Soil Moisture Active Passive (SMAP)

Algorithm Theoretical Basis Document
L2 & L3 Radar/Radiometer Soil Moisture (Active/Passive) Data Products

Revision A
December 9, 2014

Dara Entekhabi
Massachusetts Institute of Technology
Cambridge, MA

Narendra Das, Eni Njoku, Simon Yueh
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA

Joel Johnson
Ohio State University
Columbus, OH

Jiancheng Shi
University of California
Santa Barbara, CA

Jet Propulsion Laboratory
California Institute of Technology

© 2014. All rights reserved.

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

The SMAP ATBDs were reviewed by a NASA Headquarters review panel in January 2012 with initial public release later in 2012. The current version is Revision A. The ATBDs may undergo additional version updates after SMAP launch.

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
EASE	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
OSSE	Observing System Simulation Experiment
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
TB	Brightness Temperature
TBC	To Be Confirmed
TBD	To Be Determined
USDA	United States Department of Agriculture
UTC	Coordinated Universal Time
VWC	Vegetation Water Content

1 INTRODUCTION

This document is the Algorithm Theoretical Basis Document (ATBD) for the surface soil moisture data product for hydrometeorology applications using combined SMAP radar and radiometer measurements. The SMAP Level 1 Requirements and Mission Success Criteria document specifies the SMAP baseline requirement for soil moisture as:

> ***4.1.1 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);*

In order to meet this requirement, the SMAP radar and radiometer measurements need to be combined. This document provides the theoretical basis and error analysis on the following data products:

1. Level 2 Soil Moisture (L2_SM_AP) in half orbit format.
2. Level 3 Soil Moisture (L3_SM_AP) in the form of global daily composites.

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 Soil Moisture and Ocean Salinity (SMOS) mission launched in December, 2009 is the first wide-swath L-band soil moisture mission, and has the potential for retrieving 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], consequently 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. Therefore surface soil moisture retrieved using these approaches are not matched suitably 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 National Research Council's Committee on Earth Science and Applications from Space recommended the implementation of the Soil Moisture Active Passive (SMAP) mission concept based on its impact on overall societal benefits and potential scientific advances in the fields of hydrology, meteorology and ecology [3].

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 prioritization 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 is late 2014 [5].

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.

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.

The SMAP instrument incorporates an L-band radar and an L-band radiometer that share 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 radiometer 3 dB antenna footprint of 40 km (root-ellipsoidal-area). The real-aperture radar footprint is 30 km, defined by the two-way antenna beamwidth. The real-aperture radar and radiometer data will 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 planned 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.

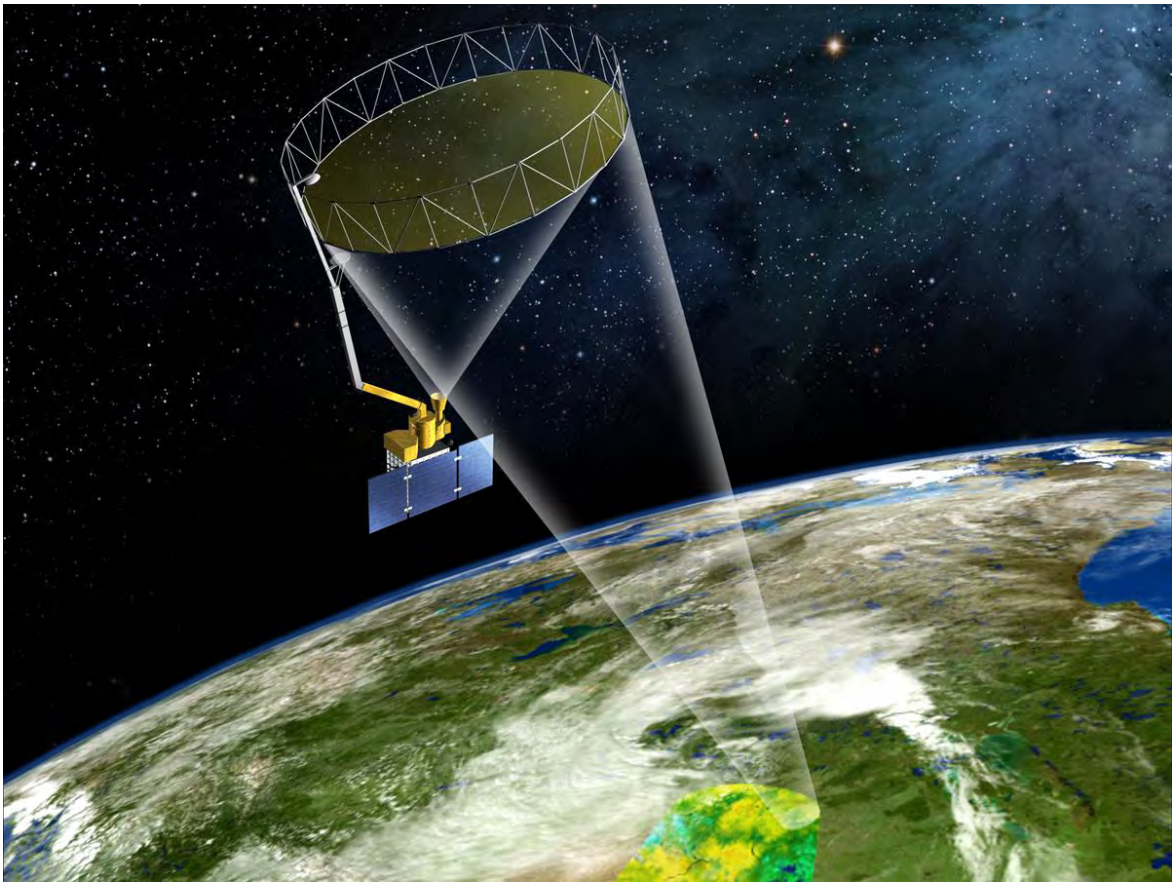


Figure 1*: The SMAP observatory is a dedicated spacecraft with a rotating 6-m light-weight deployable mesh reflector. The radar and radiometer share a common feed.*

Table 2. SMAP Data Products Table.

Product	Description	Gridding (Resolution)	Latency**	
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	
L1C_S0_HiRes	High Resolution Radar σ_o in Half-Orbits	1 km (1-3 km)	12 hrs	
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	
L3_FT_A	Freeze/Thaw State (Radar)	3 km	50 hrs	Science Data (Daily Composite)
L3_SM_A	Soil Moisture (Radar)	3 km	50 hrs	
L3_SM_P	Soil Moisture (Radiometer)	36 km	50 hrs	
L3_SM_AP	Soil Moisture (Radar + Radiometer)	9 km	50 hrs	
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	

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. Using the same antenna system but with synthetic aperture radar (SAR) processing, the SMAP L-band radar provides higher resolution (~3 km) 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 [6]. 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 SMAP radar backscatter cross-section [7]. With a typical instrument sensitivity of about 1 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 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 baseline and option algorithms that overcome these limitations by merging the active (radar) and passive (radiometer) measurements to derive a 9 km soil moisture product L2_SM_AP that meets the SMAP Level 1 requirements.

Relative to one another, the SMAP radiometer brightness temperature measurements are coarser resolution but with higher sensitivity to soil moisture and the SMAP radar backscatter cross-section measurements are higher resolution but with somewhat reduced sensitivity to soil moisture especially over densely vegetated land surfaces. The L2_SM_AP soil moisture product merges the two measurements to produce soil moisture retrieval with intermediate resolution that meets the Level 1 requirements.

The baseline suite of products from the SMAP mission is shown in Table 2. The Level 2 radiometer-only soil moisture product (L2_SM_P) is derived principally from the brightness temperature product (L1B_TB). This data product L2_SM_P is posted on a 36 km Earth fixed grid. The L2_SM_P product also includes a land brightness temperature estimate that is corrected for water bodies within the 36 km. This brightness temperature is available through a 'pipe-bend' during the production of the L2_SM_P processing. This water-body corrected brightness temperature is an input for the L2_SM_AP data processing.

The radar-only soil moisture (L2_SM_A) is a fine-resolution (3 km) soil moisture estimate derived from the Hi-Res radar backscatter data (L1C_S0_HiRes). The L2_SM_A data product is unlikely to meet the L1 accuracy requirements, although soil moisture information is expected to be achieved at a reduced accuracy, but at higher resolution. The L2_SM_A produces radar backscatter cross-section values aggregated to 3 km during the early stages of its processing. Through another 'pipe-bend' this data set along with water body and freeze/thaw flags are made available to the L2_SM_AP data processing as input.

The data product L2_SM_AP is posted on a 9 km EASE grid that is nested consistently with the 36 km brightness temperature and 3 km radar backscatter cross-section data sets (see Figure 2 and its discussion in Section 3.1).

1.4 Historical Perspective

In the past, numerous studies [8, 9, 10, 11] 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 infra red 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 [12]. From all these approaches, one common aspect is the use of static and dynamic ancillary geophysical data in the downscaling/disaggregation approaches. The ancillary geophysical data come from different sources with inherent systematic and unsystematic 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 [13]; 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 [14], where a Bayesian method is used to downscale radiometer observations using radar measurements. Kim et al [15] 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. [16] presented a change detection scheme within an 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 [13] this approach also suffers from accumulation of errors over time. A spatial variability technique developed by [17] to blend SMAP radar measurement and radiometer-based soil moisture data also takes advantage of 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 [13] and [15], the spatial variability technique used in [17] does not require the previous satellite overpass observations to estimate the current soil moisture value. A new active-passive algorithm is developed by [18] that draws from all the above algorithms and techniques ([13], [15], [16], and [17]). The new active-passive algorithm [18] 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 [18] is the current baseline L2_SM_AP algorithm, and is discussed in Section 3 in detail.

1.5 Product Characteristics

The L2_SM_AP product is based on the merger of the SMAP radiometer and radar instrument products at two discrete grid resolutions i.e., 36 km and 3 km, respectively. The Equal-Area-Scalable-Earth (EASE) grid cells of the radiometer and radar products nest perfectly (refer L2_SM_P ATBD), and therefore L2_SM_AP 9 km soil moisture product have 16:1 and 1:9 correspondence with the radiometer and radar products, respectively. The grid definition used in the algorithms is illustrated in Fig. 2. The baseline and optional algorithms disaggregate the coarse resolution radiometer brightness temperature product based on the spatial variation in high resolution radar backscatter. In addition, the algorithms require static and dynamic ancillary data. These ancillary data are resampled to the same EASE grid prior to ingestion in the L2_SM_AP processing. The dynamic ancillary data used to retrieve soil moisture for a particular 9 km grid cell at a specific point in time will be listed in the SMAP L2_SM_AP and L3_SM_AP 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 at L-band; Section 3 provides description and formulation of the L2_SM_AP baseline and option retrieval algorithms; Section 4 presents the results from the tests conducted and the error budget for the L2_SM_AP baseline algorithm; Section 5 discusses the practical consideration for implementing the baseline algorithm and generating the L2_SM_AP product; Section 6 provides the product specifications table; Section

7 provides a list of references; Section 8 is the glossary (under development). This ATBD will be updated as additional work is completed during the pre- and post-launch periods.

2 PHYSICS OF THE PROBLEM

The L2_SM_AP baseline 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. 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 SMAP makes coincident and constant look-angle radar and radiometer measurements, the co-variations over specified (short relative to plant phenology) periods of time 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.

Once the disaggregated brightness temperature at 9 km is produced, the brightness temperature inverse algorithms developed for the L2_SM_P product are then applied with ancillary information at 9 km to produce the L2_SM_AP product.

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 outlined in this section. This analysis is meant to simply demonstrate the dependencies and it is not directly (i.e., algebraically) part of the baseline L2_SM_AP algorithm formulation.

The brightness temperature at polarization p 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):

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

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 [6].

At the morning nodal crossing overpass isothermal near subsurface-to-surface-canopy thermal conditions are expected so that $T_s \approx T_c \equiv T$. Under low vegetation cover conditions 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:

$$T_{B_p} = T\left(1 - r_p \cdot e^{-2\tau_p/cos\theta}\right) \quad (2)$$

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 roughness parameter related to the root-mean-square (RMS) of surface roughness. Now (2) becomes:

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

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

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

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 [19] and [20], 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:

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

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:

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

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:

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

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:

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

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}:

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

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:

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

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 EASE grid radiometer brightness temperature (36 km), EASE grid radar backscatter cross-section (3 km), and desired merged active-passive L2_SM_AP (9 km) products. For convenience in mathematical formulation, the naming convention of '*C*' (coarse), '*F*' (fine), and '*M*' (medium) for the L1B_TB/L2_SM_P, L1C_S0_HiRes, and L2_SM_AP grid scales, respectively, is used throughout the following sections. It is evident from the grid topology (Fig. 2) that within a single (*nc* = 1) 36 km x 36 km grid cell of *C* there are *nm* = 16 grid cells of *M* and *nf* = 144 grid cells of *F*. Radar backscatter cross-section at coarse resolutions (*M* and *C*) are obtained by aggregating fine resolution radar backscatter cross-section in power.

