], the surface contribution σ_{pp}^{surf} is conceptualized as the product of polarization magnitude $|\alpha_{pp}|^2$ and surface roughness characteristics as captured in a function $f_1(roughness)$ as in:

$$\sigma_{pp}^{surf} = f_1(roughness) \cdot |\alpha_{pp}|^2 \tag{5}$$

The polarization magnitude $|\alpha_{pp}|^2$ is a function of soil dielectric properties and incidence angle. It is related to the soil reflectivity r_{s_p} in the horizontal co-polarization if the center-frequency of the radar and radiometer are close. In the vertical co-polarization, the polarization magnitude and soil reflectivity are near-linearly proportional but not equivalent. Given the proportionality $|\alpha_{pp}|^2 \propto r_{s_p}$ through conservation of energy, the linear coefficients of the relationship may be absorbed in the empirical function $f_1(roughness)$ so that:

$$\sigma_{pp}^{surf} = f_1(roughness) \cdot r_{s_p} \tag{6}$$

The interaction term σ_{pp}^{int} is a complex function of vegetation properties, soil roughness characteristics as well as surface reflectivity. The interaction term may be written as:

$$\sigma_{pp}^{int} = f_2(roughness, vegetation) \cdot r_{s_p} \tag{7}$$

using a function $f_2(roughness, vegetation)$ that depends on surface roughness and vegetation in complex ways. The vegetation volume scattering σ_{pp}^{vol} is a complex function of vegetation alone through a third function $f_3(vegetation)$ as in:

$$\sigma_{pp}^{vol} = f_3(vegetation) \tag{8}$$

Now the radar backscatter cross-section for co-polarization *pp* is:

$$
\begin{aligned}
\sigma_{pp}^t = \\
f_1(roughness) \cdot r_{s_p} \cdot e^{-2\tau_p/cos\theta} \\
+ f_2(roughness, vegetation) \cdot r_{s_p} \\
+ f_3(vegetation)
\end{aligned}
$$

Solve for r_{s_p}:

$$r_{s_p} = \frac{1}{f_1 \cdot e^{-2\tau_p/cos\theta} + f_2} \cdot (\sigma_{pp}^t - f_3) \tag{9}$$

The dependencies of the functions f_1, f_2 and f_3 are not carried forward in order to simplify the notation in (9). They depend on surface roughness, vegetation characteristics and incidence angle in complex ways. Owing to the conical scan strategy adopted by SMAP, they are, however, not dependent on incidence angle which is usually a confounding factor in radar backscatter modeling and retrievals.

Substituting (9) into (3) yields:

$$T_{B_p} = T\left(1 - \frac{e^{-h}e^{-2\tau_p/cos\theta}}{f_1 \cdot e^{-2\tau_p/cos\theta} + f_2} \cdot (\sigma_{pp}^t - f_3)\right)$$

or:

$$\frac{T_{B_p}}{T} = \left[1 + \frac{e^{-h}e^{-2\tau_p/cos\theta}}{f_1 \cdot e^{-2\tau_p/cos\theta} + f_2} * f_3\right] + \left[-\frac{e^{-h}e^{-2\tau_p/cos\theta}}{f_1 \cdot e^{-2\tau_p/cos\theta} + f_2}\right] \cdot \sigma_{pp} \tag{10}$$

which suggests a linear functional dependence of brightness temperature and radar backscatter cross-section in the presence of vegetation canopy. The slope $\left[\beta \equiv -\frac{e^{-h}e^{-2\tau_p/cos\theta}}{f_1 \cdot e^{-2\tau_p/cos\theta} + f_2}\right]$ and intercept $\alpha \equiv \left[1 + \frac{e^{-h}e^{-2\tau_p/cos\theta}}{f_1 \cdot e^{-2\tau_p/cos\theta} + f_2} * f_3\right]$ are dependent on vegetation, surface roughness characteristics, and viewing angle.

3   RETRIEVAL ALGORITHM

3.1   Grid Definition

Figure 2 shows the nested grid topology of the EASE2 grid radiometer brightness temperature (33 km), EASE2 grid radar backscatter cross-section (1 km), and desired merged active-passive L2_SM_SP (3 km) products. For convenience in mathematical formulation, the naming convention of 'C' (coarse), 'F' (fine), and 'M' (medium) for the L3_SM_P_E (33 km resolution, however, gridded at 9 km), S0_HiRes (Sentinel-1A/1B processed and upscaled to 1 km), and L2_SM_SP grid scales, respectively, is used throughout the following sections. It is evident from the grid topology

(Fig. 2) that within a single ($nc = 1$) 33 km x 33 km grid cell of C there are $nm = 121$ grid cells of M, and $nf = 1089$ grid cells of F.
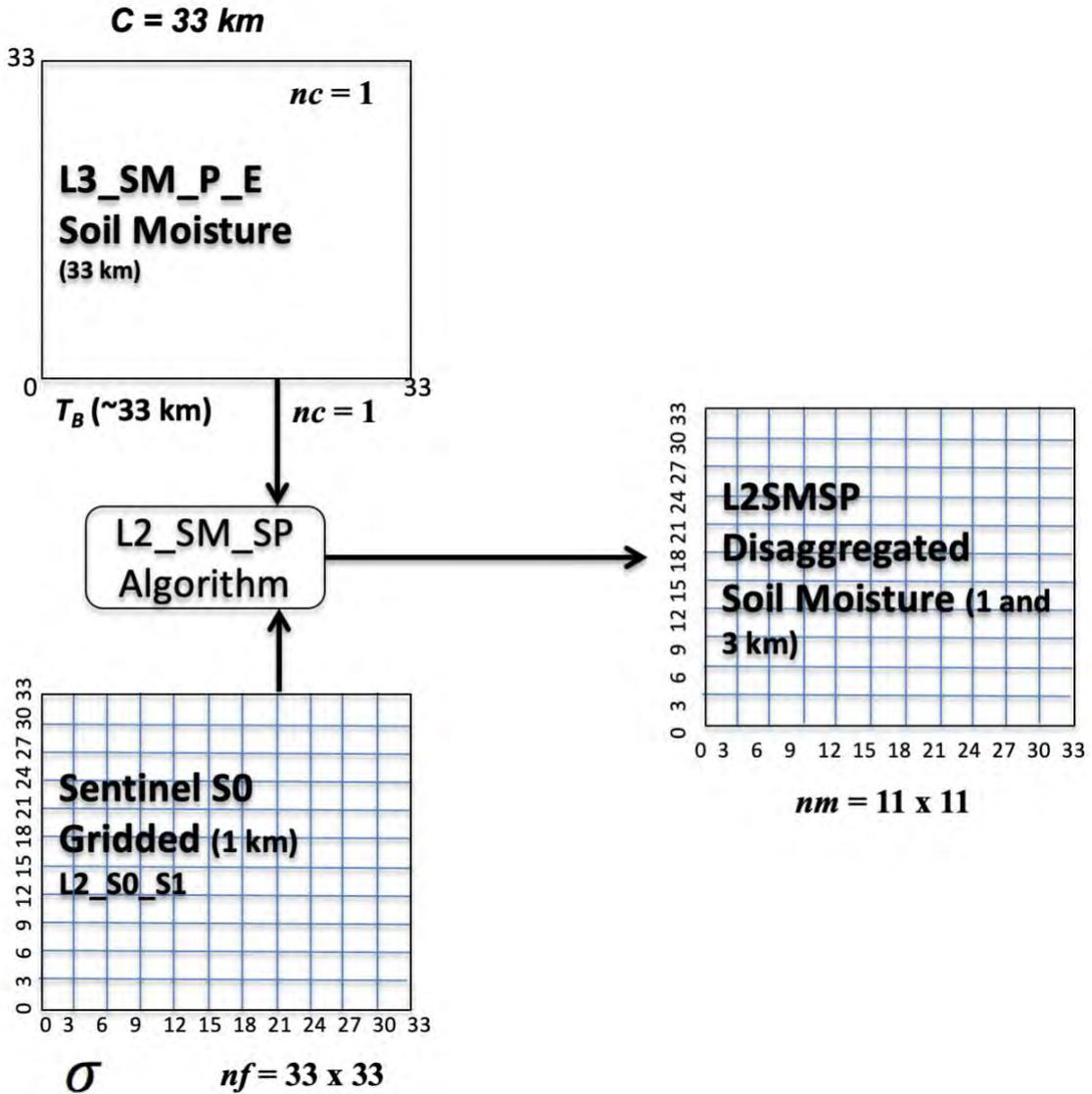

Figure 2: Grid definition of radiometer, radar, and merge product where nf and nm are the number of area pixels of radar and merged product, respectively, within one radiometer area pixel nc.

3.2 Formulation of SMAP-Sentinel L2_SM_SP Algorithm

The SMAP L-band radiometer will measure the natural microwave emission in the form of the brightness temperature (T_{B_p}) of the land surface, while the C-band radar will measure the energy backscattered (σ_{pp}) from the land surface after transmission of an electromagnetic pulse. With both SMAP radiometer and Sentinel-1A/1B radar data over a particular region on the Earth over a short period of time, an increase of surface soil moisture or soil dielectric constant will lead to a decrease in radiometer T_B [10] and an increase in radar σ [11] measurements, and vice-versa. During this short

time period T_B and σ are negatively correlated due to soil moisture variations in time. The time period is generally shorter than seasonal phenology of vegetation. The land surface vegetation and surface roughness factors vary on time scales longer that those associated with soil moisture. It should be noted that in some agricultural land use regions the vegetation can grow and change attributes rapidly over a few days that may be a source of error. Also, precipitation and associated surface disturbances can change the soil roughness characteristics that may introduce another source of error. Despite these sources of uncertainty, within this region of interest over a short period of time the SMAP measured T_{B_p} and Sentinel-1A/1B σ_{pp} are expected to have a functional relationship, which based on (10) is likely a linear functional relationship:

$$T_{B_p} = \alpha + \beta \cdot \sigma_{pp} \tag{11}$$

The unknown parameters α and β are dependent on the dominant vegetation and soil roughness characteristics (see discussion following Equation 10). Energy conservation that yields the linear dependence in (11) is based on units of power for the radar backscatter cross-section. However, the linear functional relationship also holds when σ_{pp} is expressed in decibel (dB). The T_B polarization can either be v or h and the σ polarization is either vv or hh. Equation (1) evaluated at scale C (33 km) is:

$$T_{B_p}(C) = \alpha(C) + \beta(C) \cdot \sigma_{pp}(C) \tag{12}$$

Here $\sigma_{pp}(C) = \frac{1}{nf} \sum_{i=1}^{nf} \sigma_{pp}(F_i)$, where $F = 1$ km grid resolution and nf is the number of F grid cell within C (Fig. 2). Equation 12 is based on the assumption that the linear relationship between the $T_{B_p}(C)$ and $\sigma_{pp}(C)$ holds. Therefore, it is also important to demonstrate that the similar linear relationship is found between the SMAP radiometer $T_{B_p}(C)$ and spatially-averaged Sentinel radar data $\sigma_{pp}(C)$. Figure 3 illustrates the scatters between the SMAP $T_{B_p}(C)$ and Sentinel $\sigma_{pp}(C)$ from various regions of the world with different land covers having varying amounts of vegetation water content (VWC).
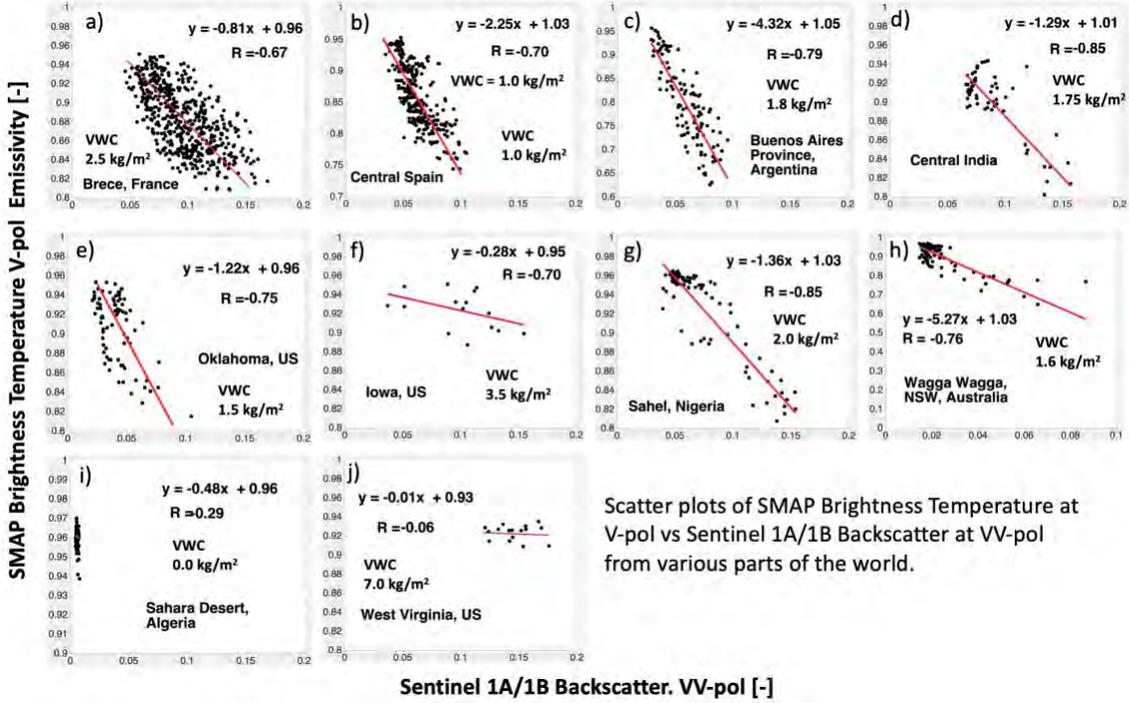
Figure 3: Scatter plots between the SMAP L-band radiometer $T_{B_p}(C)$ and spatially-averaged Sentinel C-band radar data $\sigma_{pp}(C)$ from various regions of the world.

Two years (2017 and 2018) of data are used to create the plots in Fig. 3 wherever there is overlap between the SMAP $T_{B_p}(C)$ and Sentinel $\sigma_{pp}(C)$. The slope of the correlation between L-band $T_{B_p}(C)$ and the C-band $\sigma_{pp}(C)$ depends on the level of VWC and the surface roughness. As expected the slope is ~0 for very highly vegetated regions such as West Virginia (Fig. 3j). It is obvious from Fig. 3 that the nearly linear relationship is valid for most of the world. However, low correlation is also visible over the dry and arid Sahara desert because the dynamic range in $T_{B_p}(C)$ and $\sigma_{pp}(C)$ is not observed during the two year period. The number of samples for any given site of Fig. 3 is dependent on the availability of the Sentinel-1A/1B granules. With the current global coverage configuration from 2016 October onwards including Sentinel-1A and Sentinel-1B, the revisit interval is nearly 12 days over most parts of the world except Europe. In Europe, the Sentinel-1A and Sentinel-1B combination acquires observations at a 6-day revisit interval. Therefore, over the European sites in Figure 3 more samples are available. The number of samples also suffers from the SMAP and the Sentinel overlap restriction of +/- 24 hours.

To develop the SMAP-Sentinel Active-Passive algorithm further, (11) can also be conceptually evaluated at the scale M (3 km) within the radiometer footprint C:

$$T_{B_p}(M_j) = \alpha(M_j) + \beta(M_j) \cdot \sigma_{pp}(M_j) \tag{13}$$

where $\sigma_{pp}(M_j) = \sum_{i=1}^{nm_j} \sigma_{pp}(F_i)$ obtained from the SMAP high-resolution (3 km) radar data product.

Here $T_{B_p}(M_j)$ is the unknown brightness temperature at scale M_j. This scale brightness temperature is not available given the SMAP radiometer instrument resolution. In fact this variable is the target of the algorithm and it is referred to as the disaggregated brightness temperature.

The first step in developing the algorithm is to subtract (12) from (13):

$$T_{B_p}(M_j) - T_{B_p}(C) = \{\alpha(M_j) - \alpha(C)\} + \{[\beta(M_j) \cdot \sigma_{pp}(M_j)] - [\beta(C) \cdot \sigma_{pp}(C)]\} \qquad (14)$$

Because $T_{B_p}(M_j)$ is not available, we cannot estimate the parameters $\alpha(M_j)$ and $\beta(M_j)$ in the manner that was followed at scale C. The path forward to incorporate the effects of the variations of these parameters at scale M_j with respect to the coarser scale C begins with algebraically rewriting (14) as:

$$
\begin{aligned}
T_{B_p}(M_j) & \\
= T_{B_p}(C) + & \qquad \text{RHS Term I} \\
\{\beta(C) \cdot [\sigma_{pp}(M_j) - \sigma_{pp}(C)]\} & \qquad \text{RHS Term II} \\
+ \{[\alpha(M_j) - \alpha(C)] + [\beta(M_j) - \beta(C)] \cdot \sigma_{pp}(M_j)\} & \qquad \text{RHS Term III} \qquad (15)
\end{aligned}
$$

The left-hand-side of (15) is the target variable of the Active-Passive algorithm, i.e. the disaggregated brightness temperature at the 3 km scale M_j.

The first term on the right-hand-side (RHS Term I), $T_{B_p}(C)$, is the radiometer-measured brightness temperature at 33 km or scale C. This is the brightness temperature corrected for water body contributions that is produced during the L2_SM_P_E processing and it is primarily based on the radiometer measurement.

The RHS Term II, $\{\beta(C) \cdot [\sigma_{pp}(M_j) - \sigma_{pp}(C)]\}$, can be calculated based on the regression parameter $\beta(C)$ that is estimated through the time-series of radiometer brightness temperature measurements and radar measurements aggregated to scale C. The remainder of this second RHS term ($[\sigma_{pp}(M_j) - \sigma_{pp}(C)]$) is also based on the radar measurements aggregated to scales M_j and C.

The RHS Term III accounts for the deviations of the parameters α and β within the grid cell C. The term $\{[\alpha(M_j) - \alpha(C)] + [\beta(M_j) - \beta(C)] \cdot \sigma_{pp}(M_j)\}$ is in units of brightness temperature and represents subgrid scale (relative to C) heterogeneity effects. The parameters α and β depend on vegetation and surface roughness. For a perfectly homogeneous region, the parameters $\alpha(M_j) = \alpha(C)$ and $\beta(M_j) = \beta(C)$, and the subgrid heterogeneity term becomes zero. However, in nature homogeneity within C rarely exists.

The SMAP radar also provided high-resolution cross-polarization radar backscatter measurements at scale F which are principally sensitive to vegetation and surface characteristics. The subgrid deviation/heterogeneity patterns in vegetation and roughness as captured by the cross-polarization backscatter at scale M_j is $[\sigma_{pq}(C) - \sigma_{pq}(M_j)]$. This indicator can be converted to variations in co-polarization backscatter through multiplications by a sensitivity parameter $\left[\frac{\partial \sigma_{pp}(M_j)}{\partial \sigma_{pq}(M_j)}\right]_C$. This sensitivity, denoted by the scale C variable $\Gamma \equiv \left[\frac{\partial \sigma_{pp}(M_j)}{\partial \sigma_{pq}(M_j)}\right]_C$, is specific to the particular grid cell C and the particular season for grid cell C. It is estimated based on the collection of co-polarization and cross-polarization radar backscatter cross-section within each grid cell C.

The term $\Gamma \cdot [\sigma_{pq}(C) - \sigma_{pq}(M_j)]$ is the projection of the cross-polarization subgrid deviation onto the co-polarization space. These variations are due to the heterogeneity in parameters

α and β in the radar co-polarization space. It can be converted to brightness temperature units for use in (15) through multiplication by $\beta(C)$, the particular radiometer grid scale C conversion factor relating co-polarization backscatter variations to brightness temperature variations. Thus the product $\beta(C) \cdot \Gamma \cdot \left[\sigma_{pq}(C) - \sigma_{pq}(M_j)\right]$ is the contribution of subgrid (subgrid to scale C) variations in α and β to the brightness temperature at scale M_j. The SMAP Active-Passive brightness temperature disaggregation algorithm is completed by substituting the term $\beta(C) \cdot \Gamma \cdot \left[\sigma_{pq}(C) - \sigma_{pq}(M_j)\right]$ to RHS Term III in (15),

$$
\begin{aligned}
T_{B_p}(M_j) = T_{B_p}(C) + \\
\{\beta(C) \cdot \left[\sigma_{pp}(M_j) - \sigma_{pp}(C)\right]\} \\
+\beta(C) \cdot \Gamma \cdot \left[\sigma_{pq}(C) - \sigma_{pq}(M_j)\right]
\end{aligned}
\tag{16}
$$

which can be written more compactly as

$$
T_{B_p}(M_j) = T_{B_p}(C) + \beta(C) \cdot \left\{\left[\sigma_{pp}(M_j) - \sigma_{pp}(C)\right] + \Gamma \cdot \left[\sigma_{pq}(C) - \sigma_{pq}(M_j)\right]\right\}
\tag{17}
$$

Equation (17) is similar to the originally developed SMAP Active-Passive algorithm ([18], [19], [20]).

For the SMAP-Sentinel1 Active-Passive algorithm, (17) is further modified to work in emissivity space instead of brightness temperature space and the Sentinel backscatter are in linear scale [-].

$$
T_{B_p}(M_j) = \left[\frac{T_{B_p}(C)}{Ts} + \beta'(C) \cdot \left\{\left[\sigma_{pp}(M_j) - \sigma_{pp}(C)\right] + \Gamma \cdot \left[\sigma_{pq}(C) - \sigma_{pq}(M_j)\right]\right\}\right] \cdot Ts
\tag{18}
$$

where Ts [K] is the emission temperature of the surface soil. The parameter Γ [-] is estimated the same way as mentioned above, although in a linear scale. The parameter $\beta'(C)$ [-] is estimated through a snapshot approach because the Sentinel revisit interval of 12 days makes the time series of the Sentinel $\sigma_{pp}(M)$ [-] and $\sigma_{pq}(M)$ [-] data very sparse. Certain aspects of implementation are changed to estimate the $\beta'(C)$ to make it more effective and applicable to accommodate the 12- day revisit interval of the Sentinel satellite. This modification is essential as with the 12-day Sentinel revisit, the $T_{B_p}(C)$ and $\sigma_{pp}(C)$ time series is too sparse, and parameter estimation through the time series approach is ineffective/unfeasible. With time-series sampling for the 12-day repeat cycle, accumulation of enough data pairs is too low to allow the statistical estimation of $\beta(C)$ over periods when the vegetation or soil roughness conditions are changing with seasons, and this does not satisfy the basic assumption of the SMAP Active-Passive algorithm. To overcome the limitation of sparse times series, a snapshot retrieval approach is adopted to estimate the co-variation parameter from the SMAP radiometer and the Sentinel radar observations. The snapshot $\beta'(C)$ is retrieved at each coarse grid cell (C) for every overlap between the SMAP and Sentinel observations, and is computed as ([25] and [26]):

$$
\beta'(C) = \frac{\frac{T_{B_p}(C)}{Ts} - (\gamma + (1-\omega)(1-\gamma))}{\left|S_{pp}(M_j)\right|^2 - \mu_{pp-pq} \cdot \left|S_{pq}(M_j)\right|^2}
\tag{19}
$$

where ω [-] is the effective single scattering albedo, $\gamma = e^{-\tau/cos\theta}$ [-] is the vegetation loss term, and θ_i [rad] is the incidence angle. $\beta'(C)$ in (19) results from eliminating smooth surface Fresnel reflectivity from the tau-omega model and variations in co-polarized backscatter that are due to soil moisture and not vegetation. The numerator is the measured surface emission minus the vegetation

volume scattering and emission. The denominator is similarly the co-polarized backscatter minus the volume scattering. The volume scattering component in the co-polarized backscatter is the total co-polarized backscatter minus the projection of the cross-polarized backscatter onto the co-polarized backscatter. The projection is

$$\mu_{pp-pq} = \partial |S_{pp}(M_j)|^2 / \partial |S_{pq}(M_j)|^2 \tag{20}$$

The nadir vegetation opacity τ [-] is related to the physical characteristics of the vegetation layer, such as the VWC. $|S_{pp}(M_j)|^2$ is co-polarized backscatter, where $|S_{pp}(M_j)|^2 \equiv \sigma_{pp}(M_j)$, and $|S_{pp}(M_j)|^2$ is cross-polarized backscatter, where $|S_{pq}(M_j)|^2 \equiv \sigma_{pq}(M_j)$. μ_{pp-pq} is the same as Γ in (2), except using a linear regression of backscattering coefficients ($\sigma_{pp}(M_j)$ [-], $\sigma_{pq}(M_j)$ in linear units) at fine scale (1 km) within each coarse-resolution TB grid cell ($T_{B_p}(C)$). This approach to estimate $\beta'(C)$ and μ_{pp-pq} does not require time series of $T_{B_p}(C)$ and $\sigma_{pp}(C)$. The snapshot approach Eq. (19) ([25] and [26]) is capable of accommodating L-band, C-band and X-band combinations of the radiometer and SAR observations at different incident angles. On any given day, the snapshot estimate of the covariance parameter (β') is unique and is dependent on the radiometer T_{B_p} (emissivity), SAR backscatter, ω [-] (the effective single scattering albedo), and $\gamma = e^{-\tau/\cos\theta}$ [-] the vegetation loss term (τ is vegetation optical density and θ is incident angle of TB).

For evaluation of $\beta'(C)$ retrieved by the snapshot approach, a comparison was made with $\beta(C)$ derived from the time series purely obtained from data of the SMAP mission (SMAP radar and radiometer). Both approaches converge with the $\beta'(C)$ values similar to $\beta(C)$ as shown in Fig. 4, except over dryland regions across the Sahara, parts of the Middle East and Central Asia. These dryland regions do not have enough soil moisture variability during the April 1 to July 7 Summer season of 2015 (when the SMAP radar data are available) to induce variations in $T_{B_p}(C)$ and $\sigma_{pp}(C)$ to allow valid time-series estimation of $\beta(C)$. Outside of these regions, the magnitudes and distribution of the covariation parameter are similar between the statistical time-series and snapshot approaches [26].
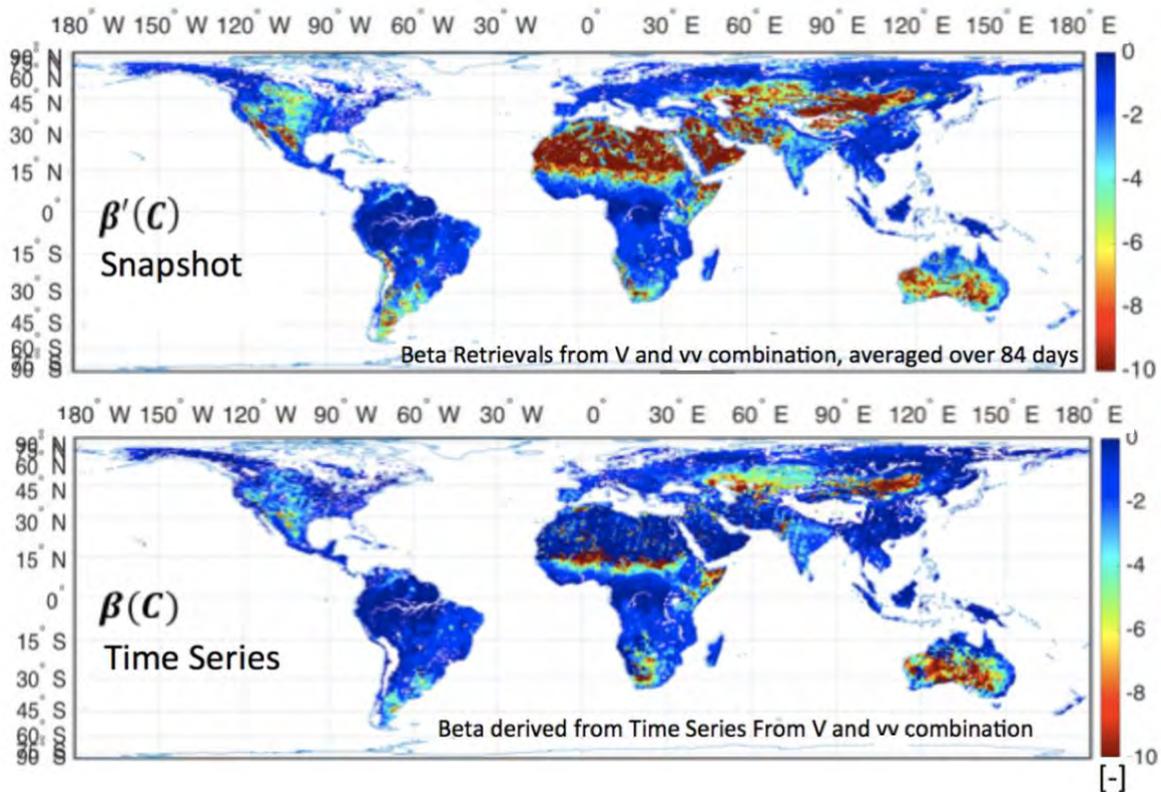
Figure 4: Comparison of snapshot-retrieved $\beta'(C)$ and time series-retrieved $\beta(C)$ at global extent for the SMAP Active-Passive period (~2.5 months in 2015).

The baseline SMAP L2_SM_SP algorithm has two parameters ($\beta'(C)$ and Γ), as shown in (18). The performance of the brightness temperature disaggregation that results in the 3 km and 1 km soil moisture retrievals is heavily dependent on robust estimates of the parameters $\beta'(C)$ and Γ. Figure 5 shows the mean and coefficient of variation (CV) of $\beta'(C)$ at global extent using SMAP radiometer and Sentinel-1A/B backscatter data from May 1, 2015 to April 30, 2017. The global evolution of mean $\beta'(C)$ (Fig. 5) shows the typical feature of reducing magnitude (approaching zero) with increasing VWC [23]. However, the CV in Fig. 5 represents high variability except over very arid regions. This is a clear indication of seasonality/variability in $\beta'(C)$ and the gradually changing values with the surface conditions, especially VWC. Some low absolute values are also observed over the Sahara desert because of local variation in roughness values leading to high backscatter even for the very dry surface.
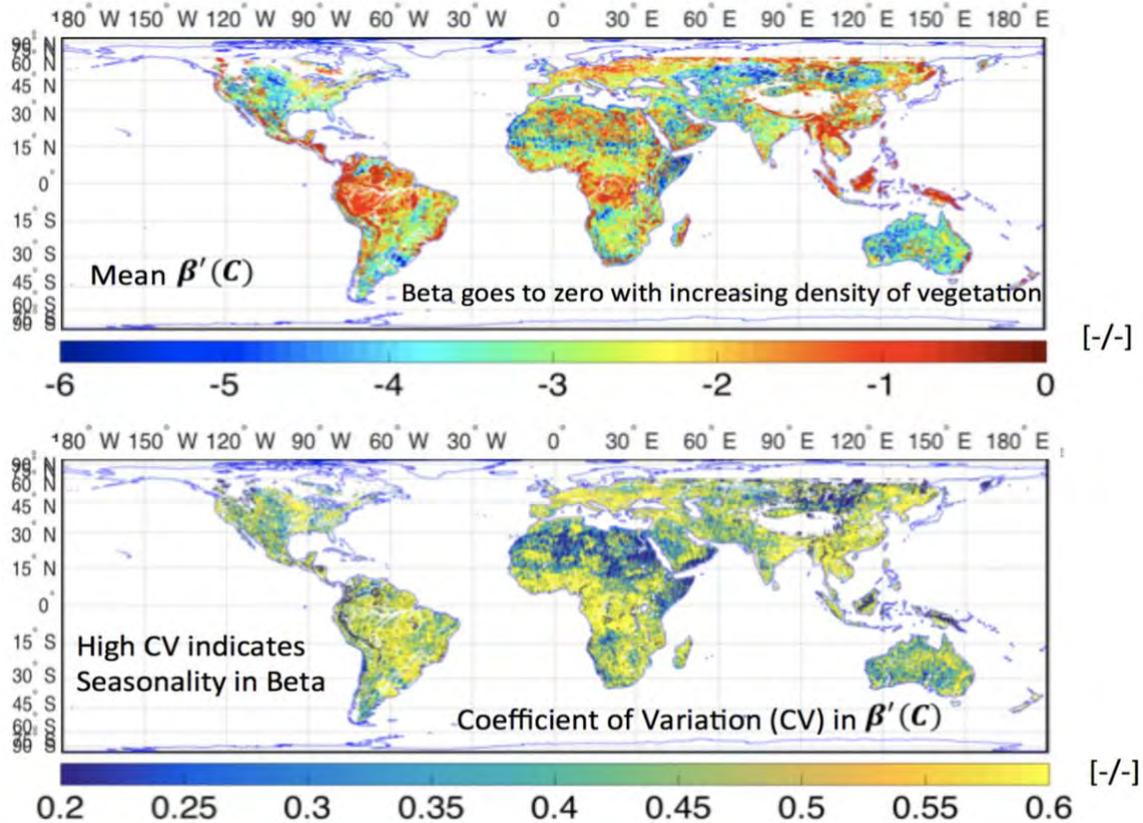
Figure 5: $\beta'(C)$ mean and CV computed using all the available SMAP radiometer data and Sentinel-1A/Sentinel-1B σ_{pp} data from May 01, 2015 to April 30, 2017.

The estimation of $\beta'(C)$ through (19) does not require time series of $T_{B_p}(C)$ and $\sigma_{pp}(C)$. Therefore, space-borne radar and radiometer acquisitions with varying incidence angle can be used and the covariation parameter $\beta'(C)$ is dependent on the angle. The range of incidence angles for Sentinel-1A/1B observations within the C scale (C-scale is 33 km resolution) is ~1 deg. Therefore, linearly averaging the Sentinel-1A/1B backscatter is quantitatively possible and valid. In Fig. 6 from [26], the dynamics of the covariation estimation with variation of incidence angle (Sentinel-1A/B: $34° – 44°$) is presented for four different ranges of vegetation water content (VWC) within the African continent. The covariation parameter $\beta'(C)$ from low to moderate amounts of vegetation (VWC < 5 [kg/m$_2$]), gradually decreases in magnitude with increasing plant moisture. However, the largest change in magnitude of $\beta'(C)$ along incidence angle (for the lowest VWC-range in Fig. 6) is around 0.5 [23]. As expected, $\beta'(C)$ shows minimum sensitivity to incidence angle variations for strongly vegetated areas (VWC > 6 [kg/m$_2$]) leveling around -1.5 [-]. This might be due to the insensitivity of both the SMAP radiometer (L-band) and Sentinel-1A/B radar (C-band) to soil moisture variations under highly moist vegetation. One interpretation is that the incidence angle variation of Active-Passive microwave covariation is increasingly masked/gets absorbed by denser/thicker vegetation [26].
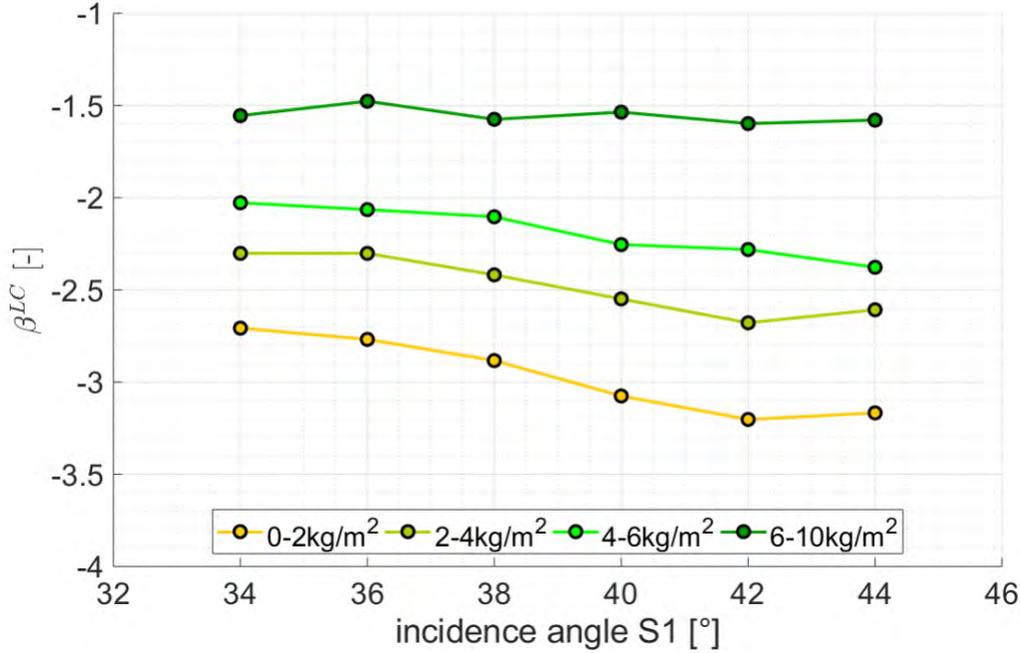
Figure 6. Time-averaged (04/2015-04/2017) $\beta'(C)$ [-] along Sentinel-1A/B incidence angle [°] for four VWC ranges in Africa; circles indicate median values for each VWC range, sourced from [26].

The parameter Γ is determined statistically for any particular overpass using the Sentinel-1A/1B radar backscatters σ_{pp} and σ_{pq} at the finest available resolution (in this case at 1 km) that are encompassed within the 33 km $T_{B_p}(C)$ grid cell. The parameter Γ is the projection of Sentinel-1A/1B σ_{pq} space into the σ_{pp} space. It is the slope of covariance between the Sentinel-1A/1B σ_{pq} and the σ_{pp} (Γ is estimated as $\equiv \left[\frac{\partial \sigma_{pp}(M_j)}{\partial \sigma_{pq}(M_j)}\right]_C$). Γ shows that the heterogeneity is captured through the spatial deviation of σ_{pq} backscatter from its mean at (C) scale. The Γ value projects this spatial deviation in σ_{pq} backscatter in the σ_{pp} backscatter space. The projection of spatial deviation in σ_{pq} backscatter can be additive or negative with the σ_{pp} backscatter depending on the vegetation and surface roughness. Therefore, one Γ value is sufficient to capture the heterogeneity of the scene within the (C) scale. In Figure 7, the values of Γ for all arid regions of the Earth surface are between 4 and 5. This is because the range of the σ_{pq} backscatter response is much lower in the arid region than any other land cover type. Figure 7 illustrates the mean and Coefficient of Variation (CV) of Γ values over the global extent using all data from May 1, 2015 to April 30, 2017. The Γ parameter is spatially and temporally more stable than $\beta'(C)$. At a global extent, the mean values range from 2.5 to 4.5. The CV in Γ is also very low for any given location, indicating temporal stability of this parameter.
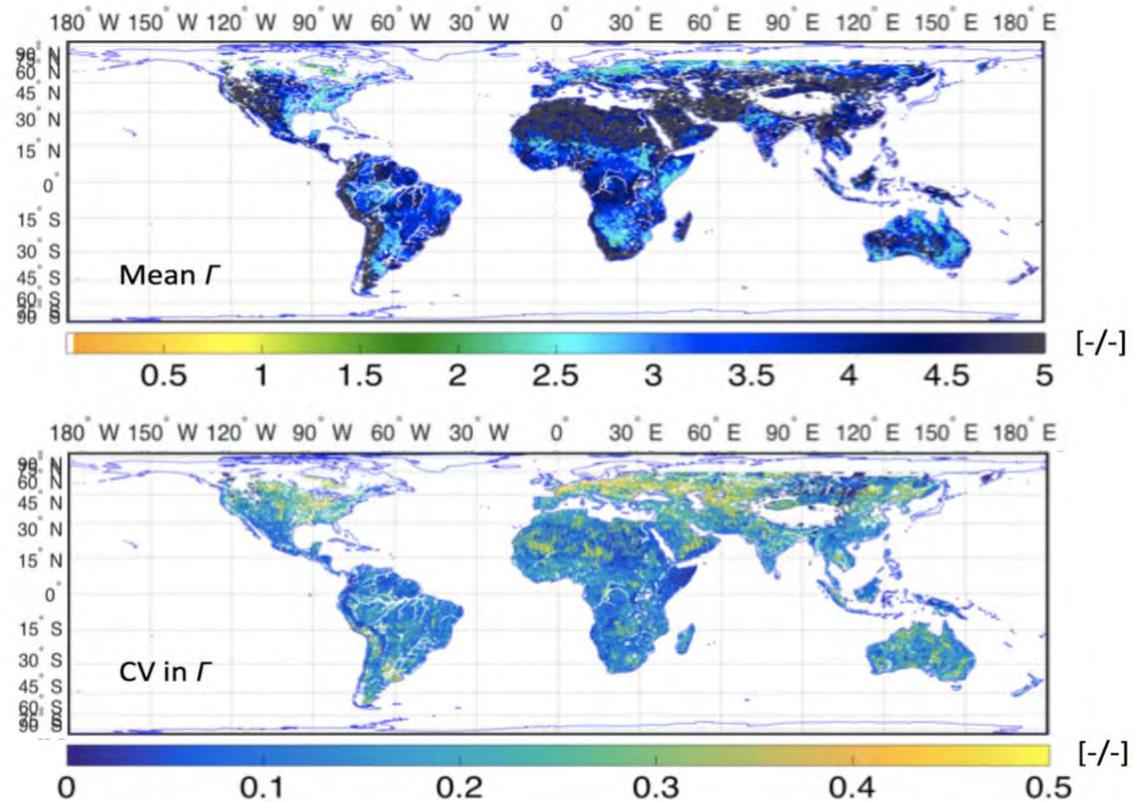
Figure 7: Γ mean and CV computed using all the available SMAP radiometer data and Sentinel-1A/Sentinel-1B σ_{pp} data (1 km resolution) from May 01, 2015 to April 30, 2017.