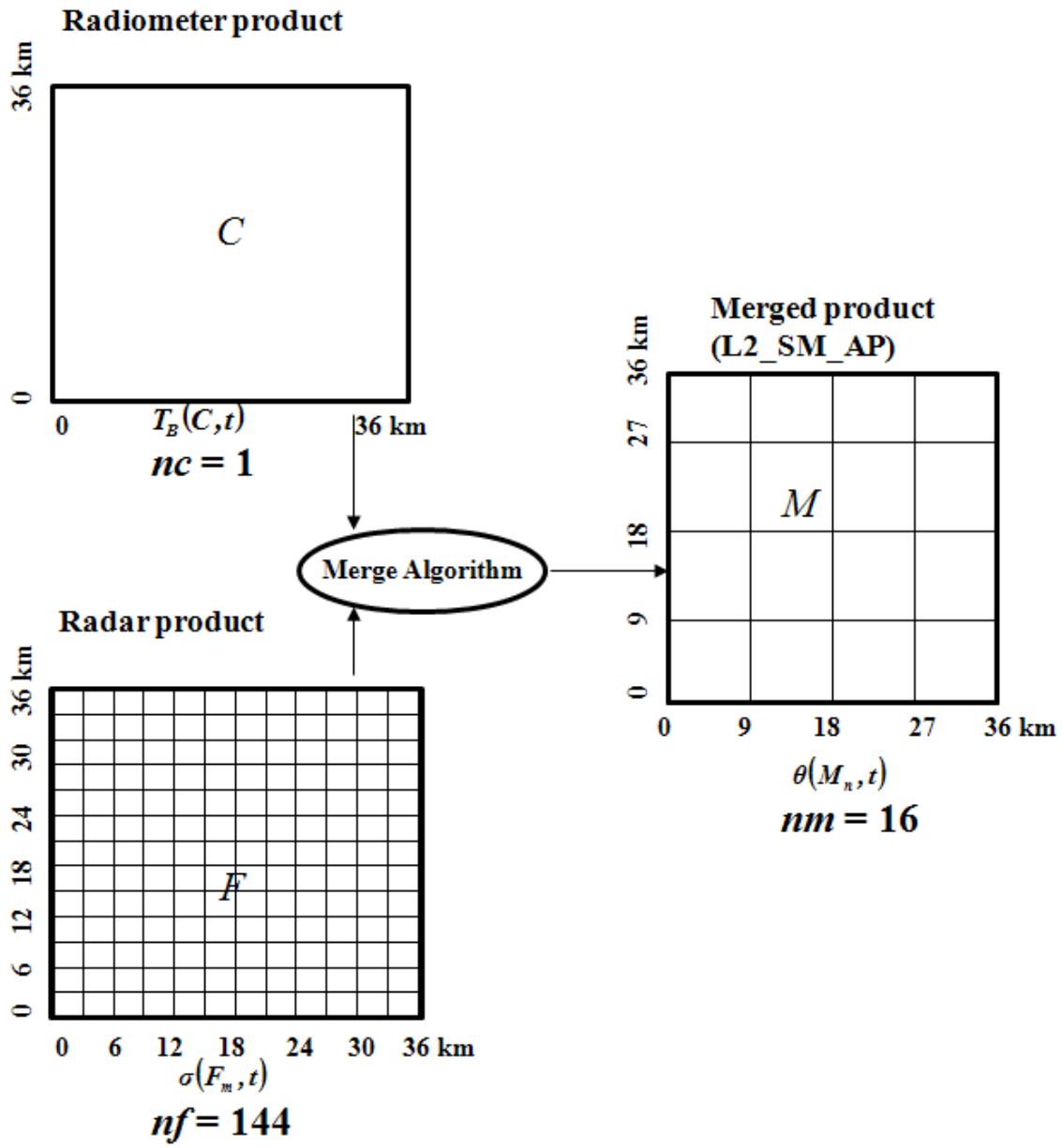


***Figure 2**: Grid definition of radiometer, radar, and merge product where nf and nm are the number of area pixels of radar and merged product, respectively, within one radiometer area pixel nc.*

3.2 Formulation of Baseline L2_SM_AP 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 L-band radar will measure the energy backscattered (σ_{pp}) from the land surface after transmission of an electromagnetic pulse. With concurrent SMAP radiometer and radar data and the SMAP constant look angle measurement approach over a particular region on the Earth, the influence of azimuthal and viewing-angle dependent factors are minimized. Over a short period of time, the increase of surface soil moisture or soil dielectric constant will lead to a decrease in radiometer T_B [6] and an increase in radar σ [7] 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 landuse 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 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), and is shown using the PALS data in upcoming sections. 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* (36 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 = 3 km grid resolution and *nf* is the number of F grid cell within C (Fig. 2). The parameter $\beta(C)$ can be statistically estimated based on the time-series regression in (12), i.e. pairs of SMAP radiometer $T_{B_p}(C)$ and spatially-averaged radar data $\sigma_{pp}(C)$ from successive overpasses over the same Earth grid are used in the statistical linear time-series regression $T_{B_p}(C) = intercept + slope \cdot \sigma_{pp}(C)$. Clearly these parameters are effective across scale C.

As a test of robustness of the assumption of linear functional relationship (12) between brightness temperature and radar backscatter cross-section, data from PALS taken during the Soil Moisture Experiment 2002 (SMEX02) are used to show the strength of linear functional dependence (i.e., R^2) between the time-series of T_{B_p} K (4 km) and σ_{pp} dB (4 km) specific to a particular location or coarse radiometer pixel (Fig. 3). The lower panel of Fig. 3 shows that the explained variance (high R^2) of the linear approximation $T_{B_h} = \alpha + \beta . \sigma_{vv}$ is between 65% and 93% for the SMEX02 PALS observations. There were only 8 days of PALS flights during SMEX02 and in some locations within the larger flight domain inadequate soil moisture changes occurred. The R^2 are slightly lower for the middle region of the PALS domain (Fig. 3) because this region

does not experience enough transition or dynamic range in surface soil moisture conditions as compared to the remainder of the PALS domain.

Another test is shown in Fig. 4 to emphasize the linear functional relationship between T_{B_p} 800 m and σ_{pp} 800 m using the SMEX02 PALS data. Pairs of T_{B_p} and σ_{pp} are binned based on Radar-Vegetation-Index (RVI[1]) for the entire PALS data of the SMEX02 campaign. Irrespective of spatial context, the scatter plots of Fig. 4 illustrate near linear trend between T_{B_p} and σ_{pp} for all the range of RVI except for heavily vegetated regions (RVI: 0.8 – 1). The parameter β that indicates backscatter change sensitivity to brightness temperature change is confirmed to be highly dependent on vegetation characteristics. Values of β for different classes of RVI show that dense vegetation cover masks the soil moisture sensitivity (β approaches near zero for RVI approaching unity). Across low vegetation cover regions (low RVI), the changes in radiometer brightness temperature are also reflected in changes in radar backscatter, leading to large (negative) values of the statistically-estimated β.

[1]The RVI is defined as

$$\mathrm{RVI} = \frac{8*\sigma_{hv}}{\sigma_{vv}+ \sigma_{hh}+ 2*\sigma_{hv}}$$

where the radar backscatter values in units of power [15]. RVI is an index that is directly proportional to the amount of vegetation on the land surface. It can be derived directly from SMAP radar measurements. When the vegetation cover is dense and there is complete volume scattering from the vegetation canopy. For complete volume scattering $\sigma_{vv} = \sigma_{hh} = 3\,.\sigma_{hv}$, therefore it makes RVI reach the upper limit of unity. For bare smooth surfaces, the cross-pol radar backscatter cross-section is insensitive to soil moisture and is much smaller than the co-pol values. This leads to a near-zero RVI. Conveniently, the RVI has a dynamic range between zero and unity.

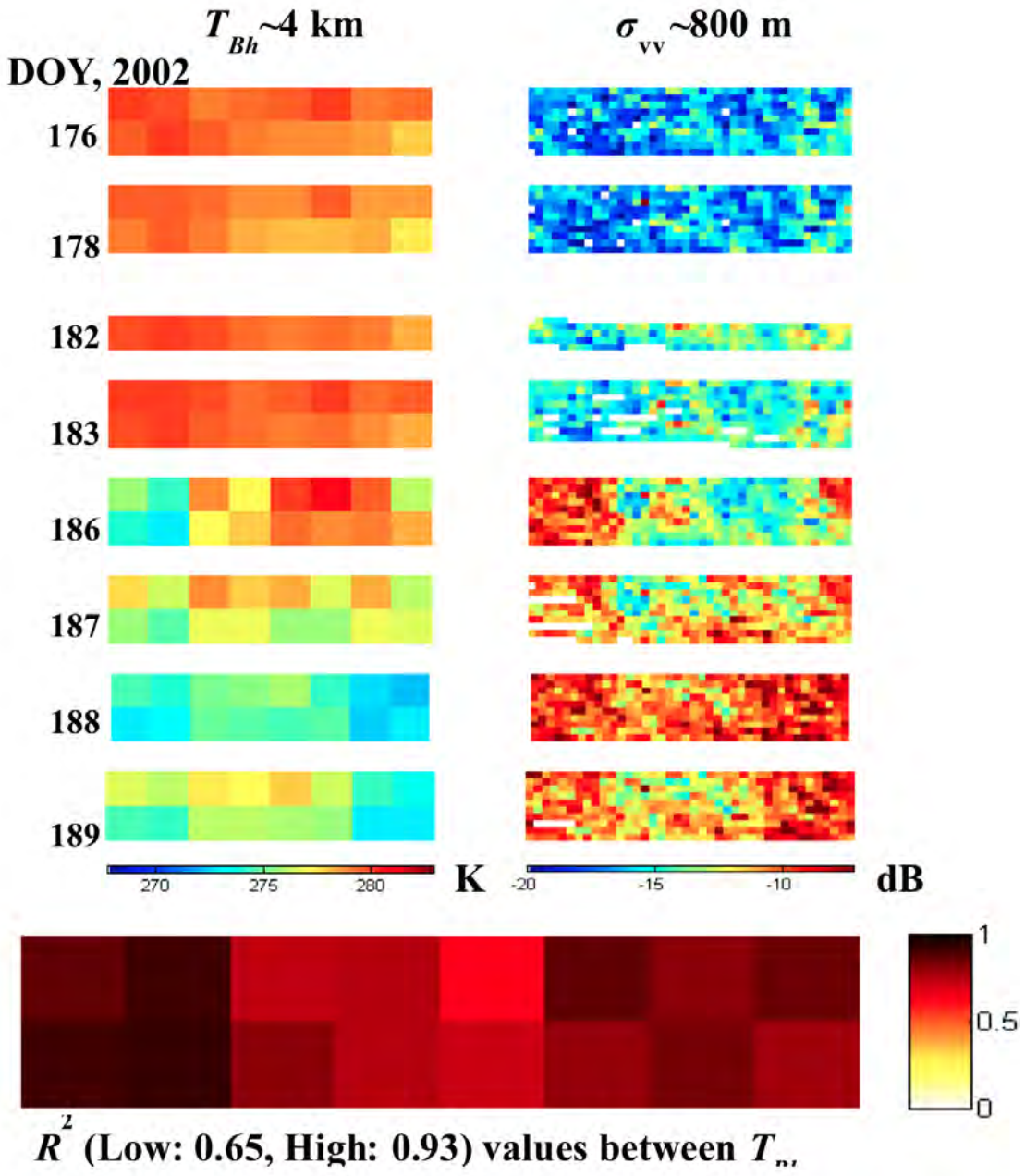


Figure 3: *Gridded SMEX02 PALS L-band radiometer T_{B_h} at ~4 km resolution and radar σ_{vv} at ~800 m resolution for eight days are shown in the top two panels. The lower panel shows the explained-variance or R^2 between the spatially-aggregated (~4 km) T_{B_h} and σ_{vv}.*