3.3   Process Flow of the L2_SM_SP Algorithm

A simplified process flow chart/processing scheme of the SMAP-Sentinel1 Active-Passive algorithm implementation is shown in Fig. 8. The input data are the Sentinel-1A/1B Interferometric Wide (IW) Swath mode backscatter σ_{pp} (co-pol *vv*) and σ_{pq} (cross-pol *vh*) at 1 km EASE2 grid resolution and the brightness temperature $T_{B_p}(C)$ from the SMAP Level-2 Enhanced product (L2_SM_P_E) at about 33 km spatial resolution in EASE2 9 km grid.
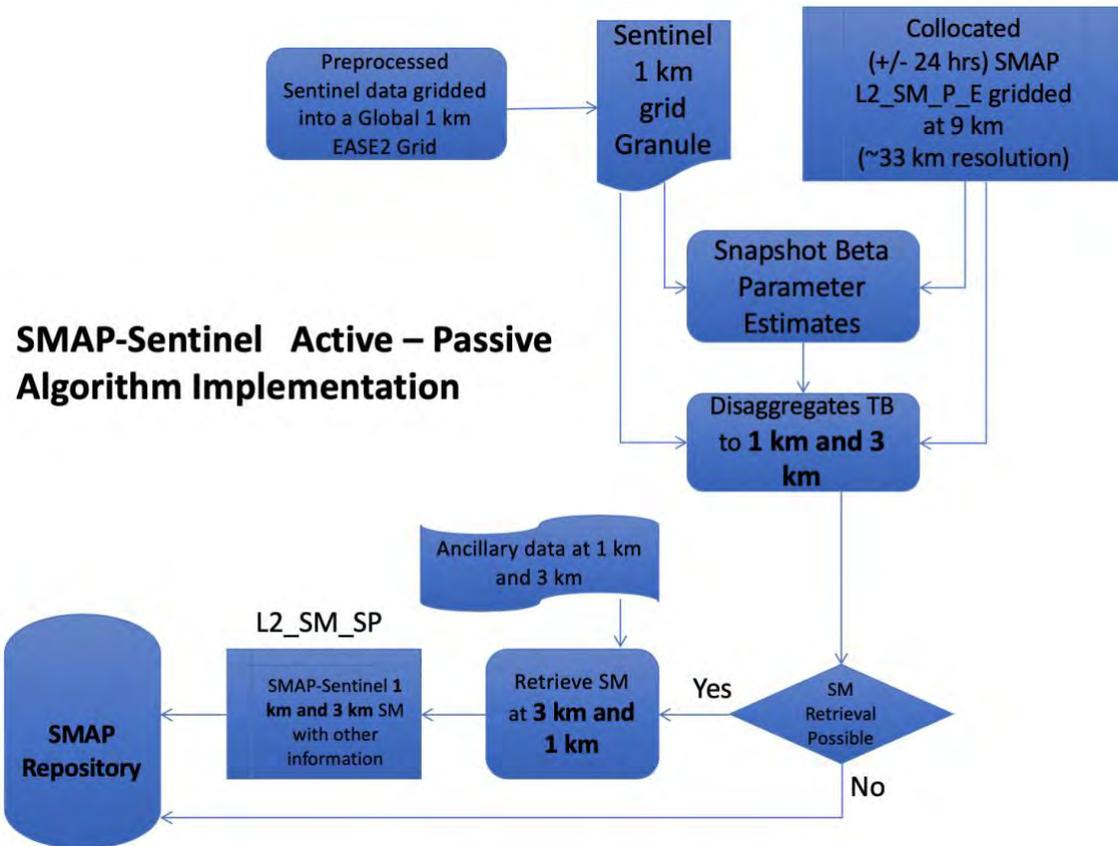
Figure 8: Process flow scheme of the SMAP-Sentinel-1A/1B Active-Passive (L2_SM_SP) algorithm in the JPL SDS.

The native resolution of Sentinel-1A/1B IW swath mode backscatter σ_{pp} (co-pol vv) and σ_{pq} (cross-pol vh) is ~25 meters. The high-resolution Sentinel-1A/1B SAR backscatter data are processed for calibration, noise subtraction, terrain correction (with SRTM DEM) using the ESA Sentinel 1 toolbox (SNAP). Thereafter, the high-resolution Sentinel-1A/1B SAR backscatter data (both σ_{pp} and σ_{pq}), were subjected to filtering and aggregation (linear averaging) to 1 km. Before aggregation of σ_{pp} and σ_{pq} from ~25 meters to 1 km, spatial filtering (hybrid spatial filtering tool) was conducted to remove the effect of urban and manmade structures from the backscatter observations. The customized hybrid spatial filtering tool was developed at NASA JPL and is not available in the SNAP toolbox.

Several factors were addressed by the hybrid spatial filtering tool: a) the tool should not affect latency; b) the tool should remove most of the unwanted measurements; c) excessive averaging should not occur, and d) image details should be preserved. Several techniques were studied. Techniques based on standard distribution threshold were efficient but for narrow distributions they showed that some desired features could be lost. Moving window median filter techniques were also efficient in removing undesired measurements, but they were computationally expensive and produced excessive averaging when a large size window was used. To overcome all the issues mentioned above, a hybrid filter (combination of median filter and filter based on standard deviation thresholds) was implemented as follows:

1) For each 1 km2 grid cell within a given Sentinel granule the mean (m_i) and the standard deviation (s_i) of the Sentinel backscatter values were computed, $i = 1 \dots Nc$, where Nc is the number of 1 km grid cells within the Sentinel granule.

2) The tool then computed the mean standard deviation $MStd$ over all the s_i with $i = 1 \dots Nc$.

3) For all 1 km2 cells with $s_i > MStd$ a moving window median filter with a 9x9 samples window size was applied.

4) For all 1 km cells with $s_i <= MStd$, we eliminated all the Sentinel samples outside the ±1 standard deviation range [m_i - $MStd$: m_i + $MStd$] (Note that the threshold $MStd$ is used to avoid affecting areas with narrow distribution).

Figure 9 illustrates the Sentinel-1A σ_{vv} data aggregated to 1 km over Southern Iowa. The high values of σ_{vv}, as highlighted in Fig. 9A, are due to non-natural scatterers (urban areas or manmade structures). These undesired high backscatter observations were filtered for the entire Sentinel granule, and then aggregated to 1 km. The filtered Sentinel-1A σ_{vv} granule is illustrated in Fig. 9B.
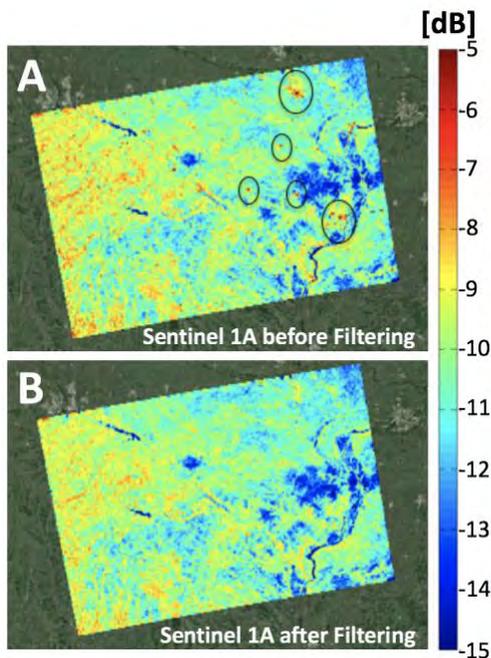

Figure 9: Sentinel-1A σ_{vv} granule from Southern Iowa on May 05, 2018. (A) σ_{vv} unprocessed data; and (B) σ_{vv} data after calibration, noise subtraction, terrain correction (using SRTM DEM), filtering, and aggregation to 1 km.

As shown in the algorithm scheme/flow (Fig. 8), the processed Sentinel-1A and Sentinel-1B data are overlapped with the SMAP observations (descending ~6:00 AM overpasses) that are closest to the Sentinel overpass within +/-24 hrs time difference. The time difference between the Sentinel-1A/1B (ascending and descending) and SMAP descending is an average of ~12 hrs. It is expected that the spatial distribution and pattern of the soil moisture does not change significantly because of the inherent memory of the soil moisture over the short period of the time difference.

The disaggregated/downscaled brightness temperature ($T_{B_p}(M_j)$) is then obtained by using the algorithm (2) on the overlapped Sentinel-1A/B (σ_{vv} and σ_{vh}) and $T_{B_v}(C)$. The implementation of (18) is conducted at 33 km resolution (C). The $T_{B_v}(C)$ values in L2_SM_P_E are gridded at/to 9 km, but keeping its inherent spatial resolution of 33 km. Therefore, the overlapped Sentinel-1A/1B data, that form a grid of 33 rows and 33 columns at 1 km resolution, is used in the process to first compute the snapshot $\beta'(C)$ and then in (18) to obtain downscaled brightness temperature $T_{B_v}(M_j)$, as illustrated in Fig 2.

The downscaled brightness temperature $T_{B_v}(M_j)$ is then injected into the tau-omega model [7, 8, 9] to retrieve surface soil moisture. Various ancillary data and lookup tables are used in the tau-omega model to retrieve soil moisture. Prominent ancillary data are NDVI climatology from MODIS, clay fraction from global soil database, and land surface temperature (LST) from NASA GMAO, and the parameters are albedo (ω), surface roughness (h), and vegetation coefficient (b) detailed for IGBP landcover classes. These ancillary data and parameters are similar to that used in the L2_SM_P/L2_SM_P_E product, however, at much finer resolutions (1 km and 3 km). The following sections describe tests of the L2_SM_SP algorithm and the L2_SM_SP soil moisture product and its characteristics.

4   TESTS OF L2_SM_SP ALGORITHM

The assessment of the L2_SM_SP was performed by comparing the disaggregated/downscaled brightness temperature with the high-resolution brightness temperature observed through an airborne platform.

Assessment of L2_SM_SP Downscaled Brightness Temperature

A primary part of the assessment for the L2_SM_SP algorithm is the comparison of disaggregated high-resolution brightness temperatures with L-band airborne remote sensing data. This assessment was done using airborne data from the SMAPEx 2015 campaign conducted in Southeastern Australia [27]. The brightness temperature data from SMAPEx 2015 has a resolution of ~1 km with varying incidence angles. For better comparison with SMAP satellite data, the SMAPEx airborne data are subjected to normalization to bring all the observations to a uniform 40° incidence angle [27]. This process introduced an error of ~4-5 K in the SMAPEx airborne data [27]. The normalized data are actually used for assessment of the L2_SM_SP disaggregated high-resolution brightness temperature. There were 2 overlapping days (May 5th, 2015 and September 13th, 2015) between the SMAP-Sentinel-1A/1B L2_SM_SP product and the Polarimetric L-band Microwave Radiometer (PLMR) airborne data from the SMAPEx field campaign. These concurrent acquisitions of data from different platforms provide an opportunity to validate the L2_SM_SP high-resolution disaggregated brightness temperatures. These specific dates of SMAPEx airborne data are also considered due to very different surface conditions in the observation domain: a) May 5th, 2015, low vegetation cover (~1 kg/m2), and b) September 13th, 2015, moderately high vegetation cover (~2.7 kg/m2). A map of the SMAPEx 2015 domain is shown in Fig. 10.
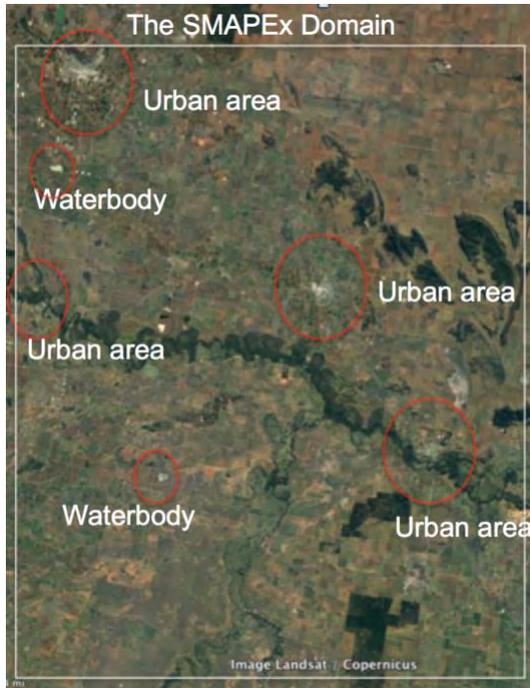
Figure 10: Study domain (white frame) of SMAPEx airborne campaign conducted in 2015 [27]. Red ovals in the figure indicate urban areas and water bodies.

As illustrated in Fig. 10, the SMAPEx study domain contains many urban areas, small manmade structures, and waterbodies. These urban areas and waterbodies were undesirable for assessment purposes. Therefore, such data need to be flagged or masked during L2_SM_SP assessments.

Figure 11a shows the PLMR airborne T_{B_v} data, Fig. 11b shows the Sentinel σ_{vv} data, and Fig. 11c shows the Sentinel σ_{vh} data from May 5, 2015 over the SMAPEx study area. It is apparent that PLMR T_{B_v} from SMAPEx are not impacted adversely by small urban areas or manmade structures, unlike the Sentinel σ_{vv} and σ_{vh} data. Figure 11b-c also show that in the Sentinel data, the large urban areas are masked and removed but the small urban areas and manmade structures are not identified and masked. These types of undesirable outliers in the Sentinel backscatter data created anomalies in the L2_SM_SP disaggregated T_{B_v} data during a first assessment. However, the combined standard deviation and median spatial filter, as discussed in Section 3.3, was successfully implemented to remove the small urban areas, manmade structures and waterbodies.
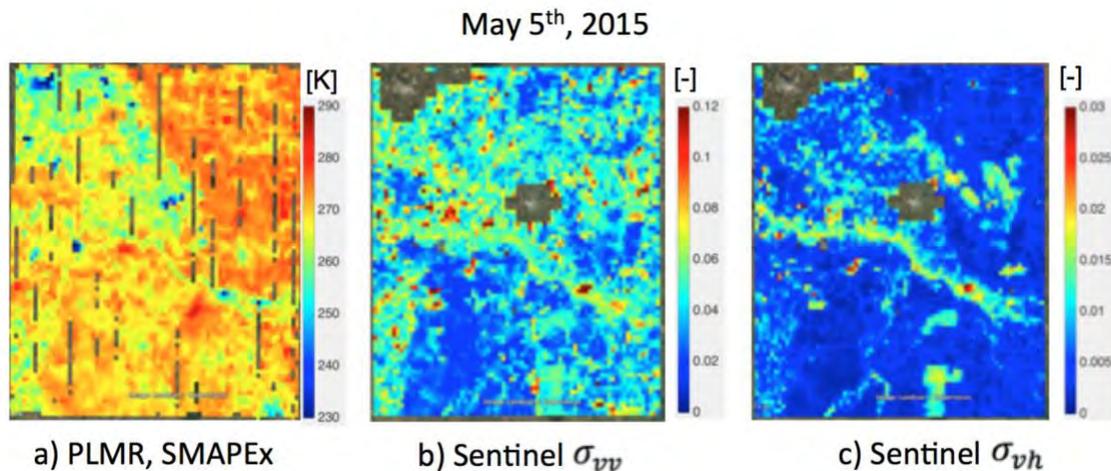
May 5th, 2015

a) PLMR, SMAPEx   b) Sentinel σ_{vv} c) Sentinel σ_{vh}

Figure 11: PLMR and the Sentinel-1A observations at EASE2 grid 1 km resolution over the SMAPEx study domain on May 5th, 2015 (prior to Sentinel filtering).

Examples of disaggregated high-resolution 3 km T_{B_v} from L2_SM_SP product are shown in Fig. 12a and Fig. 12b, and compared against the SMAPEx PLMR data and the SMAP L2_SM_P_E (T_{B_v} data corrected for presence of water) product gridded at 9 km for May 5th, 2015 and Sep. 13th, 2015, respectively. The plots in Fig. 12 show the finer details captured by the L2_SM_SP Active-Passive algorithm by incorporating the Sentinel-1A backscatter observations. In addition, the finer spatial features are very similar to the PLMR T_{B_v} data. To evaluate the SMAP-Sentinel1A/1B Active-Passive

algorithm performance, the L2_SM_SP high-resolution disaggregated T_{B_v} are compared against Minimum Performance criteria to determine the value of combining Sentinel-1A/1B SAR data with SMAP L2_SM_P_E brightness temperature data. The Minimum Performance is the SMAP L2_SM_P_E $T_{B_v}(C)$ that is applied to all the 3 km EASE2 grid cells within the overlapping 9 km EASE2 grid cell; it can be obtained by setting $\beta'(C) = 0$ in Eq. 18. Ideally, the slope and correlation between the L2_SM_SP downscaled brightness temperature and airborne high-resolution brightness temperature should be close to one (unity). In Fig. 13, we show the slope and correlation between Minimum Performance and airborne data, between L2_SM_SP and airborne data and ideal performance. In the two available airborne images (May 5th, 2015and Sep 13th, 2015) the slope and correlation between L2_SM_SP downscaled brightness temperature and airborne data are higher than the Minimum Performance. A similar analysis conducted at EASE2 grid 9 km in Fig. 13b also shows that the L2_SM_SP $T_{B_v}(M_j)$ aggregated to 9 km has better slopes and correlations when compared against L2_SM_P_E $T_{B_v}(C)$. These results (Figs. 13a and 13b) clearly demonstrate that Sentinel-1A/1B σ_{vv} and σ_{vh} data add valuable information to disaggregate the coarse-resolution L2_SM_P_E $T_{B_v}(C)$ to obtain L2_SM_SP $T_{B_p}(M_j)$ that matches better with the high-resolution spatial features observed by the SMAPEx PLMR platform.
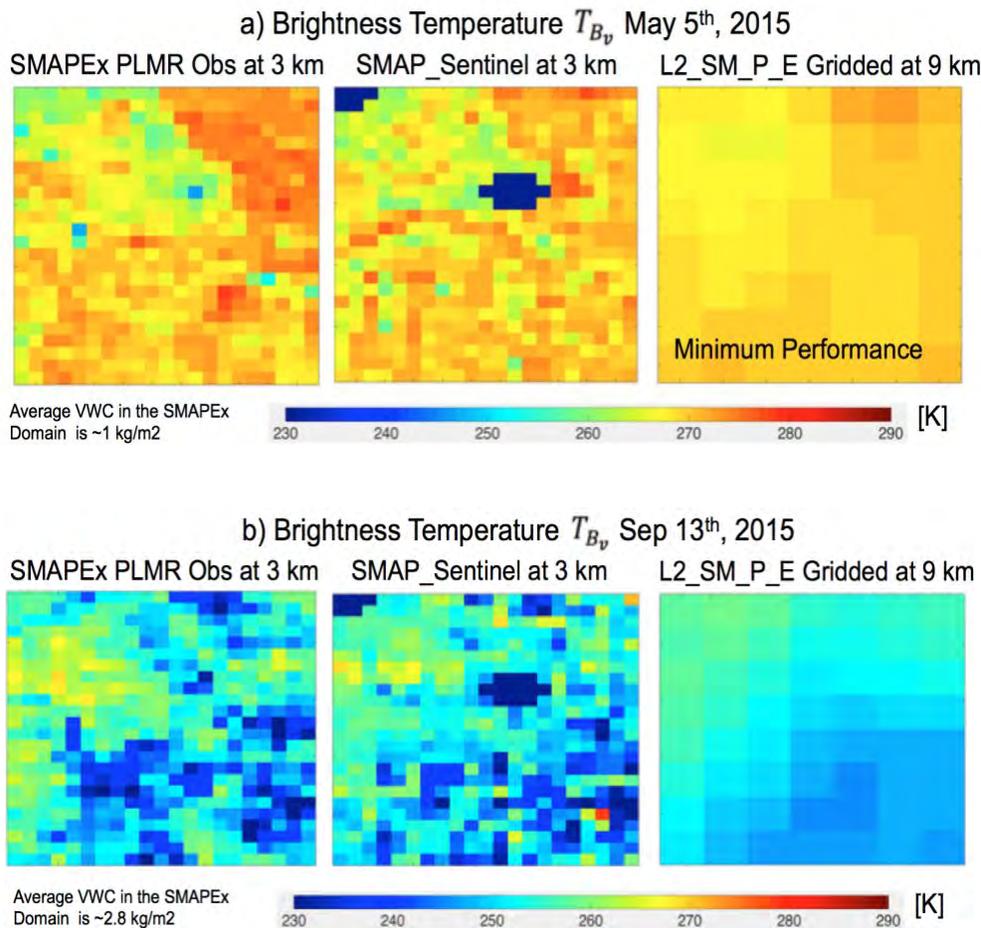

Figure 12: Output of L2_SM_SP compared against PLMR T_{B_v}data from SMAPEx and the Minimum Performance (T_{Bv} from L2_SM_P_E at 9 km).
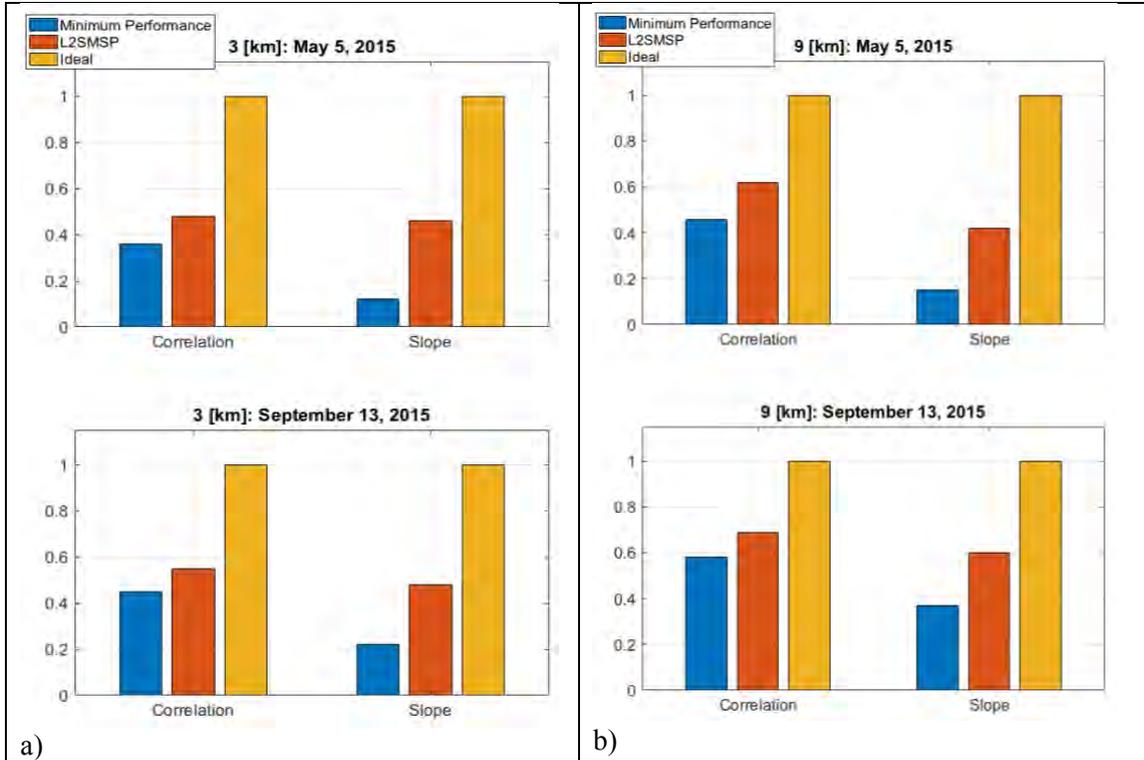
Figure 13: a) Bar plots of SMAPEx PLMR observations against L2_SM_SP T_{B_p} at 3 km and Minimum Performance (T_{Bv} from L2_SM_P_E) at 3 km. b) Bar plots of SMAPEx PLMR observations against L2_SM_SP T_{B_v} gridded at 9 km and Minimum Performance (T_{Bv} from L2_SM_P_E) gridded at 9 km.

The overall RMSE of T_{Bv} is ~3.4 K for the L2_SM_SP product and ~4.6 K for the minimum performance L2_SM_P_E at 3 km resolution, and ~2.5 K for the L2_SM_SP product and ~3.3 K for the minimum performance L2_SM_P_E at gridded 9 km resolution.

5   SMAP-SENTINEL-1 ACTIVE-PASSIVE (L2_SM_SP) PRODUCT

The L2_SM_SP data from April 15th, 2015 through current are available from the National Snow and Ice Data Center (NSIDC), the NASA Distributed Active Archive Center (DAAC) assigned to the SMAP mission at URL https://nsidc.org/data/spl2smap_s. The coverage/overlap of SMAP and Sentinel-1 platforms is from April 2015 onwards. Sentinel-1A is available from March 2015 to current, and Sentinel-1B is available from October 2016 to current. The 12-day global revisit is possible only when Sentinel-1A and Sentinel-1B are composited with the present data feed from ESA. However, over Europe the revisit interval is ~ 6 days.

5.1   Patterns and Features in the L2SMSP product

The L2_SM_SP product is available at 3 km and 1 km resolution. In this section, prior to the quantitative assessments that follow, the general features of global images are reviewed for the

L2_SM_SP product. With the current orbit configuration and data acquisition plan in the IW swath mode, the Sentinel-1A and Sentinel-1B spacecraft have a revisit interval of 6 days to 12 days at different regions of the world. Therefore, the composite of L2_SM_SP for 12 days should cover most parts of the Earth. Figure 14 shows a 12-day composite of L2_SM_SP granules from 1st May 2017 to 12th May 2017 which illustrates the global coverage between $+60°$ and $-60°$ latitudes. Figure 14 also provides a complete global extent of soil moisture evolution over different biomes and landcovers. Assessment of global soil moisture from the SMAP-Sentinel-1A/1B Active-Passive retrievals shows consistency in the soil moisture range (0.02 m3/m3 to 0.6 m3/m3), and probable values. For example, the regions that are very dry (i.e., the Sahara desert) and wet (i.e., the Amazon Basin) reflect the nature of the soil moisture distribution and expected variability as influenced by geophysical factors (soil types, vegetation, weather, and terrain) and landcovers.
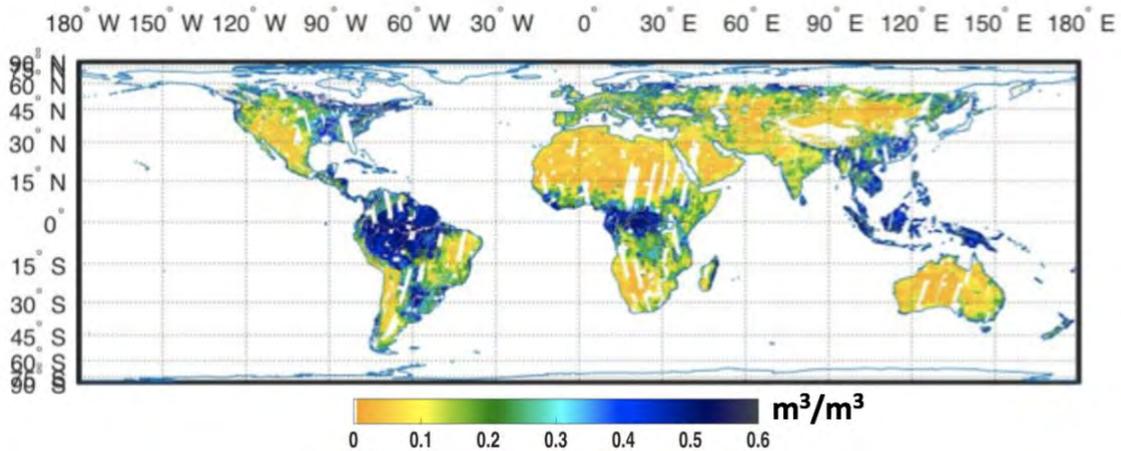

Figure 14: Twelve-day coverage of SMAP-Sentinel1 L2_SM_SP high-resolution (3 km) soil moisture data from 1st May, 2017 to 12th May, 2017.

There are a number of quality flags that are applied to L2_SM_SP products. These flags imply that the data should be used with caution while others indicate that the data should not be used in any geophysical application. A complete description of the flags and flag thresholds used in L2_SM_SP processing can be found in the L2_SM_SP Product Specification Document available at NSIDC [(https://nsidc.org/sites/nsidc.org/files/technical-references/SMAP%20L2_SM_SP%20PSD_20180531.pdf)]. The reliability of soil moisture retrieval algorithms is known to decrease when the VWC exceeds a certain threshold. For the L2_SM_SP product, a 3 kg/m2 VWC value is used as a flag threshold to indicate areas of moderate vegetation where soil moisture retrievals are possibly less accurate (in comparison, the SMAP passive-only L2SMP and L2SMP_E products use a VWC threshold of 5 kg/m2). A quality flag value of 0 represents good quality and any value greater than 0 represents substandard quality due to surface flags or due to a quality flag associated with the disaggregated T_{B_p} or due to the quality of the input data ($T_{B_p}(C)$ and σ_{pp} and σ_{pq}. A surface flag is also associated with each and every soil moisture retrieval data field. The surface flags are stored in two byte integers. There are 16 bits in each two byte integer. For example, the first bit position indicates presence of a waterbody. The first bit position is set to 0 if the water fraction is less than or equal to a threshold value ($<= 0.1$); otherwise, the first bit position is set to 1 if the water fraction is greater than the threshold value (> 0.1). Similarly, the other bits are assigned 0 or 1 based on the threshold values of urban area, mountainous region, VWC, etc.

It is anticipated that some of the flag thresholds may be relaxed in time as the algorithms are improved for the presence of certain currently problematic surface conditions. Other areas that are

flagged include regions with varied topography features (for example, mountain ranges) and near large water bodies (coastal regions and areas near large lakes).

   The variability within the radiometer coarse grid cell is mostly due to soil moisture, vegetation and soil roughness [7, 8], and is captured by high-resolution Sentinel-1A/Sentinel-1B backscatter values of σ_{vv} and σ_{vh} at the finest available resolution (in this case at ~1 km). For illustration, Fig. 15 shows the primary inputs to the algorithm, the brightness temperature $T_{B_v}(C)$ values in L2_SM_P_E gridded at 9 km (~33 km resolution) and the Sentinel-1A/1B processed σ_{vv} and σ_{vh} backscatter data at 1 km. Figure 16 illustrates the L2_SM_SP algorithm's capability to capture high-resolution spatial features of soil moisture possible through Sentinel-1A/1B backscatter observations (Fig. 15b-15c) that disaggregate the brightness temperature $T_{B_v}(C)$ values (Fig. 15a). Figure 17 shows another perspective to highlight the dynamic range of brightness temperature and soil moisture present in the SMAP-based soil moisture products. The plot clearly shows the increase in variability and dynamic range in the L2_SM_SP product at 3 km and 1 km resolution when compared to the 9 km gridded L2_SM_P_E data that have an effective resolution of ~33 km.

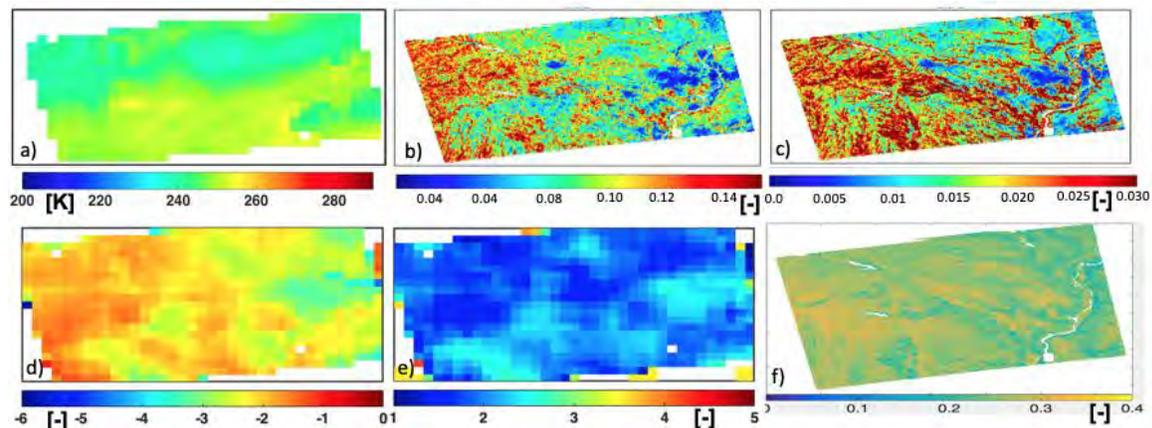**Figure 15**: An example of primary inputs to the SMAP-Sentinel Active-Passive algorithm. From Southern Iowa, 5th May, 2018: a)SMAP radiometer brightness temperature $T_{B_v}(C)$ at about 33 km resolution but gridded at 9 km; b) Sentinel-1A/B co-polarized backscatter (σ_{vv}) at 1 km; c) Sentinel-1A/B cross-polarized backscatter (σ_{vh})at 1 km; d) Parameter $\beta'(C)$; e) Parameter Γ, and; f) Clay fraction.
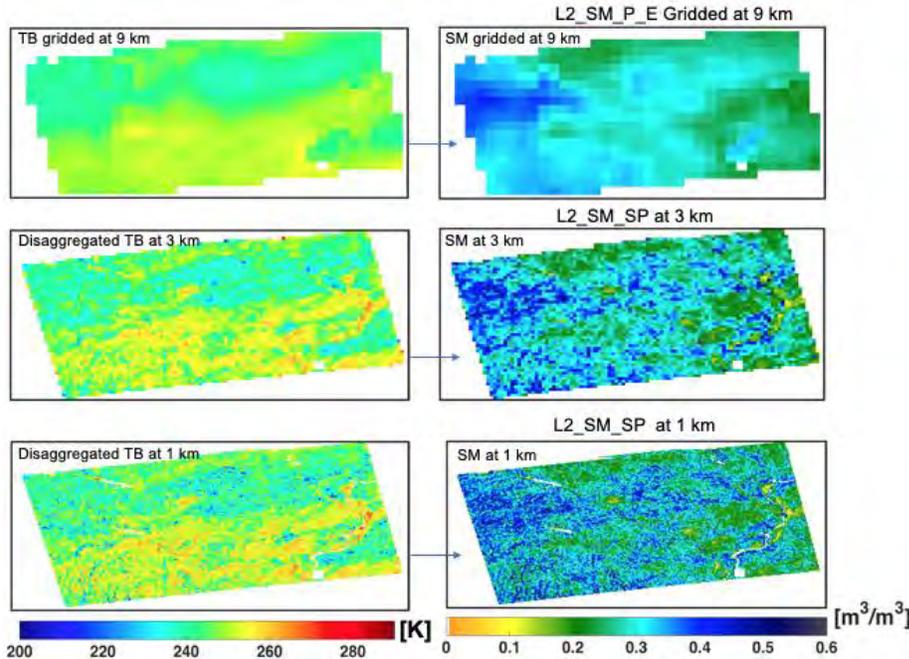
Figure 16: Comparison of L2_SM_SP product at 1 km and 3 km resolution with the corresponding L2_SM_P_E product gridded at 9 km. The L2_SM_SP image is from Southern Iowa, May 5th, 2018.
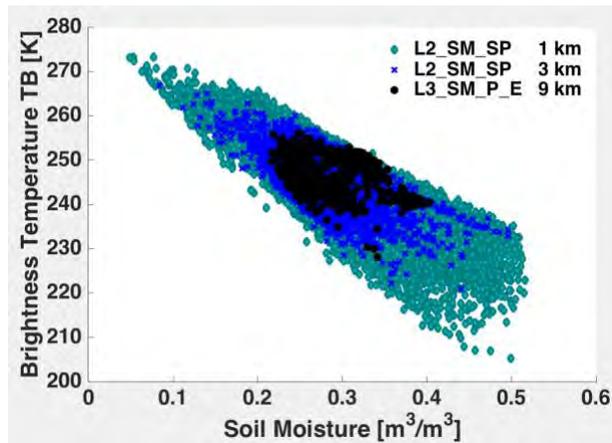
Figure 17: Distribution of data in the soil moisture and the brightness temperature space for the L2_SM_SP product at 1 km and 3 km, and L2_SM_P_E product gridded at 9 km (~33 km effective resolution) over the Southern part of Iowa on 5th May, 2018.

5.2   Error Budget of L2_SM_SP Algorithm

The algorithm (Section 3.2) error budget is estimated using the basic input uncertainty and parameter perturbations. The input data are the water body corrected brightness temperature and the radar backscatter cross-sections that are averaged to 3 km. Radar pixels that include water bodies are excluded.

Table 3 lists the various contributions to the disaggregated brightness temperature at 3 km resulting from the Active-Passive algorithm. The first numbered row is the estimated error in the coarse resolution radiometer brightness temperature which is due to the instrument, geophysical contributions of Earth, and gridding. Effects of water bodies are removed from the brightness

temperature. Assuming a nominal 5% error in the estimation of inland water bodies, the estimated contribution of error is about 0.45 K. The errors due to mis-specification of inland water bodies are dependent on the absolute percent of water fraction. A 5% error is assumed on a condition with 5% water body fraction. It should be noted that this source of error can be very large. For example, if a pixel contains 10% inland water and there is 10% error on its specification, the impact on brightness temperature correction can be as large as ~2.0 K uncertainty. As a nominal case, 5% error on 5% water coverage is considered. The permanent water bodies within a radiometer pixel are estimated from existing data such as the Shuttle Radar Topography Mission (SRTM).

The waterbody-adjusted brightness temperature root-sum-of-squares (RSS) is reported in row three of Table 3. The baseline algorithm uses the radar backscatter cross-section and brightness temperature time-series to estimate a disaggregated 3 km brightness temperature. The radar noise and uncertainty contributions to the disaggregated 3 km brightness temperature are estimated to be 2.5 K based on the algorithm models. This uncertainty is shown in row four of Table 3 error budget. Beside radar backscatter cross-calibration and contamination noise, other important sources of errors in Eq. 17 are the uncertainties in algorithm parameters (β' and Γ). Nominal values of 20% uncertainties are used for the algorithm parameters to evaluate the error contribution in the disaggregated 3 km brightness temperature, and the estimated value is 1.89 K (shown in row 5 of Table 3). The total 3 km disaggregated brightness temperature error of ~3.4 K is shown as an RSS in the sixth row of Table 3.

Table 3: Error budget in degrees Kelvin.

	Error Sources at 36 km EASE2 Grid	Estimated Error
1	Radiometer precision and calibration stability, Faraday rotation, atmospheric gases, non-precipitating clouds, and gridding	1.3* K
2	Waterbody fraction surface heterogeneity 5% error	0.45 K
3	**Adjusted Corrected** T_B **RSS**	**1.47 K**
4	Radar calibration and contamination error	2.5 K
5	Algorithm Parameter error	1.89 K
6	**Disaggregated** T_B **(9 km) estimation RSS**	**3.42 K**

*T_B error requirement of 1.3 K is based on a 30 km swath grid.

6   CALIBRATION AND VALIDATION OF L2_SM_SP PRODUCT

The calibration/validation (cal/val) plan for the L2_SM_SP product has heritage from the SMAP-only Active-Passive product L2_SM_AP [22] and [24]. The L2_SM_SP product cal/val was performed using two different approaches: 1) by comparing the soil moisture retrievals at 3 km from L2_SM_SP against upscaled *in situ* soil moisture observations from the core sites.; and 2) comparing the soil moisture retrievals from L2_SM_SP at 1 km against soil moisture measurements from sparse networks.

Core Validation Sites: The Core Validation Sites (CVS) are of primary importance. The CVS provide spatial averages of soil moisture at 3 km spatial resolution with adequate replications, with minimal latency, and with well-known error quantification (verified against gravimetric measurements). NASA has established agreements with many partners to provide CVS data. There are nearly ~30 candidate sites [24]; out of these, a few are selected for CVS based on a strict set of requirements and their capability to provide spatial average of soil moisture at 3 km. The CVS sites for L2_SM_SP primary validation also come from various land covers and biomes.

Sparse Network: Sparse soil moisture measurement networks exist in the United States and other parts of the world. The limitation of such networks is low measurement density that results in one measurement site within a 3 km grid cell. It is challenging and difficult to use measurements from one *in situ* site within a 3 km grid cell for validation. Therefore, research to identify strategic measurement sites that are temporally stable (least affected and manipulated over a longer time period) and that represent the grid average are critical for the success of extensive validation of the L2_SM_SP product. Examples of sparse measurement networks in the United States are the USDA Soil Climate Analysis Network (SCAN), the NOAA Climate Research Network (CRN), the Oklahoma Mesonet, SoilSCAPE network at Tonzi Ranch, California and Canton, Oklahoma, and the Illinois Soil Moisture Data from the Illinois State Water Survey. Access to sparse measurement networks located outside the United States is also important for global validation efforts. Data from soil moisture networks in Canada, Mongolia, China, Australia, and Europe are accessible through the Global Soil Moisture Data bank. To rationally use the sparse measurement network, cal/val of L2_SM_SP has to deal with data latency, verification of calibration, and spatial scaling. At present cal/val relies primarily on CVS and only secondarily on sparse networks.