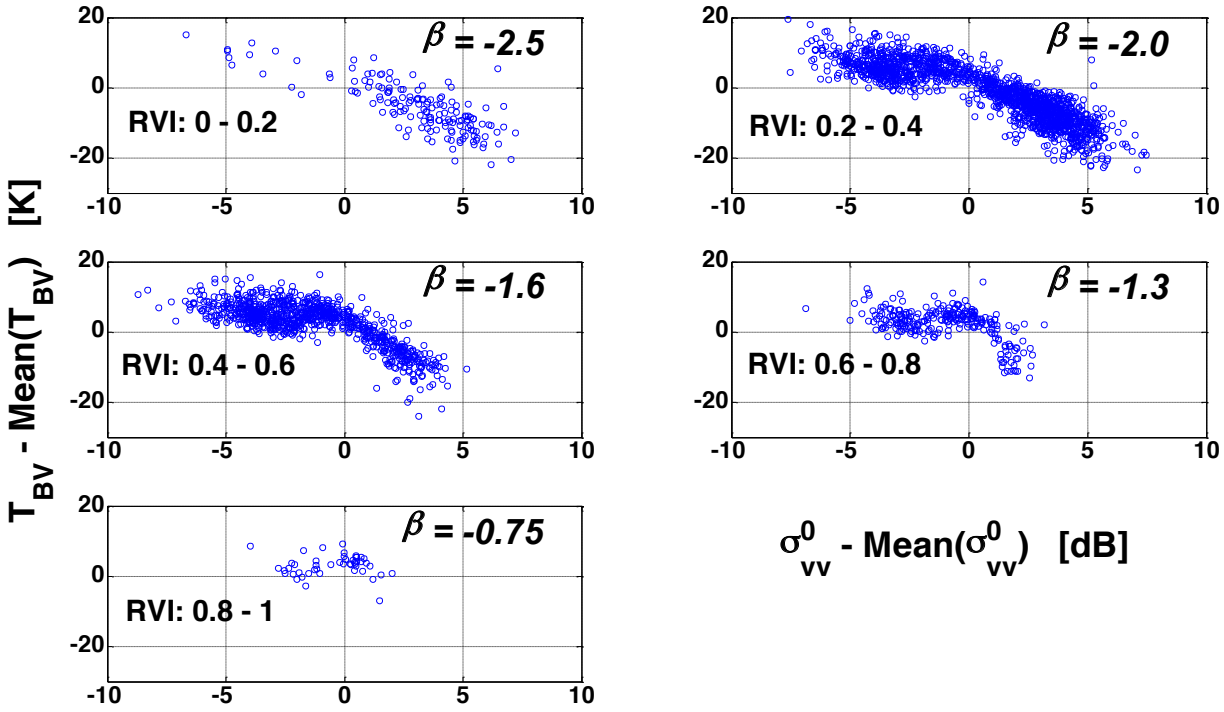


***Figure 4**: Scatter plots of SMEX02 T_{B_v} K and σ_{vv} dB anomalies categorized based on range of RVI.*

To further confirm the fidelity of the linear functional relationship between T_{B_p} and σ_{pp} for different hydroclimatic regions and various landcovers, PALS data from multiple field campaigns (SGP99, SMEX02, CLASIC, and SMAPVEX08) are consolidated and analyzed. During consolidation, campaign-average values of T_{B_p} and σ_{pp} are removed to eliminate campaign-to-campaign variations in PALS instrument and data processing calibration biases. Correlations between different combinations of T_{B_p} and σ_{pp} are then computed. The results are binned based on Radar-Vegetation-Index (RVI), and are shown in Fig. 5. Given the SMAP instrument design, various combinations of T_{B_p} and σ_{pp} are possible for the development of the active-passive algorithm. From Fig. 5, it is apparent that the linear functional dependence between T_{B_v} and σ_{vv} exhibits the highest correlation, and therefore is a preferable combination considered for the proposed active-passive algorithm. Nonetheless further testing of the polarization choices will be carried out during prelaunch phase based on expected airborne data and simulated datasets.

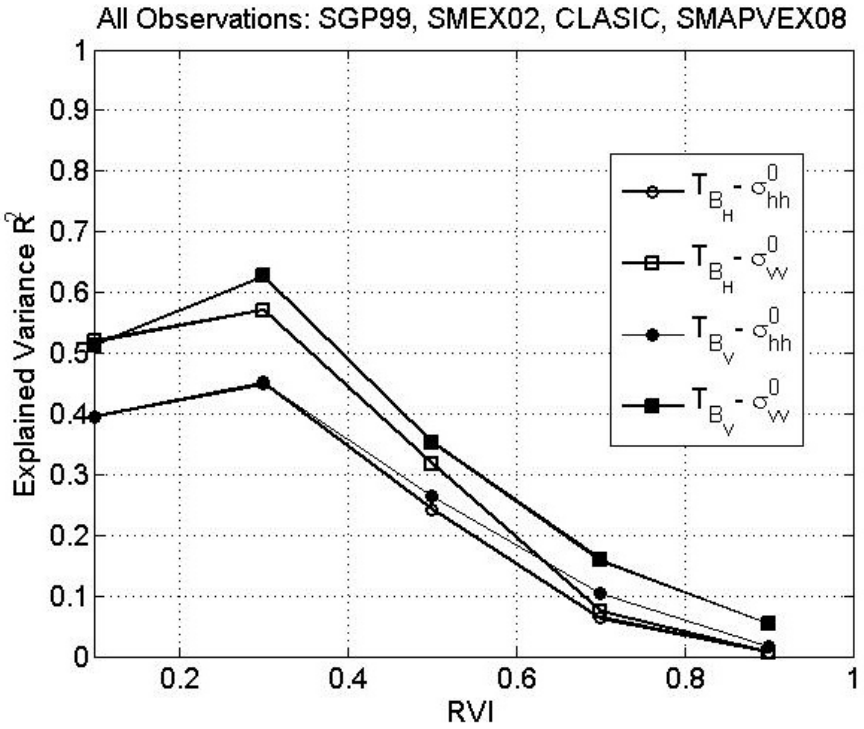


Figure 5*: Correlations between different combinations of T_{B_p} and σ_{pp} evaluated from PALS data taken over four field experiments.*

The slopes (β) of (11) and standard error in β are also estimated from the consolidated PALS data (SGP99, SMEX02, CLASIC, and SMAPVEX08) for different landcover classes (Fig. 6 and Fig 7). The trends of β for all the landcover classes are similar, and the systematic progression in magnitude of β with respect to RVI follows the basic understanding of relationship between T_{B_p} and σ_{pp}. The slope between brightness temperature and radar backscatter changes over time is negative in sign and the magnitude decreases for denser vegetation conditions. Figure 6 also highlights the differences in β due to vegetation types and for different stages of vegetation phenology. High standard errors are observed at RVI value of 0.1 because of low counts of T_{B_p} and σ_{pp} datasets.

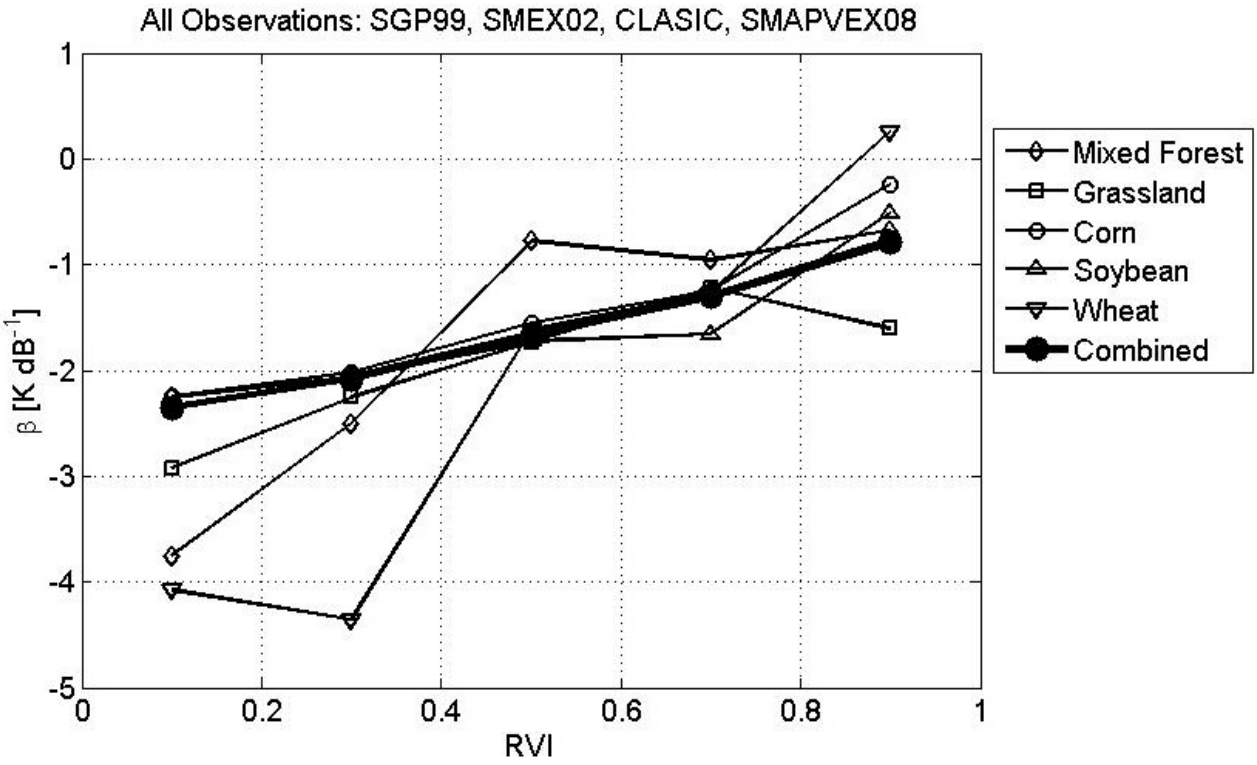


***Figure 6**: β estimates from consolidated PALS data taken over four field experiments (SGP99, SMEX02, CLASIC, SMAPVEX08) for different landcover.*

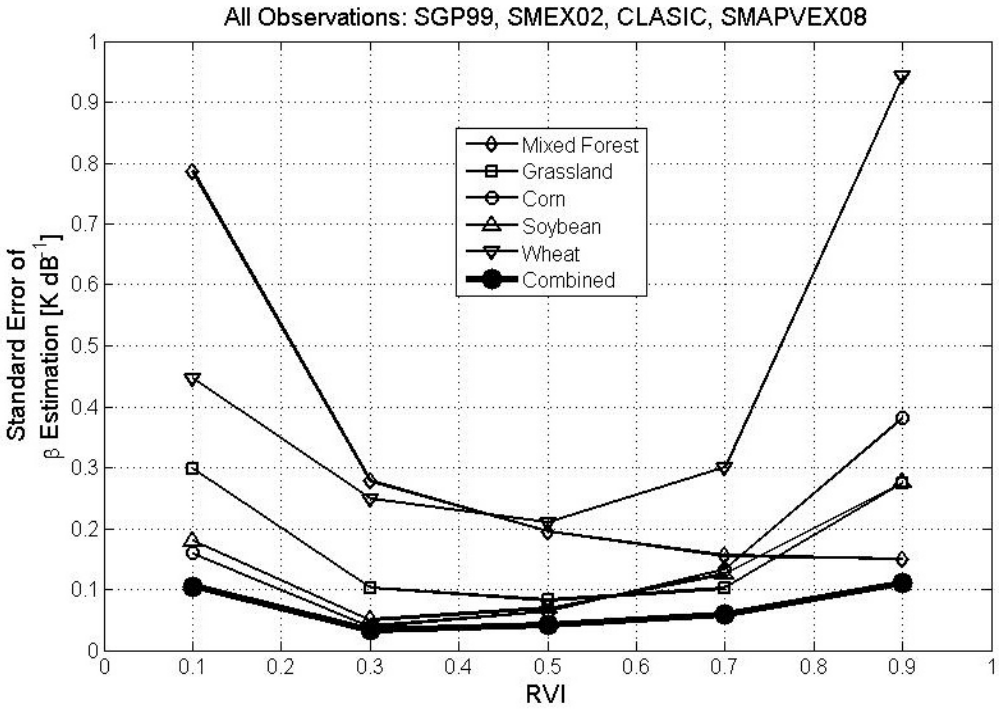


***Figure 7**: Standard errors in estimated β for different landcovers with respect to RVI.*

The statistically-estimated slope parameter $\beta(C)$ in (12) (when based on SMAP measurements) is specific for a given location. This is because $\beta(C)$ is a sensitivity parameter relating $T_{B_p}(C)$ and $\sigma_{pp}(C)$ and it is a function of surface characteristics like the local vegetation cover and soil roughness for a particular period of time. The parameter varies seasonally as well as geographically depending upon landcover. Therefore the time-series pairs of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ used for a location in the regression span a moving-window period over which vegetation phenology and surface characteristics can be considered constant. The length of the temporal window is discussed in next section.

To develop the satellite-based active-passive algorithm further, (11) can also be conceptually evaluated at the scale M (9 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):

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

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

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

The left-hand-side of (15) is the target variable of the active-passive algorithm, i.e. the disaggregated brightness temperature at the 9 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 36 km or scale C. This is the brightness temperature corrected for water body contributions that is produced during the L2_SM_P 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 provides 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.

Consolidated PALS data are used to test the strength of relationship between the co-polarization and cross-polarization backscatter. Significant correlation is observed between the co-pol and cross-pol backscatter. The parameter Γ and standard error in Γ categorized based on RVI are illustrated in Fig. 8 and Fig 9, respectively, for different landcover classes. Γ can be estimated using high-resolution SMAP co-polarization and cross-polarization radar backscatter data within a scale *C* through statistical regression. For any scale C data granule there will be a reasonable number of scale M_j radar data pairs to estimate the sensitivity parameter Γ.

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 $\beta(C) \cdot \Gamma \cdot [\sigma_{pq}(C) - \sigma_{pq}(M_j)]$ 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 $\beta(C) \cdot \Gamma \cdot [\sigma_{pq}(C) - \sigma_{pq}(M_j)]$ to RHS Term III in (15),

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

which can be written more compactly as

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

The disaggregated brightness temperature $T_B(M_j)$ at 9 km is an intermediate product of the proposed active-passive algorithm. Averaging (16) $[\frac{1}{nm}\sum_{j=1}^{nm} T_{B_p}(M_j)]$ over nm pixels result in (18) since the second and third terms are perturbations around mean values. This ensures that the brightness temperature from the active-passive algorithm is consistent with the radiometer brightness temperature. This further ensures close consistency of soil moisture products from passive and active-passive algorithm.

$$T_{B_p}(C) \approx \frac{1}{nm}\sum_{j=1}^{nm} T_{B_p}(M_j) \tag{18}$$

From this point tau-omega (τ-ω) brightness temperature retrieval algorithms with considerable heritage can be used to retrieve surface soil moisture [6, 21, and 22]. Here for preliminary testing the single channel retrieval algorithm in conjunction with high-resolution ancillary data is used to retrieve soil moisture at 9 km.

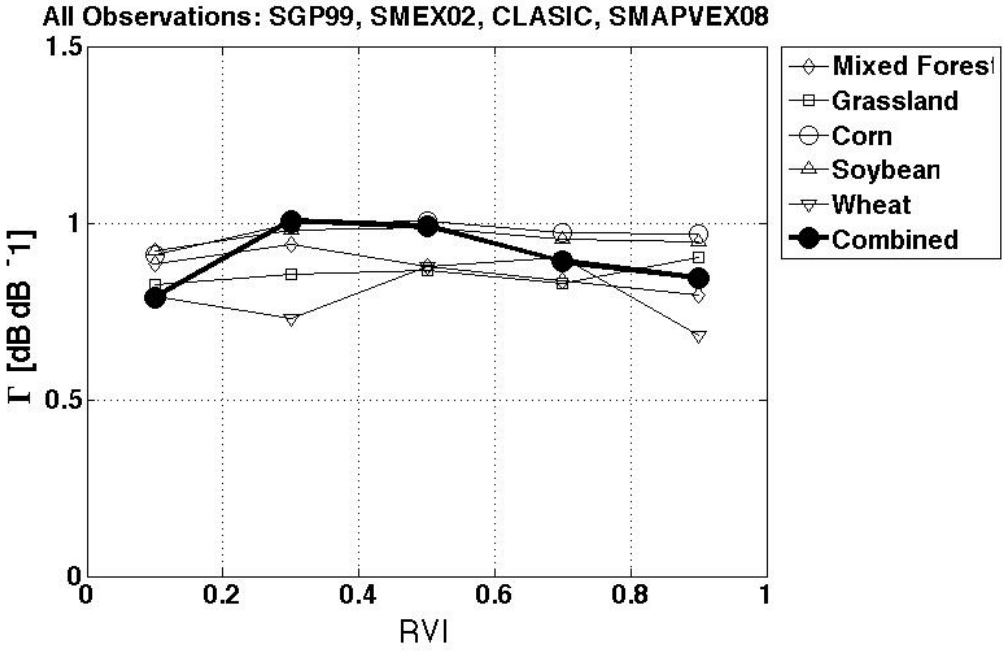


***Figure 8**: Parameter Γ determined using σ_{pp} and σ_{pq} from consolidated PALS measurements plotted with respect to RVI for different landcovers.*

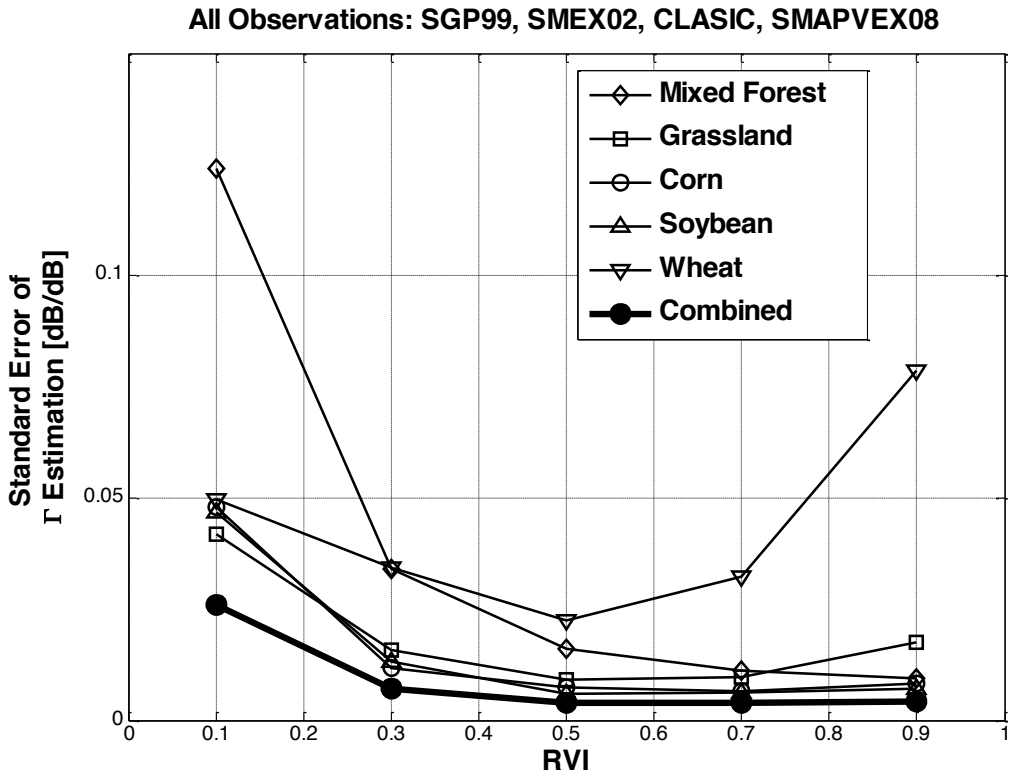


***Figure 9**: Standard errors in estimated Γ for different landcovers with respect to RVI.*

Estimation of the β and Γ Parameters

The performance of the brightness temperature disaggregation that results in the 9 km soil moisture retrievals is heavily dependent on robust estimates of the parameters β and Γ in (17). Regression of the time-series (formed based on multiple overpasses) for $T_{B_p}(C)$ and $\sigma_{pp}(C)$ are used to statistically estimate β. The statistically-estimated slope parameters are specific for a given location and reflect the local roughness and vegetation cover conditions. The parameters vary seasonally, therefore the time-series pairs of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ used for a location in the regression should be limited to a finite-length moving-window over which vegetation phenology can be considered constant. Depending on the dominant land cover vegetation, this may be as short as a few weeks for croplands and a few months for natural landscapes, especially those with evergreen (tropical or boreal) plant types. The robustness will be discussed further below. This translates to anywhere from two to about twenty or at most thirty pair of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ for the regression analysis and statistical estimation of β. The issue is further complicated by the fact that robust estimation is possible only if there is adequate soil moisture variation within the window period (drydown or wetting event) to cause variations in both $T_{B_p}(C)$ and $\sigma_{pp}(C)$ to allow estimation of a slope (β). Similarly Γ is statistically estimated using pairs of $\sigma_{pp}(M_j)$ and $\sigma_{pq}(M_j)$.

In the implementation of the brightness temperature disaggregation algorithm, the regression-based values of β and Γ will be estimated in a moving time-window, and will be combined statistically with a prior estimates to yield a robust value for this parameter. The prior estimate of β is based on β-RVI relationships established at the same and other pixels over time. Observations from airborne field experiments show that RVI is indeed a unique and reliable estimator of β (Fig. 6). More importantly, RVI isolates the impact of vegetation and separates the effect of surface roughness. This is shown in Fig. 10 (a-c) where all PALS active and passive airborne observations from the SGP99, SMEX02, CLASIC and SMAPVEX08 are pooled together to statistically estimate β at every pixel within the experiment region. The β parameters are then stratified according to three indicators for vegetation cover. The first indicator used is the VWC estimated from optical and near-infrared vegetation classification. In Fig. 10a the β parameter stratified based on the dominant optical-based vegetation index does not show any sensitivity; all classes of optically derived VWC yield essentially the same mean β. In Fig. 10b the PALS-observed cross-pol radar backscatter cross-section σ_{hv} is used as a vegetation classifier. Higher values of this cross-pol measurement indicate heavier vegetation cover. As evident in Fig. 10b, the β parameter is indeed closer to zero at high σ_{hv} indicating strong attenuation of the radar and radiometer signals as expected with denser vegetation. The magnitude of β decreases as the value of σ_{hv} is reduced indicating less vegetation cover. However at very low values of σ_{hv} indicating nearly bare surface conditions, the value of β again increase and approaches near zero. Thus, the fluctuations in $T_{B_p}(C)$ are apparently not evident in $\sigma_{pp}(C)$. Such behavior is expected of 'smooth' surfaces for the radar measurements. Based on the field experiment measurements summarized in Fig. 10b, it is evident that the cross-pol σ_{hv} alone is not a unique indicator of a prior value for β. Fig. 10c shows the same β values but now stratified according to the radar-based RVI. Clearly, RVI is a unique indicator of the vegetation cover contributions to β across the dynamic range of RVI. Similarly the Γ parameter priors are related to RVI in the L2_SM_AP implementation.

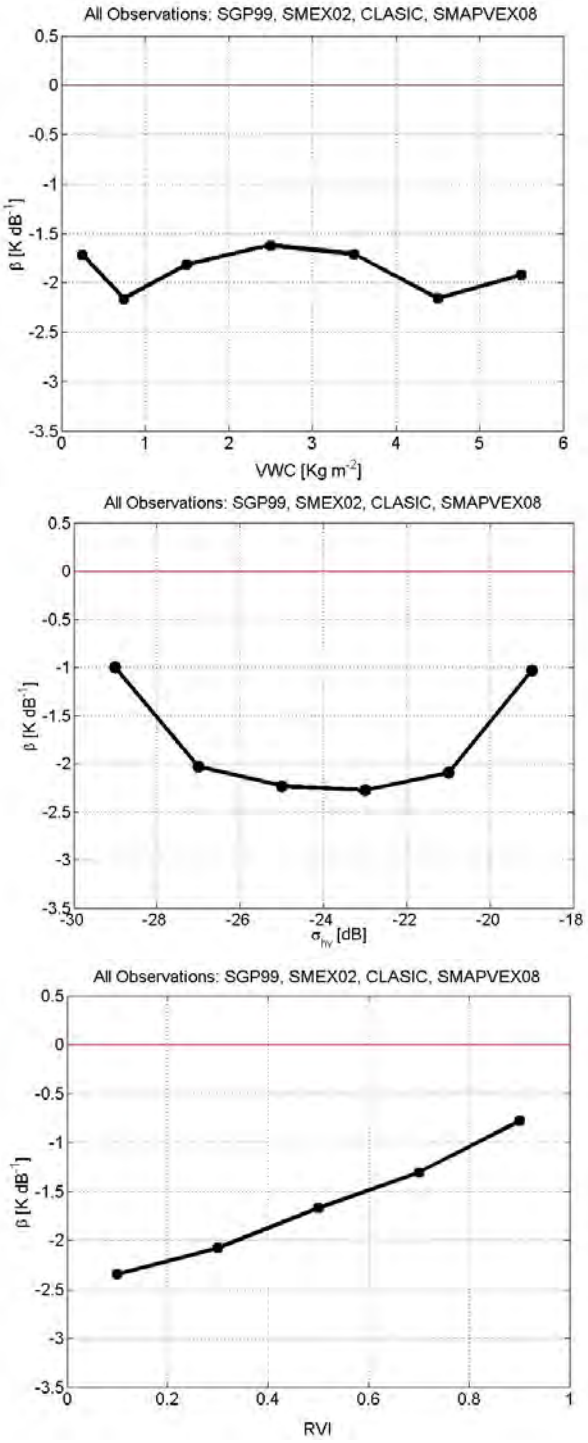


Figure 10*: Values of β estimated from PALS radar and radiometer measurements during SGP99, SMEX02, CLASIC and SMAPVEX08 segmented and plotted versus: a) optical- and infrared-derived vegetation classification and VWC, b) cross-pol radar backscatter cross-section, and c) RVI as in (11).*

In the baseline algorithm the values of β and Γ are estimated using regression of pairs of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ over the finite and moving window of past SMAP radar and radiometer measurements and pairs of $\sigma_{pp}(M_j)$ and $\sigma_{pq}(M_j)$, respectively, over land regions that are within the

retrieval mask. The regression yields an error of estimation as well as the expected statistical slopes β and Γ. An example of temporal window created using MODIS landcover data is illustrated in Fig. 11. The values of Fig. 11 act as an initial estimate (priors) of temporal window length that is used before the SMAP launch. It is expected that every *C*-scale grid cell will have its unique temporal window length due to it natural characteristics of vegetation types and heterogeneity, topography and soil types. The temporal window length will be updated accordingly with the availability of SMAP data.