Complete details of cal/val approaches are found in the SMAP Calibration and Validation Plan (SMAP Science Data Calibration and Validation Plan. SMAP Project, JPL D-52544, Jet Propulsion Laboratory, Pasadena, CA).

6.1   L2_SM_SP Validation using the 3 km Core Site data

The SMAP L2_SM_SP product validation was based primarily on comparison of retrievals with *in situ* soil moisture measurements [28]. The *in situ* measurements for the top ~5 cm soil moisture from networks with an acceptable sensor density within a 3 km EASE2 grid are the primary validation locations for the L2_SM_SP product. The SMAP project collaborated with various partners from around the world to identify such locations and established CVS [28]. These CVS have been verified as providing a spatial average of soil moisture at 3 km (with at least 3 *in situ* sites) and 9 km (with at least 5 *in situ* sites) spatial resolution. However, the spatial averages of soil moisture from CVS are not without issues because of inherent upscaling errors. Table 4 lists the CVS used for validation of the L2_SM_SP product. Beside the CVS, sparse networks [29] were also used as a supporting tool/option to validate the L2_SM_SP product.

Table 4: SMAP Cal/Val Partner Sites Providing Validation Data for the L2_SM_SP product

Site Name	Site PI	Area	Climate regime	IGBP Land Cover	Status
Walnut Gulch***	C. Holifield Collins	USA (Arizona)	Arid	Shrub open	Valid for 3 km and 9 km
Fort Cobb**	P. Starks	USA (Oklahoma)	Temperate	Grasslands	Valid for 9 km
Little Washita**	P. Starks	USA (Oklahoma)	Temperate	Grasslands	Valid for 9 km
South Fork**	M. Cosh	USA (Iowa)	Cold	Croplands	Valid for 9 km

				Cropland/natural mosaic	
Little River**	D. Bosch	USA (Georgia)	Temperate		Valid for 9 km
TxSON***	T. Caldwell	USA (Texas)	Temperate	Grasslands	Valid for 3 km and 9 km
Kenaston***	A. Berg	Canada	Cold	Croplands	Valid for 9 km
Carman**	H. McNairn	Canada	Cold	Croplands	Valid for 9 km
Monte Buey***	M. Thibeault	Argentina	Arid	Croplands	Valid for 3 km and 9 km
REMEDHUS***	J. Martinez	Spain	Temperate	Croplands	Valid for 3 km and 9 km
Valencia***	E. Lopez-Beaza	Spain	Arid	Shrub (open)	Valid for 3 km and 9 km
St Josephs**	M. Cosh	USA (Indiana)	Cold	Croplands	Valid for 9 km
Yanco***	J. Walker	Australia	Arid	Croplands	Valid for 3 km and 9 km
	=CVS used in assessment at 9 km, * = CVS used for both 3 km and 9 km.				

The *in situ* data obtained from the SMAP Cal/Val Partner sites (Table 4) are subjected to quality control (QC) before using them to validate the SMAP products. A QC software tool was developed at JPL using the approached presented in [30] for QC of the *in situ* soil moisture data. Figures 18-21 illustrate time series and scatter plot comparisons of the L2_SM_SP product at 3 km grid cells against four CVS: TxSON, Monte Buey, Valencia, and Yanco. A total number of twelve 3 km grid cells from the 7 CVS were used to compute statistics for primary validation of the L2_SM_SP product. Table 5 shows the performance statistics/metrics for all the CVS used for validation. The time series plot in Fig. 18 for the TxSON site shows a good match between soil moisture trends, with some bias in soil moisture estimation compared to *in situ* measurements that is possible due to differences in soil texture used in the retrieval process. The performance of the L2_SM_SP product over most of the CVS with non-crop land covers is reasonable as illustrated in Fig. 18 for TxSON and Fig. 19 for Valencia. However, the performance of the L2_SM_SP over CVS with crop cover is inferior, as shown in Fig. 20, possibly because of being out of sync with the vegetation attribute information and strong C-band interactions with vegetation which might cause patterns not totally attributable to soil moisture but to vegetation cover instead. The retrieval process uses Vegetation Water Content (VWC) derived from NDVI climatology (developed from 10 years of MODIS data), which might lead to a mismatch with the actual status of VWC. Therefore, it is likely that in Fig. 20 (Monte Buey CVS) the lack of a consistent bias and higher errors may be caused by this type of mismatch. In Figs. 18-21 red color for in situ data represents good quality, and the purple color is when the in situ data quality is not satisfactory. The black dots are the L2_SM_SP data used in the scatter plot and computation of RMSE. The grey dots are the L2_SM_SP data that match on a given day with the inferior quality in situ data and are not used in calculation of the RMSE and R values.
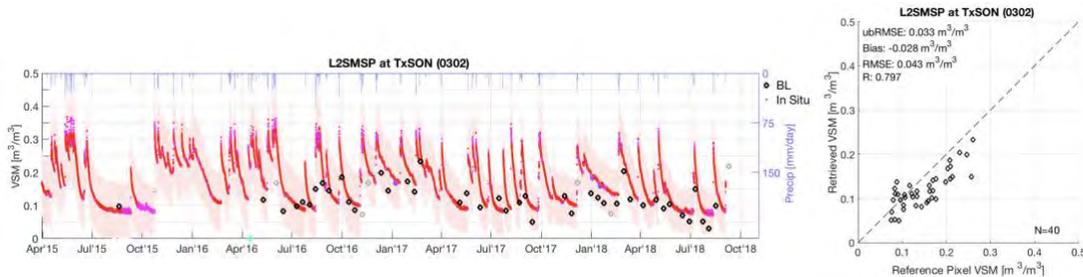

Figure 18: L2_SM_SP assessment at 3 km (40 data point) for TxSON, Texas, USA. (BL: L2_SM_SP)
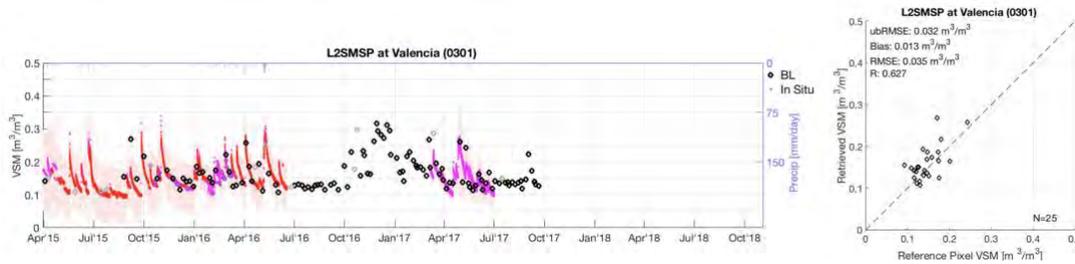
Figure 19: L2_SM_SP assessment at 3 km (25 data point) for Valencia, Spain. (BL: L2_SM_SP)
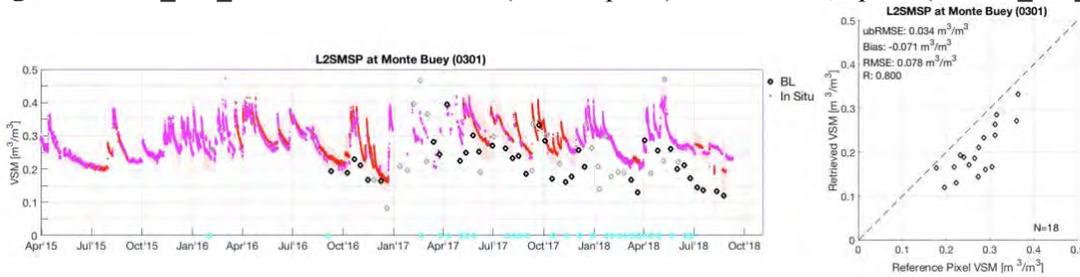

Figure 20: L2_SM_SP assessment at 3 km (18 data point) for Monte Buey, Argentina. (BL: L2_SM_SP)
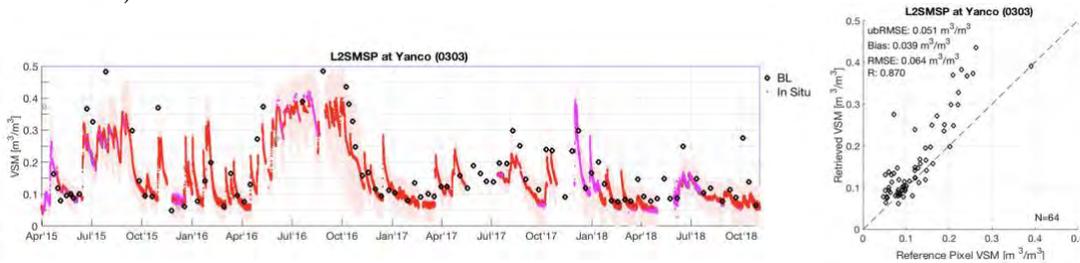
Figure 21: L2_SM_SP Assessment at 3 km (55 data point) for Yanco, Australia. (BL: L2_SM_SP)

The validation results at 3 km resolution in Figs. 18-21 and Table 5 come from a very limited number of CVS. Thus, another strategy was developed to overcome this limitation: an upscaled L2_SM_SP product at 9 km is constructed by aggregating all nine L2_SM_SP 3 km EASE2 grid cells within the 9 km EASE2 grid. The upscaled 9 km product is then used for the 17 CVS already established and operating for the SMAP-only Active-Passive L2_SM_AP 9 km product [19]. This approach optimizes the CVS usage and to assess the performance of the spatially upscaled L2_SM_SP 3 km product at 9 km spatial resolution. The results and performance of the upscaled L2_SM_SP product at 9 km in Table 6 are encouraging. The overall ubRMSE of 0.036 m_3/m_3 for the L2_SM_SP product meets the SMAP mission accuracy goal of 0.04 m_3/m_3 previously established as a benchmark for the SMAP-only L2_SM_AP product. In Table 6 most of the R-values are relatively high (R>0.829), indicating a considerable match between estimates and *in situ* measurements.

Table 5: SMAP L2_SM_SP Assessment Statistics against CVS at 3 km.

Site name	ubRMSE	Bias	RMSE	R	#Samples
Walnut Gulch	0.033	0.033	0.047	0.950	23
Walnut Gulch	0.029	0.063	0.069	0.929	21
TxSON	0.041	-0.039	0.056	0.895	24
TxSON	0.033	-0.028	0.043	0.797	40
Kenaston	0.065	-0.052	0.083	0.216	31

Kenaston	0.053	-0.047	0.071	0.603	24
Monte Buey	0.034	-0.071	0.078	0.800	20
Valencia	0.032	0.013	0.035	0.627	25
Yanco	0.082	0.013	0.085	0.489	38
Yanco	0.075	0.018	0.077	0.635	40
Yanco	0.048	0.037	0.060	0.898	64
Yanco	0.065	0.070	0.096	0.761	48
SMAP Average	**0.049**	**0.001**	**0.067**	**0.717**	**Total: 398**

The term RMSE in the analysis is interchangeably used for root-mean-square-difference (RMSD). Nearly 30 to 50 time-matching samples are found in core sites and are used in computing the statistics.

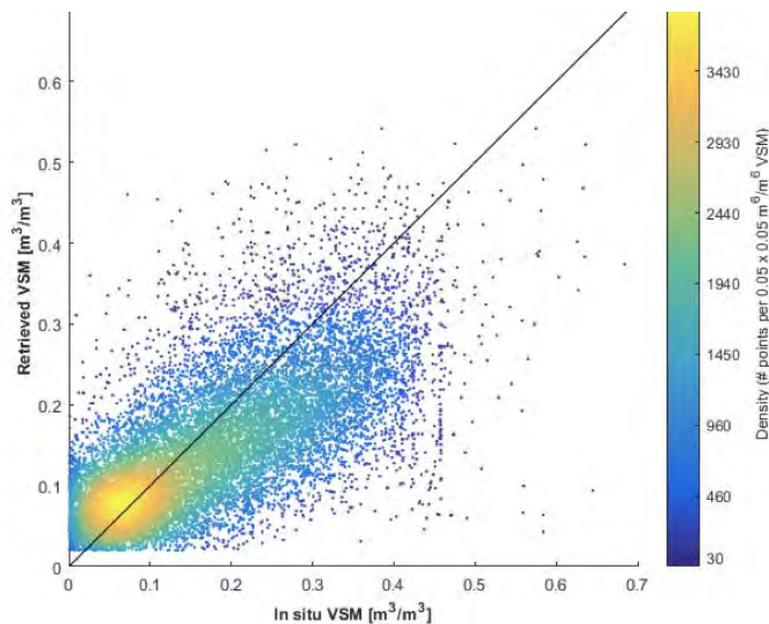**Table 6**: SMAP L2_SM_SP Assessment Statistics against CVS measurements at 9 km.

Site name	ubRMSE	Bias	RMSE	R	#Samples
Walnut Gulch	0.025	0.022	0.033	0.871	37
Walnut Gulch	0.028	0.049	0.056	0.863	47
TxSON	0.022	0.010	0.025	0.882	23
TxSON	0.030	0.012	0.032	0.904	42
Fort Cobb	0.030	-0.023	0.038	0.847	48
Little Washita	0.039	-0.032	0.051	0.771	93
South Fork	0.060	-0.031	0.067	0.802	39
St Josephs	0.022	-0.042	0.048	0.913	24
Little River	0.030	0.086	0.091	0.799	22
Kenaston	0.038	-0.061	0.072	0.764	28
Kenaston	0.031	-0.074	0.080	0.836	27
Carman	0.049	-0.069	0.085	0.590	20
Monte Buey	0.014	-0.049	0.051	0.967	23
REMEDHUS	0.059	0.112	0.126	0.831	63
Valencia	0.027	0.012	0.029	0.746	24
Yanco	0.055	-0.005	0.055	0.877	44
Yanco	0.049	0.037	0.061	0.831	67
SMAP Average	**0.036**	**-0.003**	**0.059**	**0.829**	**Total: 671**

6.2 L2_SM_SP Validation at 1 km using the Sparse Network

The intensive CVS validation performed for the SMAP L2_SM_SP product can be complemented by sparse networks as well as by newly developed soil moisture networks. The important difference in interpreting these data is that they involve only one *in situ* point in a grid cell. Thus, whatever reservations might exist on the upscaling of CVS *in situ* measurements to the coarser resolution cells of remote sensing products, these reservations might be of even greater concern for sparse networks of *in situ* soil moisture measurements. However, sparse networks do offer many sites in different environments for comparison.

The established sparse soil moisture networks utilized for the SMAP L2_SM_SP product comparison are the NOAA Climate Reference Network (CRN), the USDA NRCS Soil Climate Analysis Network (SCAN), the Oklahoma Mesonet, the MAHASRI network (in Mongolia), the SMOSMania network (in southwest Europe), the Pampas network (in Argentina), and soil moisture estimates derived from surface reflectance at Global Position Stations (in the western US). From these sparse soil moisture networks, ~375 sites were found to be suitable for direct comparison with the SMAP L2_SM_SP overlapping grid cells. The ~375 sites were selected based on *in situ* measurement data quality and continuity of the observations during the ~3 year period (April, 2015 to Oct, 2018). The defining feature of these networks are the low spatial density of *in situ* measurement locations that usually resulted in one point per L2_SM_SP 3 km and 1 km grid cells. This leads to large upscaling errors due to spatial representativeness and the inability of a single *in situ* site location to represent mean soil moisture within a 3 km or 1 km grid cell. However, despite this scaling bias, sparse networks can adequately describe relative errors.

Figure 22A-B illustrates comparisons of the L2_SM_SP product retrievals with the measurements available from ~375 *in situ* sparse networks from many different land covers at 3 km and 1 km,. Despite the potential errors associated with spatial representativeness, the agreement between the *in situ* soil moisture and the L2_SM_SP is reasonably good (see Table 7). The ubRMSE and bias values obtained from these sparse networks are similar to those obtained from the CVS. These results (Fig. 22) provide further confidence in the previous conclusions based on the CVS.

A)
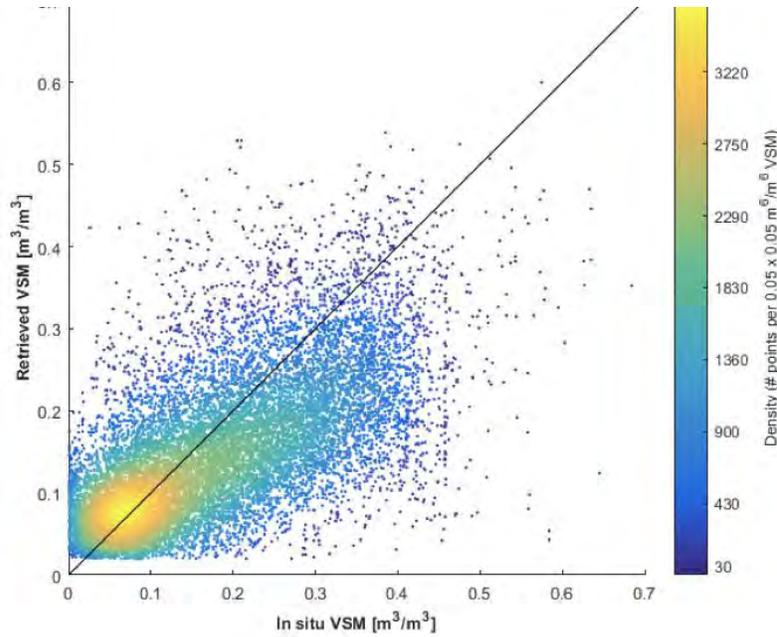
B)
Figure 22: Results of comparison between L2_SM_SP with the sparse network sites (~375 *in situ* sites): A) at 3 km resolution; and B) at 1 km resolution.

Table 7: SMAP L2_SM_SP Assessment Statistics against Sparse Networks at 3 km and 1 km resolution.

L2_SM_SP (3 km)	ubRMSE [m³/m³]	Bias [m³/m³]	RMSE [m³/m³]	R [-]	N
Open shrublands	0.04	0.017	0.045	0.506	34
Woody savannas	0.053	0.031	0.063	0.657	4
Savannas	0.04	-0.001	0.06	0.789	6
Grasslands	0.051	-0.032	0.064	0.647	230
Croplands	0.072	-0.033	0.087	0.531	69
Crop/Natural Vegetation Mosaic	0.067	-0.023	0.076	0.469	14
Barren/Sparse	0.026	0.031	0.04	0.514	9
Average	**0.05**	**-0.01**	**0.062**	**0.587**	**370**

L2_SM_SP (1 km)	ubRMSE [m³/m³]	Bias [m³/m³]	RMSE [m³/m³]	R [-]	N
Open shrublands	0.046	0.008	0.046	0.544	43
Woody savannas	0.056	-0.001	0.065	0.489	7
Savannas	0.038	0.016	0.061	0.827	4
Grasslands	0.06	-0.036	0.069	0.647	236
Croplands	0.076	-0.041	0.094	0.468	80
Crop/Natural Vegetation Mosaic	0.068	-0.008	0.077	0.349	8
Barren/Sparse	0.023	0.018	0.036	0.592	6
Average	**0.052**	**-0.028**	**0.064**	**0.548**	**384**

6.3   Ancillary Data

The data required other than SMAP and Sentinel-1A/1B observations to process and retrieve near surface soil moisture are termed as ancillary data. The soil moisture estimates are retrieved from

the disaggregated (downscaled) T_B at 3 km and 1 km spatial resolutions obtained from the Active-Passive algorithm (L2_SM_SP). Therefore, the required ancillary data needed to process the disaggregated T_B are also required at 3 km and 1 km spatial resolution. The L2_SM_SP process needs both static and dynamic ancillary data. Static ancillary data are those data that normally do not change during the mission lifetime. The L2_SM_SP process needs permanent masks and geophysical parameters as static ancillary data. Permanent masks of water bodies, mountainous regions, forest cover (VWC > 3 kg/m2), urban areas, and land area at high-resolution (1 km and 3 km) resampled in Earth fixed grid is desirable for L2_SM_SP processing and quality control. Sand fraction and clay fraction data at 1 km spatial resolution are examples of static ancillary data essential to retrieve soil moisture. The L2_SM_SP process uses static ancillary data archived by the SMAP mission's Science Data System (SDS) at JPL. Dynamic ancillary data pose more challenges because they need frequent updates (daily, biweekly, monthly, and seasonally). The dynamic data required for L2_SM_SP processing are effective surface soil temperature, VWC (for vegetation opacity τ), and land use-land cover.