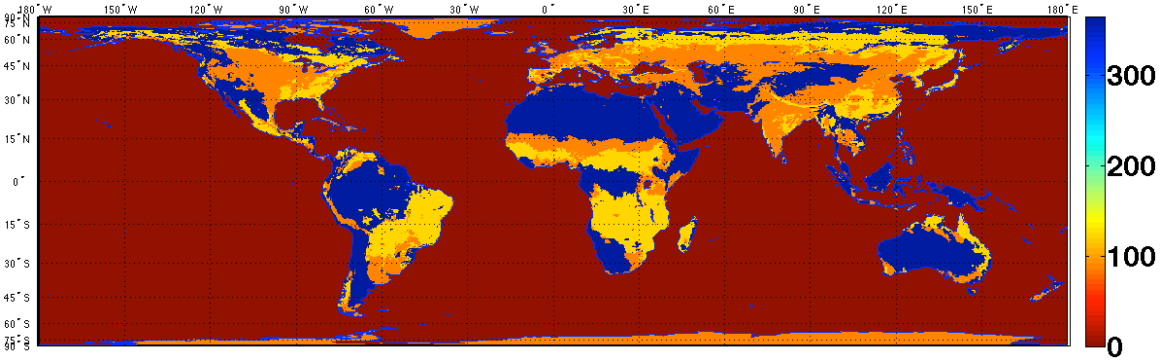


Figure 11: *Initial estimate of temporal window length based on MODIS landcover data used to determine the time series pairs of* $T_{B_p}(C)$ *and* $\sigma_{pp}(C)$ *required for* β *estimation.*

There are also prior estimates for β based on the RVI for the pixel and β-RVI relations established with field campaign datasets (Fig. 6). The error of estimation of this prior estimate is $MSE(\beta_{prior})$ and it is shown in Fig. 7. These errors of estimate will be updated with further field experiment data as they become available. The prior β and the time-series estimates of β are merged together in a Bayesian framework. Essentially they are combined to form a posterior estimate of β, ie., $\beta_{posterior}$ as defined in (19). The weights for the merger are the ratios of the errors of estimation to the sum of the errors of estimation

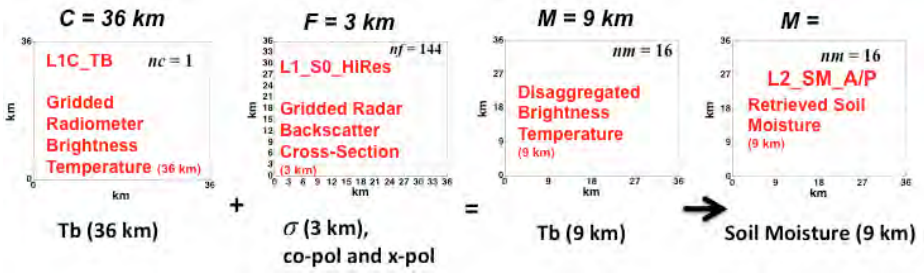
$$\beta_{posterior} = \frac{MSE(\beta_{prior})}{MSE(\beta_{reg})+MSE(\beta_{prior})} * \beta_{reg} + \frac{MSE(\beta_{reg})}{MSE(\beta_{reg})+MSE(\beta_{prior})} * \beta_{prior} \qquad (19)$$

where MSE is Mean-Square-Error, β_{prior} is a prior estimate of β that is derived from all relevant field experiments data (Fig. 7 and Fig. 9), and β_{reg} is the statistically estimated slope using regression on pairs of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ over the finite and moving window of past SMAP radar and radiometer measurements. Ultimately, $\beta_{\text{posterior}}$ will be used in (17). It is expected that the accuracy of $\beta_{\text{posterior}}$ will improve over the SMAP mission duration due to more availability of relevant data for $T_{B_p}(C)$ and $\sigma_{pp}(C)$. A similar procedure is followed for and Γ.

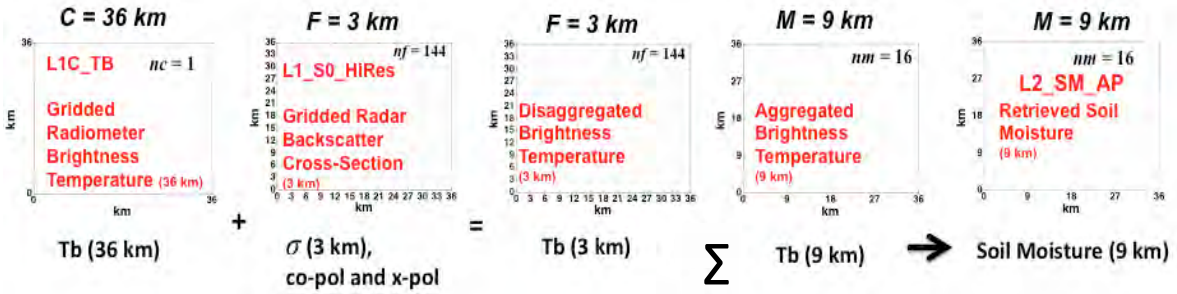
3.3 Option Algorithms

The current option algorithms are variants of baseline algorithm. The differences are in the way baseline algorithm is implemented. Schematics of process flow are elaborated below that show differences of option algorithms from baseline algorithm implementation.

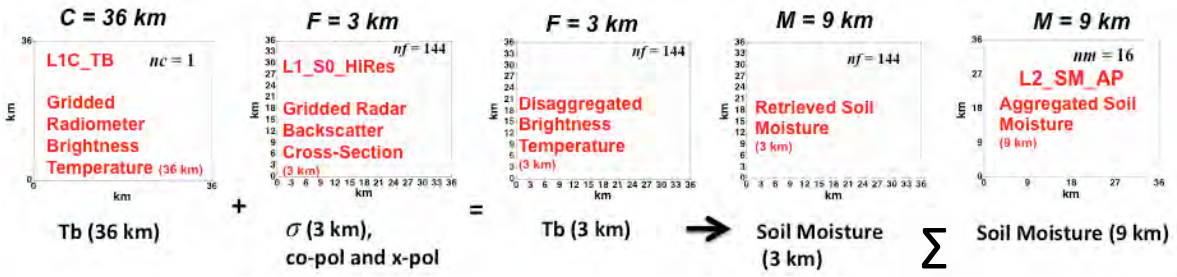
Schematic of baseline algorithm implementation:



Schematic of Option1 algorithm implementation:

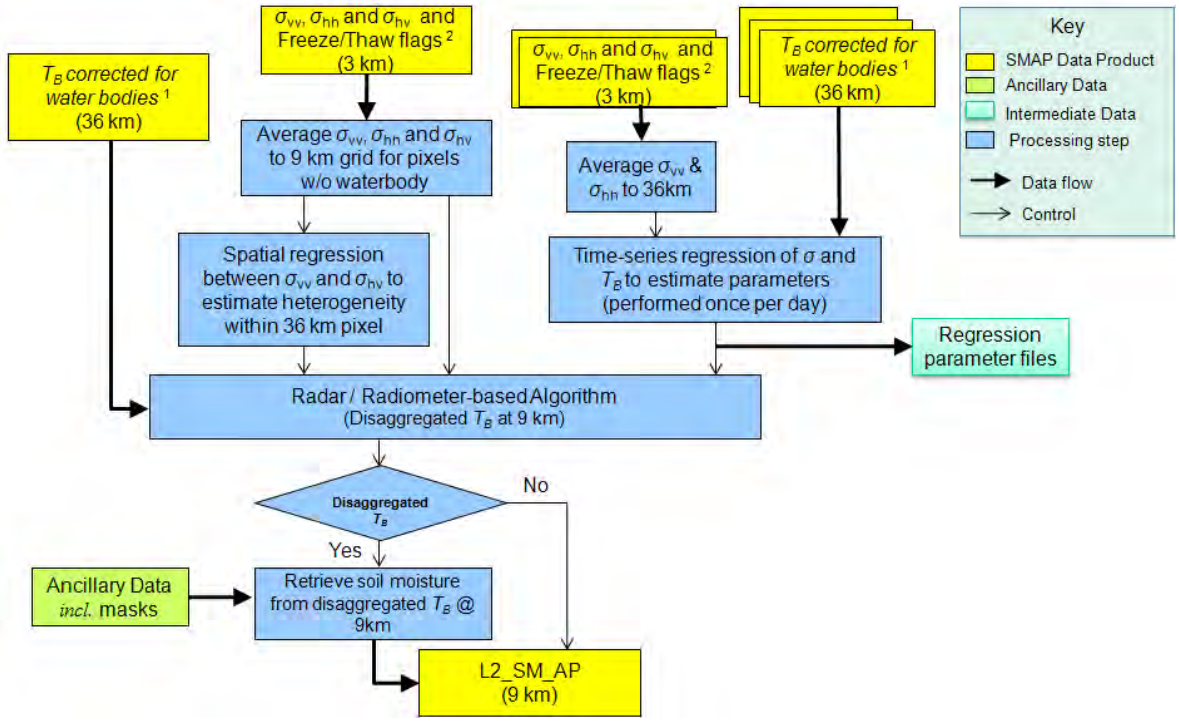


Schematic of Option2 algorithm implementation:



3.4 Process Flow of the Baseline L2_SM_AP Algorithm

Schematic below illustrates a simplified version of process flow of the current L2_SM_AP baseline algorithm. The process flow captures important processes that involve input data streams and static and ancillary data used in the algorithm to generate the L2_SM_AP soil moisture product.



Process flow diagram of the baseline L2_SM_AP algorithm. The disaggregated TB at 9 km are converted to soil moisture using algorithms described in L2_SM_P but based on 9 km resolution ancillary data.

4 TESTS OF ALGORITHMS

The performance of the baseline algorithm is evaluated using two distinct types of datasets to assess different aspects of the algorithm. Airborne active-passive PALS instrument has been flown as a SMAP simulator in several field campaigns. The campaign data are used to test the disaggregation algorithm (17). Simulated datasets are used to demonstrate the feasibility of ingesting the inputs based on the SMAP configuration. The following subsections elaborate these tests.

4.1 Test of Baseline Algorithm Using PALS Data

This study applies the real field data from PALS to the L2_SM_AP baseline algorithm. Two field campaigns (SMEX02 and SMAPVEX12) airborne data from PALS are used to evaluate the baseline active-passive algorithm. The PALS data from these field campaigns represent two distinct scenarios. The SMEX02 study site in Walnut Creek, Iowa is topographically flat with vegetation (mostly maize) and soil type (clayey-loam) that are relatively homogeneous. In contrast, the SMAPVEX12 study site in Manitoba, Canada is heterogeneous in vegetation (many different crop types and forested region) with diverse soil types (sandy to clayey). Therefore testing and successful implementation is essential to evaluate the merits and limitations of L2_SM_AP baseline algorithm.

4.1.1 **Test Using PALS Data from SMEX02**

PALS L-band radar and radiometer have similar frequencies like SMAP; however, PALS data from SMEX02 have much finer spatial resolution (~0.8 km). To apply the baseline algorithm to PALS data, the data are gridded for the radiometer at ~4 km and the radar at ~0.8 km (Fig. 3). Equation (17) is implemented to obtain the downscaled T_{B_v}at ~0.8 km and single channel algorithm retrieved soil moisture at ~0.8 km resolution as shown in Fig. 12. To validate the retrieved soil moisture estimates at ~0.8 km resolution, the field averaged soil moisture calculated from *in situ* measurements in 31 fields over 4 days are used (Fig. 13). The representative spatial resolution of a field is nearly ~0.8 km making the comparison spatially compatible.

Minimum Performance Algorithm

To evaluate the baseline algorithm against a reference, minimum performance soil moisture retrieval is used. The minimum performance soil moisture values are retrieval from T_{B_v} at ~0.8 km that are obtained by direct re-sampling (simple assignment) of T_{B_v}at ~4 km to high resolution ~0.8 km pixels. It is essentially a resampling of brightness temperature to finer scale without use of information from the radar. Fig. 13 shows that the baseline algorithm with RMSE 0.033 [cm^3 cm^{-3}] outperforms the minimum performance with RMSE: 0.056 [cm^3 cm^{-3}]. The difference between the baseline algorithm and the minimum performance algorithm is a measure of the value of radar measurements in the active-passive data product.

Another test was performed to evaluate the contribution of radar cross-polarization backscatter measurements (σ_{pq}) to the algorithm (17). The radar σ_{pq} addresses the heterogeneity especially due to vegetation within the coarse radiometer footprint as discussed in Section 3.2. The test was conducted by ignoring the radar cross-pol measurement (σ_{pq}) in the algorithm (17) or essentially setting $\Gamma = 0$. Therefore (17) becomes

$$T_{B_p}(M_j) = T_{B_p}(C) + \beta(C) \cdot \left[\left(\sigma_{pp}(M_j) - \sigma_{pp}(C)\right)\right] \quad (22)$$

The retrieval using (22) is shown in Fig. 14. The 0.043 [cm^3 cm^{-3}] RMSE of this scenario is greater than the 0.033 [cm^3 cm^{-3}] RMSE obtained from the baseline algorithm (17). This test clearly illustrates the important contribution of σ_{pq} in capturing sub-radiometer measurement scale vegetation heterogeneity for the baseline algorithm.

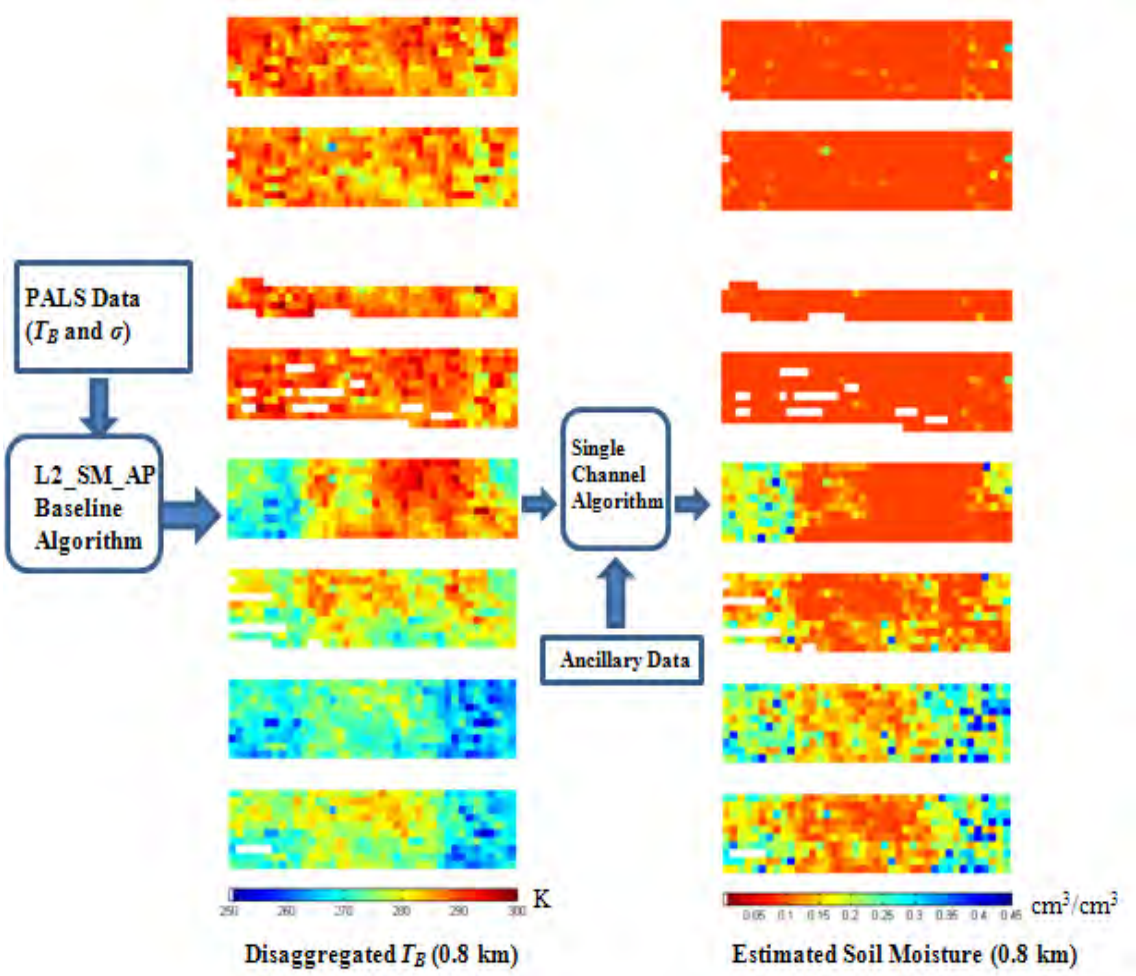
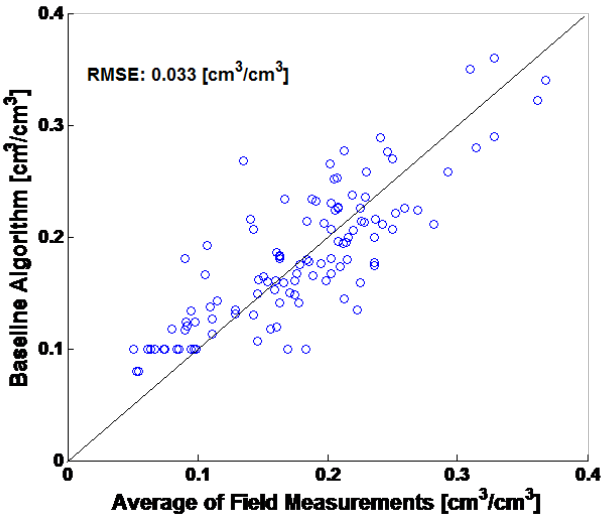
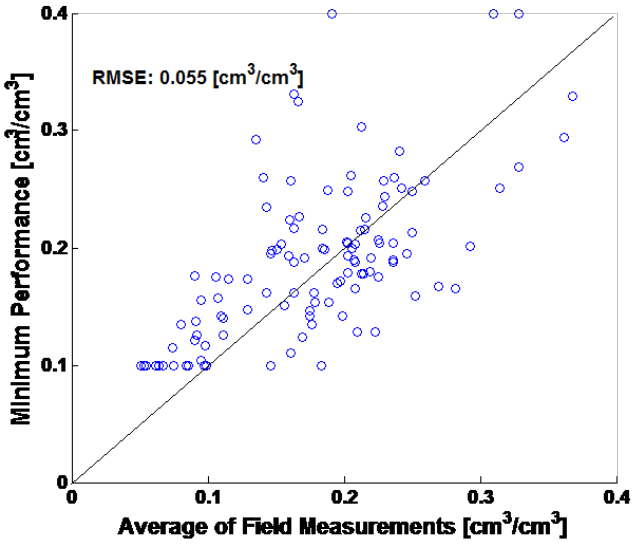


Figure 12: *Disaggregated T_{B_v} and retrieved soil moisture estimates at 0.8 km resolution.*



A)



B)

Figure 13*: Plots of averaged soil moisture from field measurements and retrieved soil moisture estimate from the SMEX02 PALS data for 8 days. A) baseline algorithm, and B) minimum performance.*

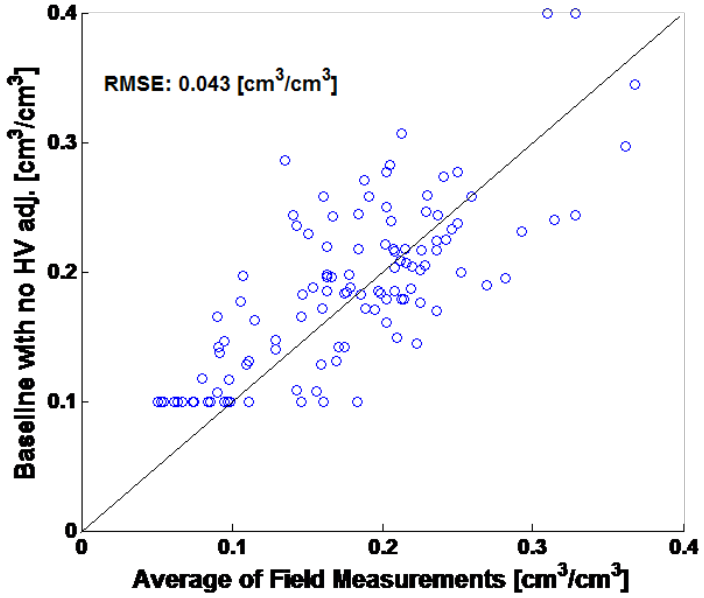


Figure 14*: Plots of averaged soil moisture from field measurements and retrieved soil moisture estimate from the SMEX02 PALS data for 8 days when the retrieval is performed with no cross-pol (hv) radar backscatter information.*

The above results show successful implementation of L2_SM_AP baseline algorithm using PALS data from SMEX02. It also highlights the potential of L2_SM_AP baseline algorithm to downscale brightness temperature over homogeneous region with rapidly changing vegetation attributes and with varying soil moisture conditions during the study period.

4.1.2 **Test Using PALS Data from SMAPVEX12**

Implementation of active-passive baseline algorithm (17) was also performed on SMAPVEX12 data. The results show reasonable performance under challenging conditions [21]. A hypothesis of temporal dependence of algorithm parameter (β) mostly due to changing vegetation attributes was verified during this implementation.

Figure 15 illustrates the correlation between T_{B_p} and σ_{pp}, when the entire duration (~2 months) of PALS data is used to compute β parameter. Low R_H^2 and R_V^2 values resemble inferior β parameter estimates. Based on the knowledge of ground conditions and vegetation status in the study, the PALS data were divided into two parts (for June and July of 2012) to improve the estimates of β parameter. It is clearly shown in Fig. 16 that R_H^2 and R_V^2 significantly improves when the time series is divided into two parts, verifying temporal dependence attribute of β parameter and better estimates that positively impacts the baseline active-passive algorithm (Eq. 17).

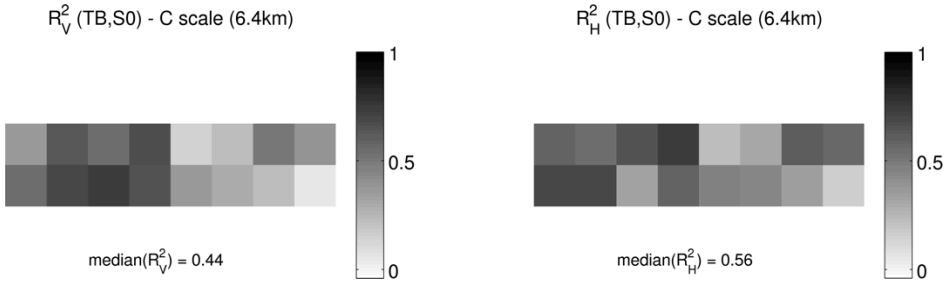


Figure 15. Explained variance of T_{B_p} based on σ_{pp} from PALS using the entire campaign time period and for V and H polarizations. R_V^2 is correlation between T_{B_v} and σ_{vv}, and R_H^2 is correlation between T_{B_h} and σ_{hh}.