Ancillary data from various resources were analyzed and selected as baseline ancillary data. SMAP ancillary data reports (https://nsidc.org/data/smap/smap-ancillary-products) were written for the individual ancillary data listed in Table 8. These reports document the rationale for the choice of the primary source of the ancillary data, and have been made available to the public. Table 8 lists most of the static and dynamic ancillary data, data source, updating frequency and desired spatial resolution required for the L2_SM_SP retrieval. The amount and type of ancillary data needed are dependent to some extent on the choice of the specific spatial resolution (3 km or 1 km). Some examples of the ancillary data used in L2_SM_SP retrieval are shown in Fig. 23-25.

Table 8: Ancillary data required to produce the SMAP L2_SM_AP product.

Parameter	Updating Frequency	Grid Resolution	Data Type	Data Source
%Sand and %Clay	Done once	3 km and 1 km	Static	Composite of soil databases (HWSD, FAO, ASRIS, STATSGO, NSDC)
Soil Texture	Done once	3 km and 1 km	Static	Composite of soil databases (HWSD, FAO, ASRIS, STATSGO, NSDC)
Vegetation (b and ω)	1-2 weeks	3 km and 1 km	Static	Values obtained from L2_SM_P ATBD look-up table
Vegetation Opacity (τ)	Daily/Weekly	3 km and 1 km	Dynamic	MODIS 1 km NDVI converted to VWC and then to $\tau = (b*VWC)$
Roughness (h)	Monthly	3 km and 1 km	Static	L2_SM_P ATBD look-up table
Effective soil temperature	Daily	9 km	Dynamic	ECMWF / MERRA (TBC)
Waterbodies (permanent)	Yearly	3 km and 1 km	Static	MOD44W – a MODIS static open water product
Precipitation flag	Daily	9 km	Dynamic	ECMWF total precipitation forecast (or GPM)
Snow/Ice flag	Daily	9 km	Dynamic	NOAA Snow and Ice Mapping System (IMS) Product
Mountainous flag	Done once	3 km and 1 km	Static	SRTM and GTOPO30 derived elevation, slope, range and variance
Urban area fraction	Done one	3 km and 1 km	Static	GRUMP data
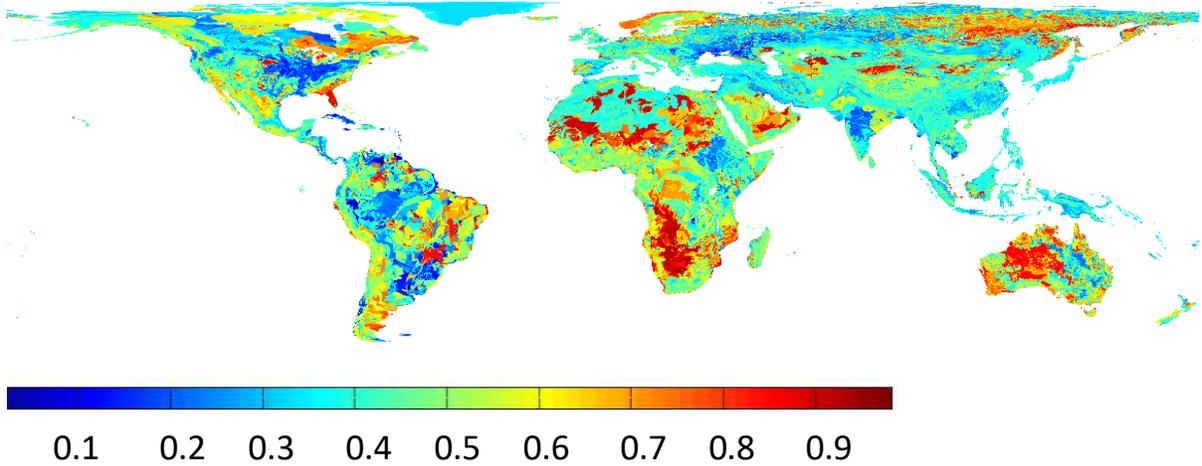
Figure 23: Global sand fraction of top soil at 3 km EASE2 grid projection.
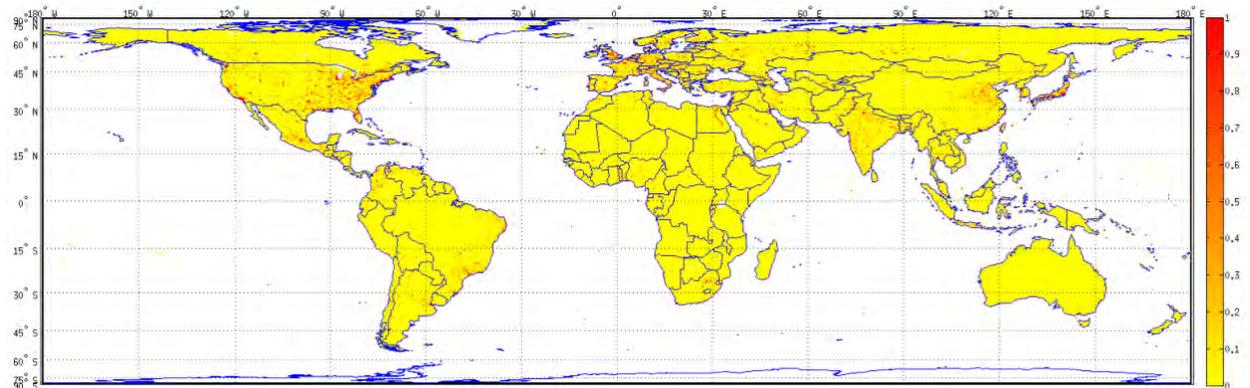

Figure 24: Urban extent fraction gridded at 3 km EASE2 grid projection.
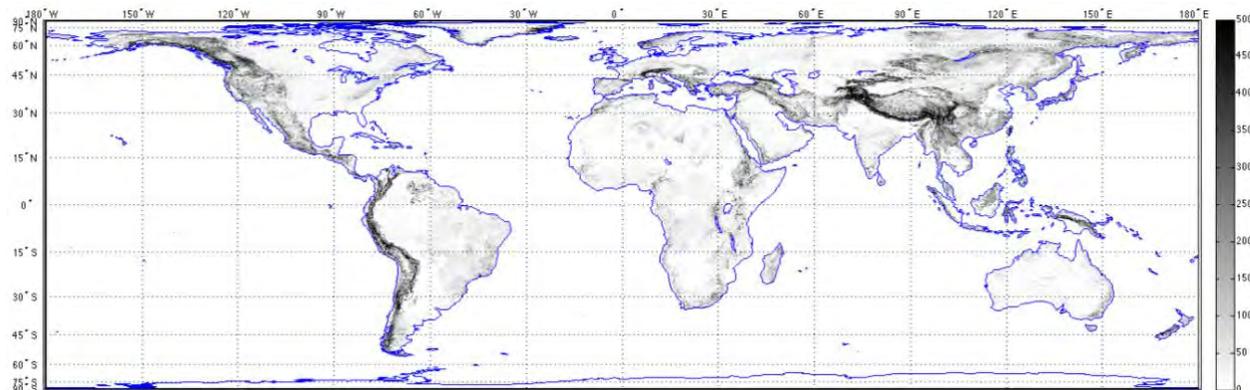

Figure 25: Standard deviation in DEM gridded at 3 km EASE2 grid projection.

Table 8 also lists the parameters (b, ω, and h) essential for soil moisture retrieval using downscaled T_B (3 km and 1 km). These parameters are dependent on the land cover types. Details of

these parameters are available in the L2_SM_P ATBD [ref?] and the associated ancillary data report [ref?].

It is important to be aware of inherent errors in the ancillary data parameters and the latency involved in acquiring dynamic ancillary data. Errors present in the ancillary data affect the performance of the L2_SM_SP algorithm and ultimately the accuracy of the output product. This highlights the fact that the L2_SM_SP algorithm should be robust to accommodate the expected inherent errors in the ancillary data. Results mentioned in the L2_SM _AP ATBD from the Monte Carlo study conducted for the Active-Passive algorithm show the effects and robustness due to errors in ancillary data.

The issue of latency involved in getting the dynamic ancillary data is critical for the timely delivery of SMAP baseline products to the DAAC. However, for the L2_SM_SP product, there are no latency requirements. Current assessment indicates that there are no major obstacles based on the ancillary data latency that would prevent the L2_SM_SP product generation after receiving the Sentinel-1A/1B granules from the European Space Agency (ESA) through the NASA Earth Science Data Gateway at the Goddard Space Flight Center (GSFC).

6.4   Quality Control and Diagnostics

Quality control (QC) is an integral part of the L2_SM_SP processing. The QC steps of L2_SM_SP processing are based on the flags that are provided with the input data streams (L2_SM_P_E, and Sentinel-1A/1B), different types of masks, flags, and fractional coverage of other variables derived from ancillary data (Table 8). The SDS at JPL processes all data from the input data streams (L2_SM_P_E, and Sentinel-1A/1B) that have favorable conditions for soil moisture retrieval (VWC <= 3 kg/m2, no rain, no snow cover, no frozen ground, no RFI, sufficient distance from open water) to generate the L2_SM_SP product. However, soil moisture retrieval will also be conducted for regions with VWC > 3 kg/m2, rain, RFI repaired data, and places closer to water bodies, but appropriate flags are added to these data points indicating their susceptibility to potentially high errors. The L2_SM_SP Product Specification Document elaborates the fields for QC bit flags. A flow diagram in Fig. 26 illustrates the decision tree to perform L2_SM_SP retrieval.

As shown in Fig. 26, the L2_SM_SP processing involves merging of two data streams, i.e., L2_SM_P_E and Sentinel-1A/1B. Therefore, the QC of L2_SM_SP output is influenced by these input data streams. In other words, the QC flags of the L2_SM_SP output are the union of QC flags from L2_SM_P_E and Sentinel-1A/1B data streams. However, due to differences in spatial resolution of the inputs (L2_SM_P_E and Sentinel-1A/1B) and output (L2_SM_SP), the assignment of QC flags in L2_SM_SP may differ from the flags associated with the inputs. The thresholds of ancillary data that initiate flagging in the L2_SM_SP product are mentioned below. For example, T_{B_p} data in L2_SM_P_E are corrected for the presence of water bodies. Studies were conducted to assess the quality of corrected T_{B_p} data that are acceptable and within the desired uncertainty level that could be used in L2_SM_SP processing. The water body fraction is reported for all land-based 3 km and 1 km grid cells in the L2_SM_SP product file, and the water body flag bit is set in the retrieval quality field if the water body fraction is greater than a threshold value. In the case of VWC, L2_SM_SP retrieval is performed at all the grid cells irrespective of VWC, but the QC flag is set only for grid cells having VWC > 3 kg/m2. No retrievals are performed for L2_SM_SP grid cells that are associated with RFI, water body fraction above a particular threshold, frozen ground, snow, and urban fraction above threshold.
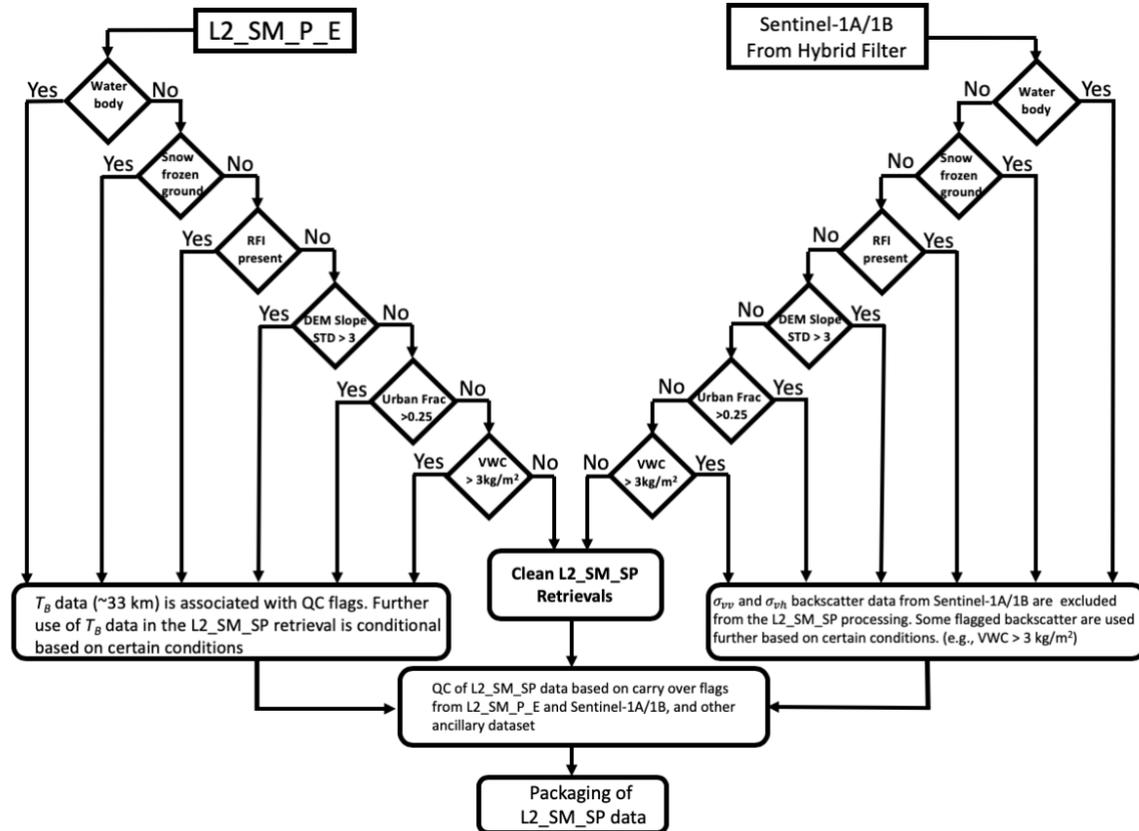
Figure 26: Decision Tree to QC σ_{vv} and σ_{vh} from Sentinel-1A/1B and T_{B_v} from SMAP.

Thresholds from masks that initiate flags and operational decisions are available in L2_SM_SP processing. The following thresholds are used to QC SMAP and Sentinel-1A/1B input streams to the L2_SM_SP processing. The thresholds from masks also initiate quality flag and surface flags in the L2_SM_SP product.

Open water body: The open water fraction map derived from the MODIS (MOD44W) database at 1 km resolution is used to flag and mask the L2_SM_P_E and the Sentinel-1A/1B input stream. This following thresholds are used to flag the L2_SM_P_E and the Sentinel-1A/1B grid cells during the L2_SM_SP soil moisture retrieval processing in the following way:

Masking and flagging of L2_SM_P_E grid cell:
        Water fraction is 0.00 – 0.05: Use in algorithm and retrieve soil moisture, do not flag.
        Water fraction is 0.05 – 0.50: Use in algorithm and retrieve soil moisture, flag.
                Water fraction is 0.5 – 1.00: Mask completely and flag, do not retrieve soil moisture

Masking and flagging of Sentinel-1A/1B grid cell:
        Water fraction is 0.00 – 0.05: Retrieve soil moisture, do not flag.
        Water fraction is 0.05 – 0.50: Flag and retrieve soil moisture.
        Water fraction is 0.50 – 1.00: Flag but do not retrieve soil moisture.

Surface flag in L2_SM_SP product is initiated when the water fraction is greater than 0.05.
Retrieval quality flag in L2_SM_SP product is initiated when the water fraction is greater than 0.05.

RFI: Presence of RFI in the SMAP T_B adversely affects the L2_SM_SP algorithm. Therefore, specific logics are built into the L2_SM_SP processor to initiate masking and flagging during soil moisture retrievals. The RFI flag is initiated as follows:

> No RFI detected in T_B: Retrieve soil moisture, do not flag.
> RFI detected in T_B and repaired: Flag and retrieve soil moisture.
> RFI detected in T_B and not repaired: Flag and do not retrieve soil moisture.

No RFI information is provided in the Sentinel-1A/1B σ_{vv} and σ_{vh} data.

Retrieval quality flag in L2_SM_SP product is initiated when RFI is detected in SMAP T_B.

Snow: The ancillary data that provide a binary indicator for presence of snow are used for flagging in the following way:

> Snow data indicate no snow: Use T_B to retrieve soil moisture, do not flag.
> Snow data indicate snow: Flag, do not retrieve soil moisture.

Surface flag in L2_SM_SP product is initiated when snow is present.
Retrieval quality flag in L2_SM_SP product is initiated when snow is present.

Frozen Ground: Besides snow, the frozen ground is another situation that impacts observations from the SMAP T_B and Sentinel-1A/1B σ_{vv} and σ_{vh}. The surface temperature data are used to detect frozen ground (temperatures <0º). SMAP T_B and Sentinel-1A/1B σ_{vv} and σ_{vh} are eliminated from the L2_SM_SP processing.

Surface flag in L2_SM_SP product is initiated when surface temperature is less than0º .
Retrieval quality flag in L2_SM_SP product is initiated when surface temperature is less than0º.

Urban Area: Presence of urban area adversely affects the L-band radiometric measurements. The presence of urban areas within the SMAP measurement and Sentinel-1A/1B data is likely to bias soil moisture retrievals. Currently the L2_SM_SP processor flags the regions having urban areas as follows:

> Urban fraction is 0.00 – 0.25: Retrieve soil moisture, do not flag.
> Urban fraction is 0.25 – 0.50: Flag T_B and σ_{vv} and σ_{vh}, and retrieve soil moisture.
> Urban fraction is 0.50 – 1.00: Flag T_B and σ_{vv} and σ_{vh}, do not retrieve soil moisture.

Surface flag in L2_SM_SP product is initiated when urban area is greater than 0.25.
Retrieval quality flag in L2_SM_SP product is initiated when urban area is greater than 0.25.

Precipitation: Presence of heavy rainfall during SMAP data acquisition may adversely affect the T_B and Sentinel-1A/1B backscatter observations. Precipitation forecasts from Goddard's GEOS5 model are used to flag the T_B and Sentinel-1A/1B backscatter observations. L2_SM_SP retrievals are performed irrespective of rainfall; however, the L2_SM_SP product grid cells are flagged in case of precipitation more than 5 mm.

Surface flag in L2_SM_SP product is initiated when precipitation is detected more than 5 mm.
Retrieval quality flag in L2_SM_SP product is initiated when precipitation is detected more than 5 mm.

VWC: L2_SM_SP retrievals are made for all the locations irrespective of VWC level. The grid cells are flagged for VWC greater than 3 kg/m2.

Surface flag in L2_SM_SP product is initiated when VWC is greater than 3 kg/m2.
Retrieval quality flag in L2_SM_SP product is initiated when VWC is greater than 3 kg/m2.

Mountainous Area: The statistical metric of mountainous regions that initiates flags and operational decisions during L2_SM_AP processing is the standard deviation of terrain slope. There are other options and threshold statistics (the range of elevation, the variance of elevation, and combination of variance of slope and elevation parameters) that were considered for the L2_SM_SP product. However, standard deviation of slope seems to be robust to detect mountainous and uneven terrain that may impact the quality of radiometric measurements. Currently the L2_SM_SP processor flags the region where standard deviation of slope is more than 3 degrees; however, the retrieval is performed for all locations.

Surface flag in L2_SM_SP product is initiated when the standard deviation of DEM slope is more than 3 degrees.
Retrieval quality flag in L2_SM_SP product is initiated when the standard deviation of DEM slope is more than 3 degrees.

Coastal Region: Large masses of water near the land such as coastal regions close to sea, ocean and large lakes influence the SMAP footprint. SMAP T_B data along the coast are substantially contaminated due to the presence of adjacent large waterbodies. Correction of the T_B data is possible and has been done in the L2_SM_P_E processing. However, the correction process for land T_B data is not perfect and the T_B data may still contain some error. Therefore, the surface flag in the L2_SM_SP product is initiated when the grid cell is within 60 km of the coast. The retrieval quality flag in the L2_SM_SP product is initiated when the grid cell is within 60 km of the coast.

6.5   Numerical Computation and Storage Consideration

The expected computational requirements of L2_SM_SP activities are moderate. With the present technology in computing and archiving of electronic files, it can easily accommodate the L2_SM_SP processing. The maximum computational demand (processor time and cache memory) is anticipated during ingestion and preliminary processing of Sentinel-1A/1B data. With the current infrastructure at the JPL Science Data System (SDS), no bottlenecks are expected while running the L2_SM_SP software PGE for the SMAP-Sentinel Active-Passive product. Data volume of one L2_SM_SP product granule is ~4 - 6 MB, and there are nearly ~200 such granules per day. This makes a yearly volume of ~365 GB for the L2_SM_SP product.