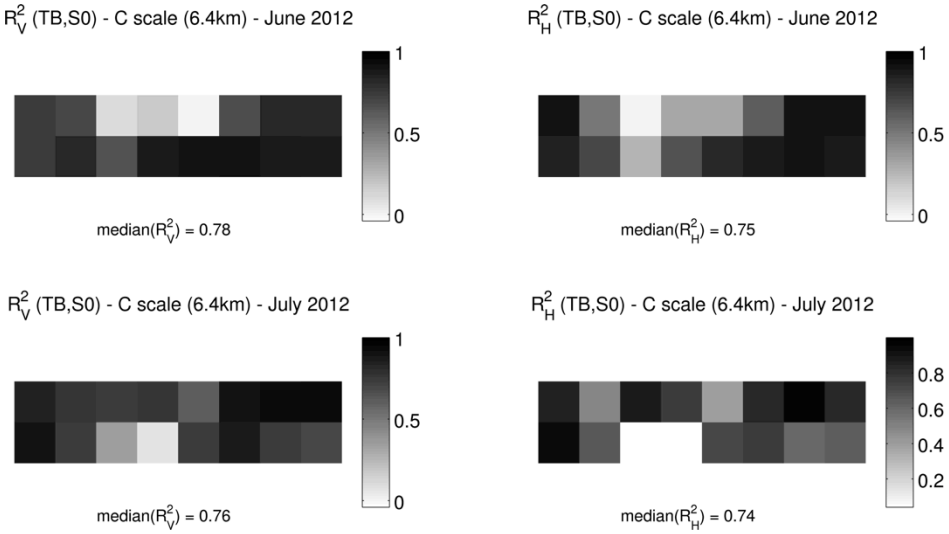


Figure 16*: Explained variance between PALS T_{B_p} and σ_{pp} after dividing the campaign time period into two: June (first row) and July (second row). The explained variance is higher when June and*

July are considered separately. R_V^2 *is correlation between* T_{B_v} *and* σ_{vv}*, and* R_H^2 *is correlation between* T_{B_h} *and* σ_{hh}*.*

The derived β parameters of the two SMAPVEX12 periods are used in active passive algorithm and some of the results are presented subsequently. Figure 17 shows the baseline active-passive algorithm clearly out performs the minimum performance most of the time. The disaggregated T_{B_p} were then subjected for soil moisture retrievals and compared against the in situ soil moisture values measured in coinciding fields. Table 3 highlights the performance statistics of baseline algorithm.

Figure 17*: Spatial distribution of the RMSE ratios between the disaggregated* T_{B_p} *and the minimum performance compared to the gridded* T_{B_v} *from high resolution PALS data. Ratio lowers than 1 (light to dark blue) denotes an improvement and bigger than 1 (light to dark red) a degradation of the disaggregation compared to the minimum performance.*

Table 3: Performance statistics of the baseline active-passive algorithm based retrieved soil moisture compared to in situ measurements.

Crop Type	Canola		Corn	
	Active-Passive	Minimum Perf.	Active-Passive	Minimum Perf.
RMSE [cm^3/cm^3]	0.04	0.046	0.029	0.035
Bias [cm^3/cm^3]	0.027	0.001	-0.004	-0.020
Unbiased RMSE [cm^3/cm^3]	0.031	0.046	0.029	0.030

Two examples of successful implementations (Section 4.1.1 and Section 4.1.2) of L2_SM_AP baseline algorithm are presented above. Future postlaunch field campaigns data will also used to further understand dynamics of L2_SM_AP algorithm and its parameters, and tune the tau-omega model for soil moisture retrievals for specific landcovers and crops.

4.2 Test of Baseline Algorithm using Simulated Data

4.2.1 Global Scale Simulation

In order to test the L2_SM_AP algorithm across a wider range of conditions than those encountered in the limited airborne field campaigns, simulation environments are also used. A global-scale simulation (GloSim) for the SMAP mission is developed and implemented on the SMAP Testbed at the SMAP Science Data System (SDS). The complete detail of Testbed mechanism and implementation is provided in SMAP Testbed document (SMAP Science Data

Management and Archive Plan. SMAP Project, JPL D-45973, Jet Propulsion Laboratory, Pasadena, CA).

The GloSim orbit simulator on SDS mimics the SMAP configuration and follows an 8-day exact repeat pattern that provides total global coverage in 2-3 days. GloSim includes the capability of generating orbital files of simulated radiometer and radar observations of T_B and σ, respectively. Geophysical data (e.g., soil moisture and soil temperature) obtained from GMAO MERRA at 9 km resolution covering one year and ancillary data (e.g., model parameters, soil texture, landcover, water bodies, and VWC) at high resolution (≤9 km; e.g., MODIS-derived VWC and land classification) are used as underlying truth maps to sample forward observations of T_B and σ to mimic SMAP-like measurements. GloSim also applies realistic instrument/antenna beam sampling and orbital sampling to simulate the footprint-averaged observations within swaths acquired by the SMAP instruments. These simulated observations, along with their noise-perturbed versions, are essential to the testing, development, and operational implementation of all SMAP Level 2 through Level 3 soil moisture and freeze/thaw algorithms.

L2_SM_AP retrievals are performed on the simulated SMAP half orbit granules of L2_SM_P and L2_SM_A generated from GloSim. The baseline algorithm of L2_SM_AP is implemented for extensive analyses. These analyses help to understand and develop solutions and risk reduction of various important operational and algorithm related issues such as:

1) Size of temporal window required over valid land pixels of radiometer and radar data to derive high fidelity algorithm parameters.
2) To identify regions of the world where a temporal window is essential due to changing vegetation phenology and ground conditions.
3) Develop and mature algorithm parameters database.
4) Develop and mature L2_SM_AP error budget table.
5) Highlight the sensitivities of various ancillary data, masks and flags on the L2_SM_AP retrievals.
6) Assess the limitations of the L2_SM_AP algorithms, and help to tune the algorithms
7) Help mature the L2_SM_AP software development.
8) Assess the computational needs of L2_SM_AP processor.
9) Demonstration of end-to-end integration of L2_SM_AP with the SMAP data process flow in the SMAP Testbed.

The L2_SM_AP processor is successfully implemented and run on the SMAP SDS as illustrated in process flow diagram Fig. 11. The water-body corrected brightness temperature $T_B(C)$ at 36 km along a swath (Fig. 18) with associated QC flags is available as part of the L2_SM_P product, and is a direct input to the L2_SM_AP baseline algorithm. The baseline algorithm also receives as input the 3 km gridded σ_{vv} data and 3 km transient water body and freeze/thaw flags produced during the L2_SM_A processing are input to the L2_SM_AP algorithm. Figure 19 shows aggregated σ_{vv} at 9 km over the same swath (as of Fig. 18) that includes appropriate noise as expected in the nadir region and outer edges of the swath. For discussion and illustration of L2_SM_AP retrievals of the particular swath as shown in Fig. 18 or Fig. 19 is selected because it covers a wide range of conditions in soil moisture state (dry – wet), soil texture (sandy – clayey), landcovers (rainforest – desert), and different hydroclimatic domain.

Along with the SMAP products as inputs, the L2_SM_AP processor implemented in the SMAP SDS also ingests static (e.g., soil) and dynamic (e.g., VWC and soil temperature) ancillary datasets, and masks (e.g., urban, inland water bodies) at 9 km Earth fixed grid over the global extent from the Testbed.

A separate processor is also developed to determine the parameters required in (17). For the GloSim retrievals, the parameter (β) estimation is conducted using time series of $T_{B_v}(C)$ and $\sigma_{vv}(C)$ for all grid cells. The parameters determined from regression of $T_{B_v}(C)$ and $\sigma_{vv}(C)$ time series are quantitatively verified with the parameters obtained using the Aquarius data (refer Section 5.2). Figure 20 illustrates the state of parameter β used over the swath in (17) that clearly exhibits dependency of β with respect to landcover. The landcover mostly influence the dynamic range of $T_{B_v}(C)$ and $\sigma_{vv}(C)$, and hence the parameter β. Another parameter Γ in (17) that detects the heterogeneity within (C) is determined on the fly over the swath and is shown in Fig. 21. Typically very high correlation is observed between σ_{vv} and σ_{hv}, and that is well captured in Fig. 21. The parameter Γ also displays mild dependency for landcover.