6.6   Programming Consideration

Processing and operational codes for the L2_SM_SP algorithm are written in Fortran to make it consistent with other algorithms. This facilitates inter-algorithm functioning of switches, data, and information transactions. Programming for the L2_SM_SP algorithm adheres to standard coding specifications to ensure consistent, maintainable, and readable code deliveries within the SMAP SDS. Fortran programming of the L2_SM_SP algorithm also meets the necessary requirements of language compliance, predictable execution, and code clarity. Adherence to these standards by Fortran routines allows efficient integration of SDS software components (Interfaces) and simplifies the Algorithm-to-PGE development process. The L2_SM_SP Fortran codes contain comments and version control information to track the changes and streamline the development of software. A software specification document was developed for documentation of all the source codes.

6.7   Exception Handling

To obtain the L2_SM_SP product from SMAP and Sentinel-1A/1B observations involves many aspects of product generation. This includes instrument performance, satellite data downlink, data preprocessing activities, quality of data (e.g., data drop-off), preceding algorithm performance, availability of ancillary data, and computation of related resources. Due to these activities, exceptions are expected while operating the L2_SM_SP algorithm on the SDS testbed. The development of L2_SM_SP software includes identification of expected exceptions. However, the formulation and computer coding of L2_SM_SP software is robust to withstand the expected exceptions and exit normally with messages in case of any exception.

6.8   Interface Assumptions

The L2_SM_SP algorithm generates data based on the input of L2_SM_P_E (water bodies corrected brightness temperature fields) and the Sentinel-1A/1B (radar backscatter cross-section) products. Masks and flags contained in the foundation products are also propagated in the L2_SM_SP. The masks include bad or missing data, inland water, coasts, vegetation and terrain flags, frozen ground flag, etc. In order to maintain consistency and streamline production, any additional masks that may be required for the L2_SM_SP product will be requested to be included in the parent SMAP data product.

6.9   Latency in L2_SM_SP Product

The L2_SM_SP product has no latency requirements. However, the SMAP project tries to make the L2_SM_SP product available to the NASA DAAC at NSIDC as early as possible. On average the data are available after ~3 days (~72 hours). The main driver of latency for the L2_SM_SP products is the availability of the Sentinel-1A/1B data and the requirement of 3 days to look for the closest match between the SMAP T_B L2_SM_P_E and Sentinel-1A/1B σ_{vv} and σ_{vh} for the Active-Passive algorithm processing. Overall ~90% of the Sentinel-1A/1B granules received by JPL SDS are processed for the L2_SM_SP product and delivered to NSIDC between ~72 – 96 hours.

7   PRODUCTS SPECIFICATIONS

The L2_SM_SP product contains relevant outputs from the Active-Passive algorithm (Section 3.2), soil moisture retrieval data, dynamic ancillary data used in the soil moisture retrieval, and metadata information. The L2_SM_SP product file comprises two distinct data groups, one for the 3 km EASE2 grid and a second one for the 1 km EASE2 grid. The L2_SM_SP product contains soil moisture retrievals for each data group. Table 9 provides the fields contained in the L2_SM_SP product file for both of the data groups. However, the primary data group of the L2_SM_SP product is at 3 km resolution (EASE2 grid) which has been validated, and the 1 km resolution (EASE2 grid) data group which is experimental. For completeness, disaggregated T_B from the current Active-Passive algorithm, averaged radar backscatter, dynamic ancillary inputs, and algorithm parameters are included. Geolocation information (latitude, longitude, and EASE2 grid-ID) and UTC time of the spacecraft overpass are also included for all the grid cells at 3 km and 1 km. Within each data group, the L2_SM_SP product file contains two parts: 1) the baseline is when only the SMAP descending pass is used in the input data stream to find an overlap with the Sentinel-1A/1B overpass, and 2) when the SMAP descending-ascending (apm) combination is used in the input data stream to find an overlap with the Sentinel-1A/1B overpass. The descending-ascending (apm) combination is the process of overlapping the SMAP T_B data from the descending pass or the ascending pass obtained from L2_SM_P_E data to find the minimum overpass time difference between the SMAP T_B data and the Sentinel-1A/1B σ data. The overlap (+/- 24 hours timeframe) is done using the 3 days of the SMAP T_B data centered around the Sentinel-1A/1B overpass time.

With the current design of the L2_SM_SP product file, the contents can be easily mapped on to the 3 km and 1 km fixed EASE2 grid cells using the geolocation information present in the product file. Complete details about the contents are provided in the L2_SM_SP product specification document. There are a total of 52 data elements at 3 km EASE2 grid resolution and 52 data elements at 1 km EASE2 grid resolution. Some of the data elements are experimental and may be removed in future releases. The metadata and the populated product fields are packaged in HDF5 file format and are made available to the public from the NASA DAAC at the National Snow and Ice Data Center (NSIDC) in Boulder, Colorado [https://nsidc.org/data/smap].

Table 9: *L2_SM_SP product file data fields (mostly in alphabetical order) at 3 km and 1 km EASE2 grid resolutions that contain geolocation information, disaggregated T_B, aggregated σ (sigma0), geophysical data fields, QC data, and dynamic ancillary data.*

L2_SM_SP 3 km Data Group			
Data Element Name	**Data type**	**Unit**	**Description**
'EASE2_column_index_3km'	integer	(-)	EASE2 grid column index at 3 km
'EASE2_column_index_apm_3km'	integer	(-)	EASE2 grid column index at 3 km for apm
'EASE2_row_index_3km'	integer	(-)	EASE2 grid row index at 3 km
'EASE2_row_index_apm_3km'	integer	(-)	EASE2 grid row index at 3 km for apm
'SMAP_Sentinel_overpass_timediff_hr_3km'	real	hours	Time difference of acquisition between SMAP and Sentinel-1A/1B observations
'SMAP_Sentinel_overpass_timediff_hr_apm_3km'	real	hours	Time difference of acquisition between SMAP and Sentinel-1A/1B observations for apm
'albedo_3km'	real	(-)	Vegetation albedo used in Tau-Omega model for soil moisture retrieval

'albedo_apm_3km'	real	(-)	Vegetation albedo used in Tau-Omega model for soil moisture retrieval for apm
'bare_soil_roughness_retrieved_3km'	real	(-)	Soil roughness provided by the lookup table Tau-Omega model for soil moisture retrieval
'bare_soil_roughness_retrieved_apm_3km'	real	(-)	Soil roughness provided by the lookup table Tau-Omega model for soil moisture retrieval for apm
'beta_tbv_vv_3km'	real	(-)	Beta parameter derived from T_{Bh} and σ_{hh} with snapshot approach used in the Active/Passive retrieval algorithm for the coarse resolution EASE2 grid cell (~33 km)
'beta_tbv_vv_apm_3km'	real	(-)	Beta parameter derived from T_{Bh} and σ_{hh} with snapshot approach used in the Active/Passive retrieval algorithm for the coarse resolution EASE2 grid cell (~33 km) for apm
'disagg_soil_moisture_3km'	real	cm3/cm3	It is soil moisture from the alternative algorithm that disaggregates the coarse resolution (~33 km) soil moisture from L2_SM_P_E
'disagg_soil_moisture_apm_3km'	real	cm3/cm3	It is soil moisture from the alternative algorithm that disaggregates the coarse resolution (~33 km) soil moisture from L2_SM_P_E for apm
'disaggregated_tb_v_qual_flag_3km'	bit flags	(-)	Bit flags that record the conditions and the quality of the disaggregated T_B at 3 km
'disaggregated_tb_v_qual_flag_apm_3km'	bit flags	(-)	Bit flags that record the conditions and the quality of the disaggregated T_B at 3 km for apm
'gamma_vv_xpol_3km'	real	(-)	Heterogeneity parameter of the Active-Passive algorithm used to obtain disaggregated T_B.
'gamma_vv_xpol_apm_3km'	real	(-)	Heterogeneity parameter of the Active-Passive algorithm used to obtain disaggregated T_B for apm
'landcover_class_3km'	integer	(-)	Enumerated data that classify MODIS derived dominant landcover at 3 km
'landcover_class_apm_3km'	integer	(-)	Enumerated data that classify MODIS derived dominant landcover at 3 km for apm
'latitude_3km'	real	degrees	Latitude of the center of the EASE2 based grid cell at 3 km
'latitude_apm_3km'	real	degrees	Latitude of the center of the EASE2 based grid cell at 3 km at apm
'longitude_3km'	real	degrees	Longitude of the center of the EASE2 based grid cell at 3 km
'longitude_apm_3km'	real	degrees	Longitude of the center of the EASE2 based grid cell at 3 km at apm
'retrieval_qual_flag_3km'	bit flags	(-)	Bit flags that record the conditions and the quality of the retrieval algorithms that generate soil moisture for the Active-Passive algorithm
'retrieval_qual_flag_apm_3km'	bit flags	(-)	Bit flags that record the conditions and the quality of the retrieval algorithms that generate soil moisture for the Active-Passive algorithm for apm
'sigma0_incidence_angle_3km'	real	degrees	Averaged Sentinel-1A/1B sigma0 incidence angle at 3 km
'sigma0_incidence_angle_apm_3km'	real	degrees	Averaged Sentinel-1A/1B sigma0 incidence angle at 3 km at apm
'sigma0_vh_aggregated_3km'	real	(-)	Average of fine resolution x-pol (vh) Sentinel-1A/1B sigma0 into at 3 km EASE2 grid resolution
'sigma0_vh_aggregated_apm_3km'	real	(-)	Average of fine resolution x-pol (vh) Sentinel-1A/1B sigma0 into at 3 km EASE2 grid resolution at apm
'sigma0_vv_aggregated_3km'	real	(-)	Average of fine resolution v-pol (vv) Sentinel-1A/1B sigma0 into at 3 km EASE2 grid resolution
'sigma0_vv_aggregated_apm_3km'	real	(-)	Average of fine resolution v-pol (vv) Sentinel-1A/1B sigma0 into at 3 km EASE2 grid resolution at apm

'soil_moisture_3km'	real	cm3/cm3	Soil moisture retrieved using disaggregated T_{Bh} for the 3 km EASE2 grid cell
'soil_moisture_apm_3km'	real	cm3/cm3	Soil moisture retrieved using disaggregated T_{Bh} for the 3 km EASE2 grid cell at apm
'soil_moisture_std_dev_3km'	real	cm3/cm3	Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 3 km EASE2 grid cell
'soil_moisture_std_dev_apm_3km'	real	cm3/cm3	Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 3 km EASE2 grid cell at apm
'spacecraft_overpass_time_seconds_3km'	real	seconds	Number of seconds since a specified epoch that represents the SMAP overpass relative to the ~33 km EASE2 grid cell that contains each 3 km EASE2 grid cell represented in this data product
'spacecraft_overpass_time_seconds_apm_3km'	real	seconds	Number of seconds since a specified epoch that represents the SMAP overpass relative to the ~33 km EASE2 grid cell that contains each 3 km EASE2 grid cell represented in this data product at apm
'surface_flag_3km'	bit flags	(-)	Bit flags that record ambient surface conditions for the 3 km EASE2 grid cell
'surface_flag_apm_3km'	bit flags	(-)	Bit flags that record ambient surface conditions for the 3 km EASE2 grid cell at apm
'surface_temperature_3km'	real	Kelvin	Temperature at land surface based on a geophysical model (GMAO) at 3 km, base resolution is 9 km
'surface_temperature_apm_3km'	real	Kelvin	Temperature at land surface based on a geophysical model (GMAO) at 3 km, base resolution is 9 km at apm
'tb_v_disaggregated_3km'	real	Kelvin	SMAP T_{Bv} disaggregated from the ~33 km grid cells into 3 km EASE2 grid cells
'tb_v_disaggregated_apm_3km'	real	Kelvin	SMAP T_{Bv} disaggregated from the ~33 km grid cells into 3 km EASE2 grid cells at apm
'tb_v_disaggregated_std_3km'	real	Kelvin	STD in SMAP T_{Bv} disaggregated from the ~33 km grid cells into 3 km EASE2 grid cells
'tb_v_disaggregated_std_apm_3km'	real	Kelvin	STD in SMAP T_{Bv} disaggregated from the ~33 km grid cells into 3 km EASE2 grid cells at apm
'vegetation_opacity_3km'	real	(-)	Vegetation opacity (tau) in the 3 km grid cell compute using the NDVI data
'vegetation_opacity_apm_3km'	real	(-)	Vegetation opacity (tau) in the 3 km grid cell compute using the NDVI data at apm
'vegetation_water_content_3km'	real	kg/m2	Vegetation-water-content (VWC) in the 3 km grid cell compute using the NDVI data
'vegetation_water_content_apm_3km'	real	kg/m2	Vegetation-water-content (VWC) in the 3 km grid cell compute using the NDVI data at apm
'water_body_fraction_3km'	real	(-)	Fraction of the area of 3 km grid cell that is a permanent or transient water body
'water_body_fraction_apm_3km'	real	(-)	Fraction of the area of 3 km grid cell that is a permanent or transient water body at apm

L2_SM_SP 1 km Data Group			
Data Element Name	**Data type**	**Unit**	**Description**
'EASE2_column_index_1km'	integer	(-)	EASE2 grid column index at 1 km
'EASE2_column_index_apm_1km'	integer	(-)	EASE2 grid column index at 1 km for apm

'EASE2_row_index_1km'	integer	(-)	EASE2 grid row index at 1 km
'EASE2_row_index_apm_1km'	integer	(-)	EASE2 grid row index at 1 km for apm
'SMAP_Sentinel_overpass_timediff_hr_1km'	real	hours	Time difference of acquisition between SMAP and Sentinel-1A/1B observations
'SMAP_Sentinel_overpass_timediff_hr_apm_1km'	real	hours	Time difference of acquisition between SMAP and Sentinel-1A/1B observations for apm
'albedo_1km'	real	(-)	Vegetation albedo used in Tau-Omega model for soil moisture retrieval
'albedo_apm_1km'	real	(-)	Vegetation albedo used in Tau-Omega model for soil moisture retrieval for apm
'bare_soil_roughness_retrieved_1km'	real	(-)	Soil roughness provided by the lookup table Tau-Omega model for soil moisture retrieval
'bare_soil_roughness_retrieved_apm_1km'	real	(-)	Soil roughness provided by the lookup table Tau-Omega model for soil moisture retrieval for apm
'beta_tbv_vv_1km'	real	(-)	Beta parameter derived from T_{Bh} and σ_{hh} with snapshot approach used in the Active/Passive retrieval algorithm for the coarse resolution EASE2 grid cell (~33 km)
'beta_tbv_vv_apm_1km'	real	(-)	Beta parameter derived from T_{Bh} and σ_{hh} with snapshot approach used in the Active/Passive retrieval algorithm for the coarse resolution EASE2 grid cell (~33 km) for apm
'disagg_soil_moisture_1km'	real	cm3/cm3	It is soil moisture from the alternative algorithm that disaggregates the coarse resolution (~33 km) soil moisture from L2_SM_P_E
'disagg_soil_moisture_apm_1km'	real	cm3/cm3	It is soil moisture from the alternative algorithm that disaggregates the coarse resolution (~33 km) soil moisture from L2_SM_P_E for apm
'disaggregated_tb_v_qual_flag_1km'	integer	(-)	Bit flags that record the conditions and the quality of the disaggregated T_B at 1 km
'disaggregated_tb_v_qual_flag_apm_1km'	integer	(-)	Bit flags that record the conditions and the quality of the disaggregated T_B at 1 km for apm
'gamma_vv_xpol_1km'	real	(-)	Heterogeneity parameter of the Active-Passive algorithm used to obtain disaggregated T_B.
'gamma_vv_xpol_apm_1km'	real	(-)	Heterogeneity parameter of the Active-Passive algorithm used to obtain disaggregated T_B for apm
'landcover_class_1km'	integer	(-)	Enumerated data that classify MODIS derived dominant landcover at 1 km
'landcover_class_apm_1km'	integer	(-)	Enumerated data that classify MODIS derived dominant landcover at 1 km for apm
'latitude_1km'	real	degrees	Latitude of the center of the EASE2 based grid cell at 1 km
'latitude_apm_1km'	real	degrees	Latitude of the center of the EASE2 based grid cell at 1 km at apm
'longitude_1km'	real	degrees	Longitude of the center of the EASE2 based grid cell at 1 km
'longitude_apm_1km'	real	degrees	Longitude of the center of the EASE2 based grid cell at 1 km at apm
'retrieval_qual_flag_1km'	integer	(-)	Bit flags that record the conditions and the quality of the retrieval algorithms that

			generate soil moisture for the Active-Passive algorithm
'retrieval_qual_flag_apm_1km'	integer	(-)	Bit flags that record the conditions and the quality of the retrieval algorithms that generate soil moisture for the Active-Passive algorithm for apm
'sigma0_incidence_angle_1km'	real	degrees	Averaged Sentinel-1A/1B sigma0 incidence angle at 1 km
'sigma0_incidence_angle_apm_1km'	real	degrees	Averaged Sentinel-1A/1B sigma0 incidence angle at 1 km at apm
'sigma0_vh_aggregated_1km'	real	(-)	Average of fine resolution x-pol (vh) Sentinel-1A/1B sigma0 into at 1 km EASE2 grid resolution
'sigma0_vh_aggregated_apm_1km'	real	(-)	Average of fine resolution x-pol (vh) Sentinel-1A/1B sigma0 into at 1 km EASE2 grid resolution at apm
'sigma0_vv_aggregated_1km'	real	(-)	Average of fine resolution v-pol (vv) Sentinel-1A/1B sigma0 into at 1 km EASE2 grid resolution
'sigma0_vv_aggregated_apm_1km'	real	(-)	Average of fine resolution v-pol (vv) Sentinel-1A/1B sigma0 into at 1 km EASE2 grid resolution at apm
'soil_moisture_1km'	real	cm3/cm3	Soil moisture retrieved using disaggregated T_{Bh} for the 1 km EASE2 grid cell
'soil_moisture_apm_1km'	real	cm3/cm3	Soil moisture retrieved using disaggregated T_{Bh} for the 1 km EASE2 grid cell at apm
'soil_moisture_std_dev_1km'	real	cm3/cm3	Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 1 km EASE2 grid cell
'soil_moisture_std_dev_apm_1km'	real	cm3/cm3	Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 1 km EASE2 grid cell at apm
'spacecraft_overpass_time_seconds_1km'	real	seconds	Number of seconds since a specified epoch that represents the SMAP overpass relative to the ~33 km EASE2 grid cell that contains each 1 km EASE2 grid cell represented in this data product
'spacecraft_overpass_time_seconds_apm_1km'	real	seconds	Number of seconds since a specified epoch that represents the SMAP overpass relative to the ~33 km EASE2 grid cell that contains each 1 km EASE2 grid cell represented in this data product at apm
'surface_flag_1km'	integer	(-)	Bit flags that record ambient surface conditions for the 1 km EASE2 grid cell
'surface_flag_apm_1km'	integer	(-)	Bit flags that record ambient surface conditions for the 1 km EASE2 grid cell at apm
'surface_temperature_1km'	real	Kelvin	Temperature at land surface based on a geophysical model (GMAO) at 1 km, base resolution is 9 km
'surface_temperature_apm_1km'	real	Kelvin	Temperature at land surface based on a geophysical model (GMAO) at 1 km, base resolution is 9 km at apm
'tb_v_disaggregated_1km'	real	Kelvin	SMAP T_{Bv} disaggregated from the ~33 km grid cells into 1 km EASE2 grid cells
'tb_v_disaggregated_apm_1km'	real	Kelvin	SMAP T_{Bv} disaggregated from the ~33 km grid cells into 1 km EASE2 grid cells at apm
'tb_v_disaggregated_std_1km'	real	Kelvin	STD in SMAP T_{Bv} disaggregated from the ~33 km grid cells into 1 km EASE2 grid cells

'tb_v_disaggregated_std_apm_1km'	real	Kelvin	STD in SMAP T_{Bv} disaggregated from the ~33 km grid cells into 1 km EASE2 grid cells at apm
'vegetation_opacity_1km'	real	(-)	Vegetation opacity (tau) in the 1 km grid cell compute using the NDVI data
'vegetation_opacity_apm_1km'	real	(-)	Vegetation opacity (tau) in the 1 km grid cell compute using the NDVI data at apm
'vegetation_water_content_1km'	real	kg/m2	Vegetation-water-content (VWC) in the 1 km grid cell compute using the NDVI data
'vegetation_water_content_apm_1km'	real	kg/m2	Vegetation-water-content (VWC) in the 1 km grid cell compute using the NDVI data at apm
'water_body_fraction_1km'	real	(-)	Fraction of the area of 1 km grid cell that is a permanent or transient water body
'water_body_fraction_apm_1km'	real	(-)	Fraction of the area of 1 km grid cell that is a permanent or transient water body at apm

8   FUTURE CONSIDERATIONS for the L2_SM_SP PRODUCT

There is further potential for improvement in the L2_SM_SP data quality by reducing the errors in soil moisture retrievals. These improvements include use of better ancillary data (e.g. optimized VWC and better soil texture data) and optimization of the tau-omega model parameters for various land covers at resolutions of 3 km and 1 km. The current operational processing of the L2_SM_SP product can be improved to attain better soil moisture accuracy over most of the land covers through the following steps:

a) Replacing the current soil texture database with the latest available high-resolution soil texture data.
b) Improving the dielectric model for higher latitudes where organic content in soil is high.
c) Improving the method/model that is used to derive VWC to obtain the vegetation opacity used in the tau-omega model to retrieve soil moisture.

8.1   Replacing Soil Texture Database

The SMAP project uses a blend of the Harmonized World Soil Database (HWSD) at ~10–25 km resolution, the State Soil Geographic (STATSGO) database for the continental United States (CONUS) at 1 km resolution, and the Australian Soil Resource Information System (ASRIS) at 1 km resolution. Apart from CONUS and Australia, the rest of the world has soil data that are very coarse (~10 to 25 km) and outdated. Recent advances in the global soil database such as available at https://openlandmap.org provide very high resolution (~250 m) with better accuracy soil texture information [31]. Including high-resolution and recent soil texture data in the L2_SM_SP retrieval process will likely improve its global performance. The impact of the coarse resolution soil texture data currently used in the SMAP operational soil moisture retrieval process is not visible in the L2_SM_SP CVS because most of CVS are confined to CONUS and Australia where high resolution soil texture data (STATSGO and ASRIS) are available. Figure 27 illustrates a comparison of texture data between the current composite soil database used by the SMAP project and the soil database from https://openlandmap.org [31]. The quality of the new database appears to be more reliable, likely due to the latest high-resolution remote sensing information, an enhancement in the number of *in situ* soil physical attribute observations, and other high-resolution pedotransfer functions used in its creation.
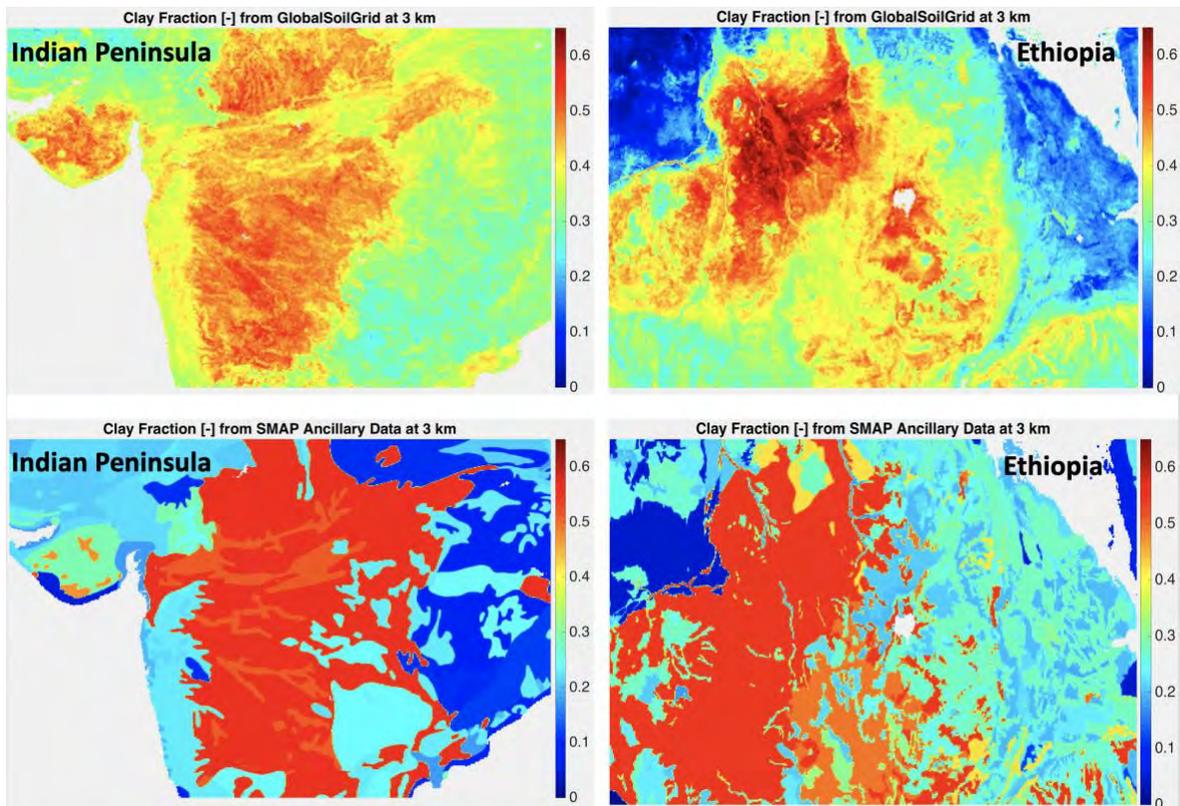
Figure 27: Comparison between the latest GlobalSoilGrid (~250 m) from https://openlandmap.org and the composite soil database used by the SMAP project. More details and natural spatial distribution of clay in the top soil is observed in the GlobalSoilGrid as compared to the artificial and unnatural contours and distinct boundaries in the current SMAP soil texture baseline.

In looking at different regions of the world, similar observations related to the soil texture data were evident in the latest GlobalSoilGrid (~250 m). Therefore, it is expected that the overall quality of the L2_SM_SP soil moisture data will improve with improved global soil texture data.

8.2   Improving the Dielectric Model for Higher Latitude

The current L2_SM_SP processing at JPL SDS uses the Mironov model [32] for estimating the real and the imaginary part of the soil effective dielectric constant. The clay fraction is the only soil physical attribute that goes into the Mironov model to compute the dielectric constant. With the clay fraction input to the Mironov model, it is optimal for most of the hydroclimatic domains and biomes, except for higher latitudes such as tundra regions where the organic component in the soil is present in high amounts. The Mironov model is not optimized for estimating the dielectric constant for top soils having a large fraction of organic matter. Other dielectric models such as [33] and [34] that consider organic content of the top soil as one of the inputs will be evaluated to determine their impact on soil moisture retrieval accuracy. If warranted, the inclusion of such a dielectric model in the JPL SDS processing should improve the L2_SM_SP soil moisture product for higher latitudes.

8.3   Improving VWC estimates

Currently, the SMAP L2_SM_SP retrievals use the same tau-omega parameters as the L2_SM_P_E retrievals [ref?]. Another important step to improving the L2_SM_SP data quality is the inclusion of retrieved vegetation optical depth (VOD) or vegetation opacity tau (τ), from dual-channel algorithms such as [35. However, such retrieved τ are at the coarse resolution of the radiometer (at C scale), and therefore, not very useful for high-resolution soil moisture retrievals. The τ values used for L2_SM_SP soil moisture retrievals are derived from a 10-year (2002−2012) climatology of NDVI-based VWC (τ =b*VWC, where b is a parameter based on land cover, typically close to 0.1 at L-band frequencies). The drawback of using VWC climatology for τ is prominently visible over CVS with cropland land cover because it may be out of sync. The accuracy of τ is critical because τ has a first order impact on the quality of the L2_SM_SP soil moisture product. Fig. 28 illustrates one such scenario where the NDVI climatology taken from Day-of-Year (DOY) 185 is compared against the actual NDVI for DOY 185, 2017 for California. Two time series plots, one from a natural landscape with shrubland cover and another with cropland, are shown in Fig. 28. The actual NDVI time series over cropland (agricultural region) does not match with the climatology (2002–2012) mostly due to crop rotation and differences in planting date for that particular year. However, the climatology and the actual time series over the shrubland are similar. Such mismatches are very possible over many of the CVS with crop land cover, hence leading to inferior performance of the L2_SM_SP product, as visible in Tables 5-6 for CVS Yanco and Kenaston. Inclusion of actual NDVI in the operational L2_SM_SP processing has potential to improve the overall accuracy of the L2SMSP soil moisture retrievals for those cases where poor accuracy is due to mismatch errors between actual and climatological NDVI . However, the 8–16 day latency of actual MODIS NDVI data is a constraint (cloudy conditions would also contribute to larger latencies). As an alternative to NDVI, the cross-polarized Sentinel-1A/B measurements could be used as a variable that is proportional to τ [36]. Future research will examine the use of Sentinel-1A/1B cross-polarized data to estimate τ and its potential to improve the quality of the L2_SM_SP product.
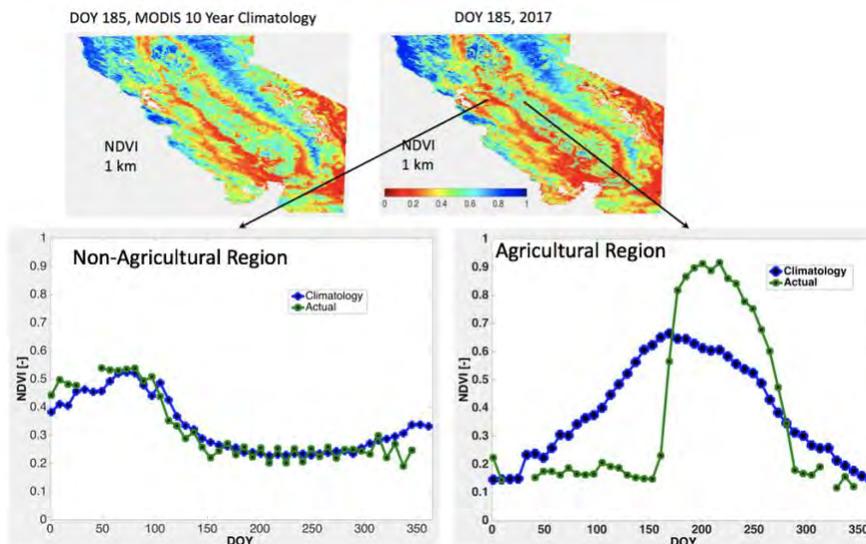

Figure 28: Comparison of actual NDVI (green curve) and climatology of NDVI (2002-2012) (blue curve) for an agricultural region (cropland) and a non-agricultural region (shrubland) in Central Valley, California.

9   REFERENCES

[1]   E. G. Njoku, T. Jackson, V. Lakshmi, T. K. Chan, and S. Nghiem, "Soil moisture retrieval from AMSR-E," IEEE Trans. Geosci. Remote Sens., vol. 41, no. 2, pp. 215–229, Feb. 2003

[2]   Y. H. Kerr et al., "Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission," IEEE Trans. Geosci. Remote Sens., vol. 39, no. 8, pp. 1729–1735, Aug. 2001.

[3]   National Research Council, "Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond," pp. 400, 2007.

[4]   D. Entekhabi, E. Njoku, P. Houser, M. Spencer, T. Doiron, J. Smith, R. Girard, S. Belair, W. Crow, T. Jackson, Y. Kerr, J. Kimball, R. Koster, K. McDonald, P. O'Neill, T. Pultz, S. Running, J. C. Shi, E. Wood, and J. van Zyl, "The Hydrosphere State (HYDROS) mission concept: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw," IEEE Trans. on Geosci. and Remote Sensing, vol. 42(10), pp. 2184-2195, 2004.

[5]   D. Entekhabi, E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, N. Martin, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, and J. V. Zyl, "The Soil Moisture Active Passive (SMAP) Mission," Proceedings of the *IEEE*, vol. 98, 704-716, 2010.

[6]   N. N. Das, D. Entekhabi, R. S. Dunbar, A. Colliander, F. Chenc, W. Crow, T. J. Jackson, A. Berg, D. D. Bosch, T. Caldwell, M. H. Cosh, C. H. Collins, E. Lopez-Baeza, M. Moghaddam, T. Rowlandson, P. J. Starks, M. Thibeault, J. P. Walker, X. Wu, P. E. O'Neill, S. Yueh, and E. G. Njoku, "The SMAP mission combined Active-Passive soil moisture product at 9 km and 3 km spatial resolutions," *Remote Sens. Environ*., vol. 211, pp. 204-217, 2018.

[7]   E. G. Njoku, and D. Entekhabi, "Passive microwave remote sensing of soil moisture," *J. Hydrol.*, vol. 184, 101-129, 1996.

[8]   Chan, S., Bindlish, R., O'Neill, P., Njoku, E.G., Jackson, T.J., Colliander, A., S. Dunbar, J. Chaubell, Chen, J. Piepmeier, Yueh S., Entekhabi D., Cosh M., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., McNairn, H., Thibeault, M., Martínez- Fernández, J., González-Zamora, Á., Seyfried, M., Bosch, D., Starks, P., Goodrich, D., Prueger, J., Palecki, M., Small, E., Calvet, J.-C., Crow, W., Kerr, Y. "Assessment of the SMAP Passive Soil Moisture Product." *IEEE Transactions on Geoscience and Remote Sensing*, 54, 4994-5007, 2016.

[9]   Chan, S., Bindlish, R., O'Neill, P., Njoku, E.G., Jackson, T.J., Colliander, A., S. Dunbar, J. Chaubell, Chen, J. Piepmeier, Yueh S., Entekhabi D., Cosh M., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., McNairn, H., Thibeault, M., Martínez- Fernández, J., González-Zamora, Á., Seyfried, M., Bosch, D., Starks, P., Goodrich, D., Prueger, J., Palecki, M., Small, E., Calvet, J.-C., Crow, W., Kerr, Y. Development and assessment of the enhanced passive soil moisture product. *Remote Sensing of Environment*, 204, 931-941, 2017.

[10]   F. T. Ulaby, P. Dubois, and J. V. Zyl, "Radar mapping of surface soil moisture," *J. Hydrol.*, vol. 184, 57-84, 1996.

[11]   G. Kim, and A. P. Barros, "Space-time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data," *Remote Sens. Environ.*, vol. 81, pp. 393–403, 2002.

[12]   G. Kim, and A. P. Barros, "Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data," *Remote Sens. Environ.*, vol. 83, pp. 400–413, 2002.

[13]   N. S. Chauhan, S. Miller, and P Ardanuy, "Spaceborne soil moisture estimation at high-resolution: A microwave-optical/IR synergistic approach," *Int. J. Remote Sens.*, vol. 24, no. 22, pp. 4599–4622, 2003.

[14]   R. H. Reichle, D. Entekhabi, and D. B. McLaughlin, "Downscaling of radio brightness measurements for soil moisture estimation: A four dimensional variational data assimilation approach," Water Resources Res., vol. 37, no. 9, pp. 2353–2364, 2001.

[15]   O. Merlin, J. P. Walker, A. Chehbouni, and Yann Kerr, "Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency," *Remote Sens. Environ.*, vol. 112, pp. 3935–3946, 2008.

[16]   U. Narayan, V. Lakshmi, and T. J. Jackson, "High resolution estimation of soil moisture using L-band radiometer and radar observations made during the SMEX02 experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, 1545-1554, 2006.

[17]   X. Zhan, P. R. Houser, J. P. Walker, and W. T. Crow, "A Method or Retrieving High-Resolution Surface Soil Moisture from Hydros L-Band Radiometer and Radar Observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, 1534-1544, 2006.

[18]   Y. Kim, and J. van Zyl, "A time series approach to estimate soil moisture using polarimetric radar data," *IEEE Trans. Geosci. Remote Sens.* vol. 47, 2519-2527, 2009.

[19]   M. Piles, D. Entekhabi, and A. Camps, "A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, 4125-4131, 2009.

[20]   N. N. Das, D. Entekhabi, and E. G. Njoku, "An Algorithm for merging SMAP radiometer and radar data for high-resolution soil moisture retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 9, 1504-1512, 2011.

[21]   N. N. Das, D. Entekhabi, E. G. Njoku, J. Johnston, J. C. Shi, and A. Colliander, "Tests of the SMAP Combined Radar and Radiometer Brightness Temperature Disaggregation Algorithm Using Airborne Field Campaign Observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, pp. 2018–2028, 2014.

[22]   N. N. Das, D. Entekhabi, R. S. Dunbar, A. Colliander, F. Chenc, W. Crow, T. J. Jackson, A. Berg, D. D. Bosch, T. Caldwell, M. H. Cosh, C. H. Collins, E. Lopez-Baeza, M. Moghaddam, T. Rowlandson, P. J. Starks, M. Thibeault, J. P. Walker, X. Wu, P. E. O'Neill, S. Yueh, and E. G. Njoku, "The SMAP mission combined Active-Passive soil moisture product at 9 km

and 3 km spatial resolutions," *Remote Sens. Environ.*, vol. 211, pp. 204-217, 2018.

[23] N. N. Das, D. Entekhabi, R. S. Dunbara, M. J. Chaubell, A. Colliander, S. Yueh, T. Jagdhuber, F. Chen, W. Crow, P. E. O'Neill, J. P. Walker, Aaron Berg, D. D. Bosch, T. Caldwell, M. H. Cosh, C. H. Collins, E. Lopez-Baeza, and M. Thibeault., "The SMAP and Copernicus Sentinel-1A/B microwave Active-Passive high-resolution surface soil moisture product," *Remote Sens. Environ.*, vol. 233, 2019.

[24] D. Entekhabi, N. N. Das, J. Johnson, and J. Shi, "Algorithm Theoretical Basis Document L2 & L3 Radar/Radiometer Soil Moisture (Active/Passive) Data Products," JPL D-66481, (https://smap.jpl.nasa.gov/system/internal_resources/details/original/277_L2_3_SM_AP_R evA_web.pdf).

[25] Jagdhuber, T., D. Entekhabi, N. N. Das, M. Link, C. Montzka, M. Baur, R. Akbar, S. Kim, S.Yueh, I. Baris, (2018). "Physics-Based Modeling of Active-Passive Microwave Covariations for Geophysical Retrievals," IGARSS 2018, pp. 250-253. doi: 10.1109/IGARSS.2018.8518975.

[26] Jagdhuber, T., M. Baur, R. Akbar, N. N. Das, M. Link, L. He and D. Entekhabi (2019) "Estimation of Active-Passive Microwave Covariation Using SMAP and Sentinel-1 Data," Remote Sensing of Environment, vol. 225, 458-468.

[27] Ye, N., Walker, J.P., Rudiger, C., 2015. "A cumulative distribution function method for normalizing variable-angle microwave observations," IEEE Trans. Geosci. Remote Sens. 55 (7), 3906–3916.

[28] Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N. N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., Lopez-Baeza, E., Martinez-Fernandez, J., Gonzalez-Zamora, A., Livingston, S., McNairn, H., Pacheco, A., Moghaddam, M., Montzka, C., Notarnicola, C., Niedrist, G., Pellarin, T., Prueger, J., Pulliainen, J., Rautiainen, K., Ramo, J., Seyfried, M., Starks, P., Su, Z., Zeng, Y., van der Velde, R., Thibeault, M., Dorigo, W., Vreugdenhil, M., Walker, J. P., Wu, X., Monerris, A., O'Neill, P. E., Entekhabi, D., Njoku, E. G., and Yueh, S. (2017). "Validation of SMAP Surface Soil Moisture Products with Core Validation Sites," Remote Sensing of Environment, 191, 215-231.

[29] Chen, F., Crow, W. T., Colliander, A., Cosh, M., Jackson, T.J., Bindlish, R., Reichle, R., Chan, S.K., Bosch, D.D., Starks, P.S., Goodrich, D.C., and Seyfried, M.S., 2017. "Application of triple collocation in ground-based validation of soil moisture active/passive (SMAP) level 2 data products," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 489-502.

[30] Dorigo, W., Xaver A., Vreudgenhil M., Gruber A., Dostálová A., Sanchis-Dufau A. D., Zamojski D., Corde C., Wagner W., Matthias D.. (2013). "Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network," Vadose Zone Journal. 12. 10.2136/vzj2012.0097.

[31]   Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., et al., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12 (2), e0169748. https://doi.org/10.1371/journal.pone.

[32]   Mironov, V. L., L. G. Kosolapova, and S. V. Fomin, "Physically and mineralogically basedspectroscopic dielectric model for moist soils," IEEE Trans. Geosci. Remote Sens., 47(7), pp. 2059–2070, 2009.

[33]   Park, C. H., Behrendt A., LeDrew E., and Wulfmeyer V., 2017. "New Approach for calculating the effective dielectric constant of the moist soil for microwaves," Remote Sens., 9, 732; doi:10.3390/rs9070732.

[34]   Mironov, V. L., De Roo R., and Savin I. V., 2010. "Temperature-dependable microwave dielectric model for an artic soil," IEEE Trans. Geosci. Remote Sens., 48(6), pp. 2544–2556.

[35]   Konings, A.G., Piles, M., Das, N.N., Entekhabi, D., 2017. "L-band vegetation optical depth and effective scattering albedo estimation from SMAP," Remote Sens. Environ. 198, 460–470.

[36]   Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rudiger, C., Strauss, P., 2018. "Sensitivity of Sentinel-1 backscatter to vegetation dynamics: an Austrian case study," Remote Sens. 10, 1–19.