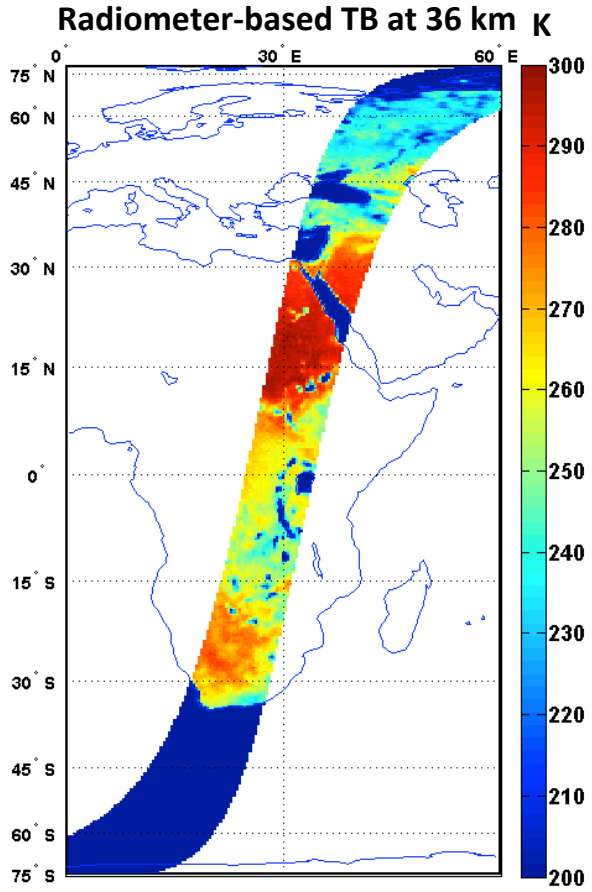


***Figure 18**: $T_{B_v}(C)_{adj}$ data gridded at 36 km from GloSim for one day in May.*

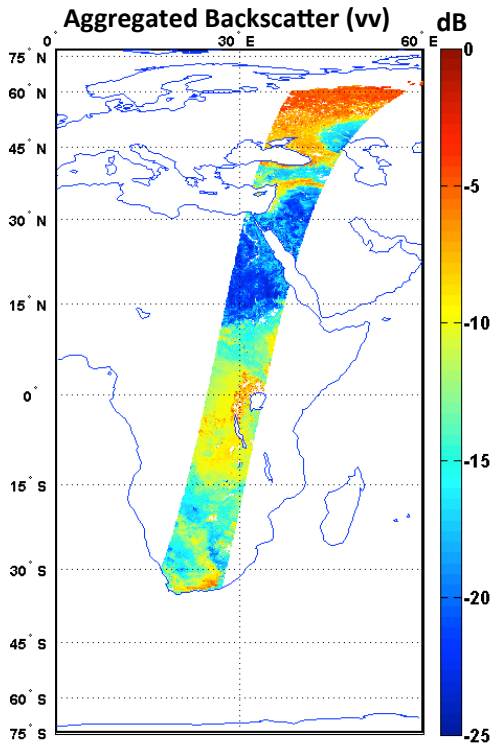


***Figure 19**: Swath of aggregated σ_{vv} at 9 km grid from GloSim for one day in May.*

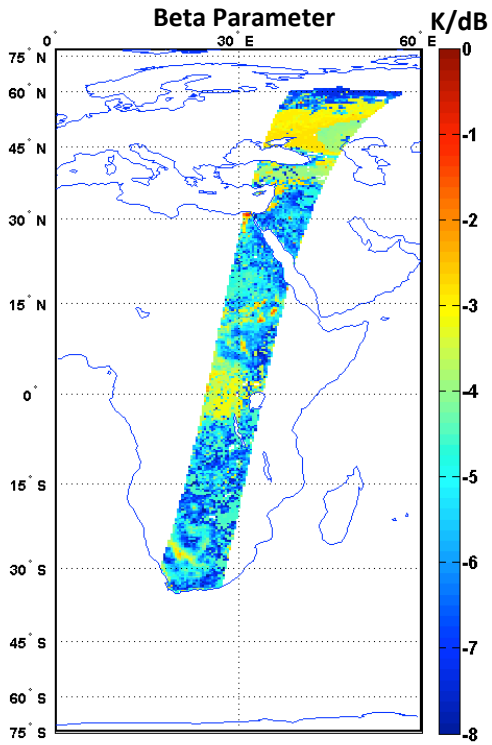


***Figure 20**: Beta (β) parameter used for the swath from GloSim for one day in May.*

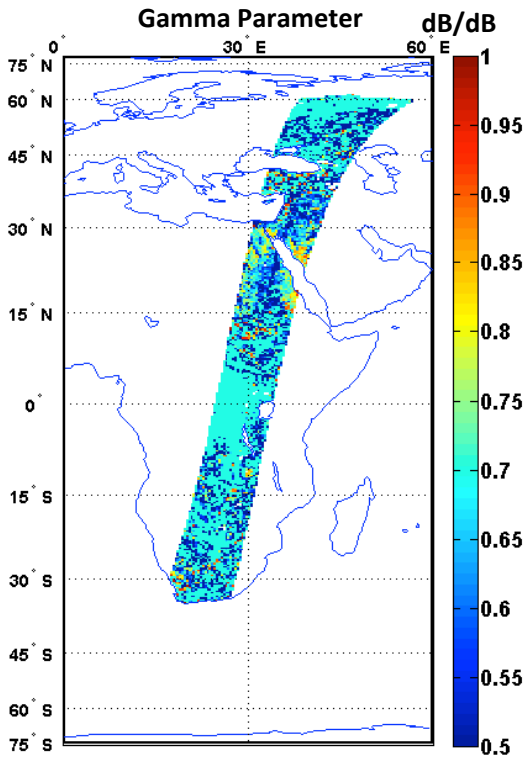


***Figure 21**: Gamma (Γ) parameter used for the swath from GloSim for one day in May.*

The baseline algorithm of L2_SM_AP (17) is applied on data obtained from L2_SM_P and L2_SM_A (Fig. 15-16). Fig. 22 shows that the disaggregated T_{B_v} at 9 km that captures the spatial heterogeneity detected by SMAP radar otherwise masked by coarse scale brightness temperature measurements. Soil moisture retrieval is conducted on the disaggregated T_{B_v} using the tau-omega (τ-ω) model (Single Channel Algorithm). Figure 23 and Fig. 24 show the retrieved and truth soil moisture at 9 km, respectively. Appropriate noise is introduced in the ancillary data while performing the retrievals. Comparison of Fig. 23 and Fig. 24 show similar spatial patterns of soil moisture for most of the regions. However, over the high-vegetated regions (e.g., rain forest) the τ-ω model did not converge and therefore a null value is assigned during the retrievals process. To get initial assessment of L2_SM_AP retrievals, errors are computed for the swath and are shown in Fig. 25, higher error for regions having high vegetation is according to the expectation. The median of errors is plotted in Fig. 26 to ascertain the error structure in soil moisture retrievals according to the stratified mean of VWC.

Nearly ~5300 half-orbit granules of L2_SM_P and L2_SM_A are processed for one year period for the GloSim L2_SM_AP retrievals. Root-Mean-Square-Errors (RMSE) are computed for each 9 km grid cell. Spatial evolution of RMSE at a global extent is shown in Fig. 27 for 6 months (April – September) period. The spatial pattern of RMSE in Fig. 27 matches with the global VWC (Fig. 28) spatial distribution. RMSE are not computed for the 9 km grid cells having more than 5% water fraction within 9 km, more than 25% urban areas, open water bodies, and mountainous region. Quantitative values of RMSE with respect to a range of VWC over the global extent is shown in Fig. 29a. The RMSE curve in this plot clearly meets the SMAP L1 requirements. A histogram is also shown in Fig. 29b that highlights the distribution of standard deviation in retrieved soil moisture obtained from one-month period (June). The median of the histogram clearly meets the SMAP L1 requirement of 0.04 cm^3/cm^3.

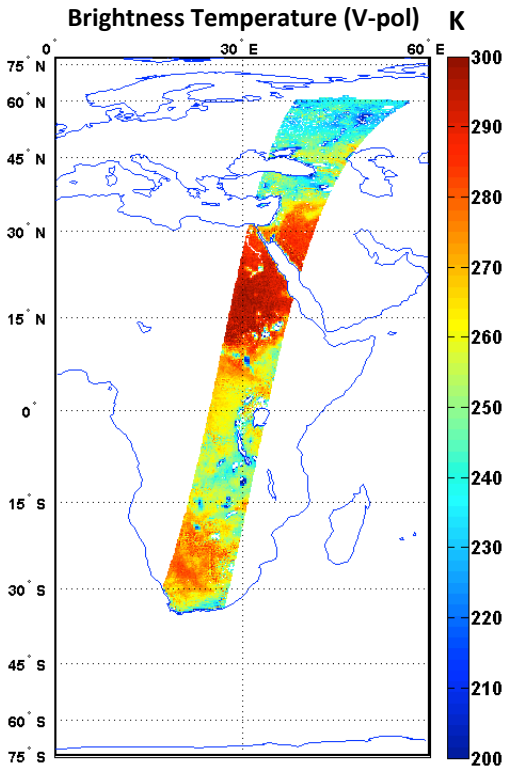


***Figure 22**: Swath of disaggregated T_{B_v} at 9 km grid from GloSim for one day in May.*

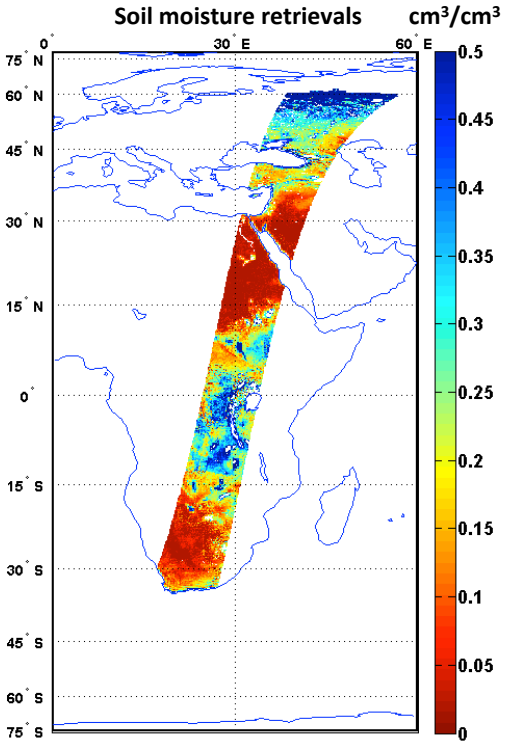


***Figure 23**: Swath of retrieved soil moisture at 9 km grid from GloSim for one day in May.*

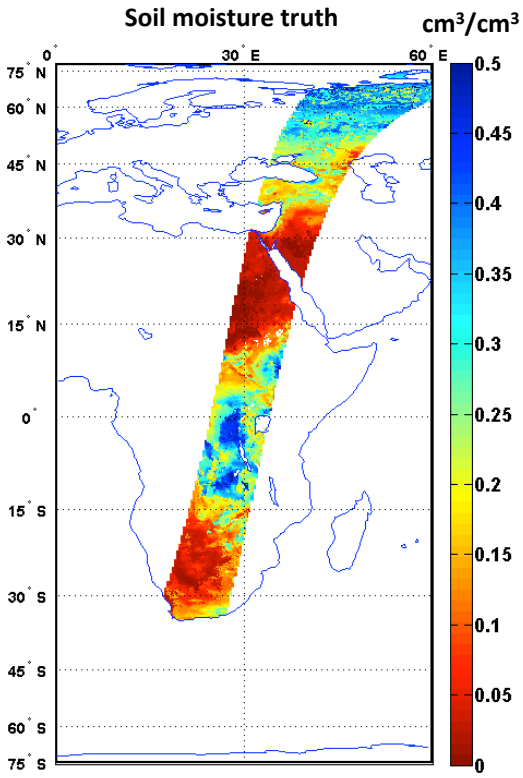


***Figure 24**: True soil moisture at 9 km grid from GloSim for one day in May.*

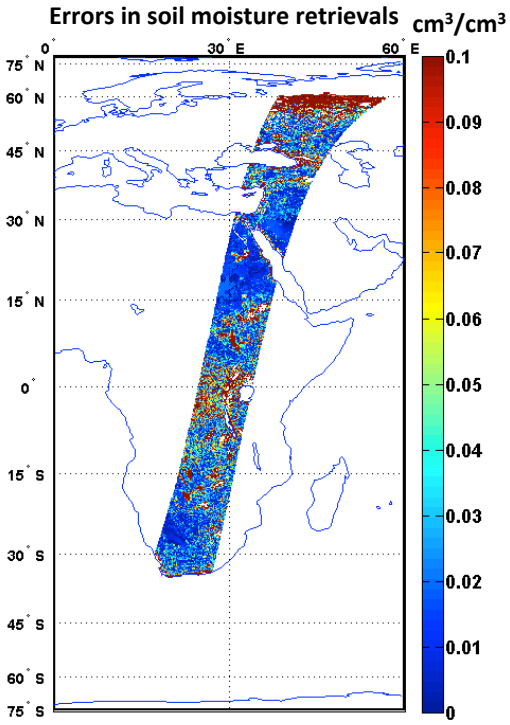


***Figure 25**: Errors in soil moisture at 9 km grid from GloSim for one day in May.*

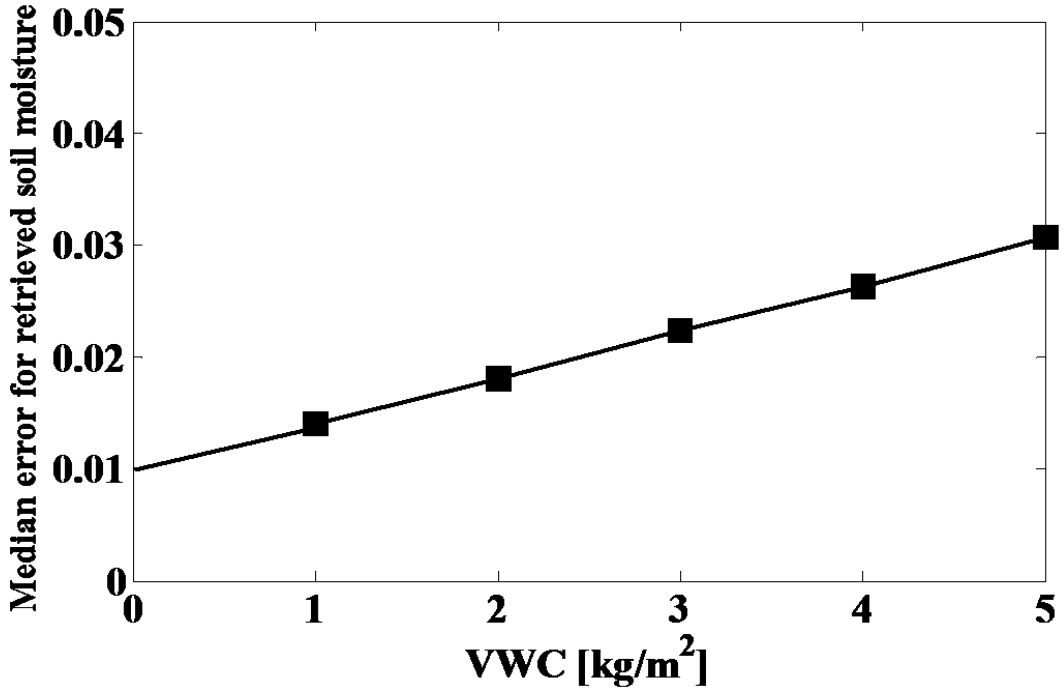


***Figure 26**: Errors in soil moisture against VWC at 9 km grid from GloSim for one day in June.*

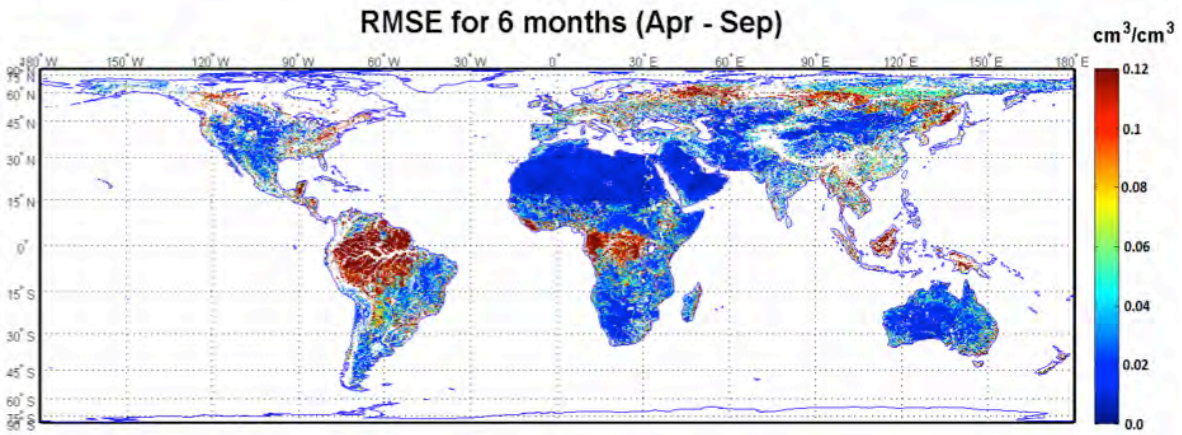


***Figure 27**: Root-Mean-Square-Errors (RMSE) in soil moisture estimates at 9 km grid from GloSim for six months period (Apr-Sep).*

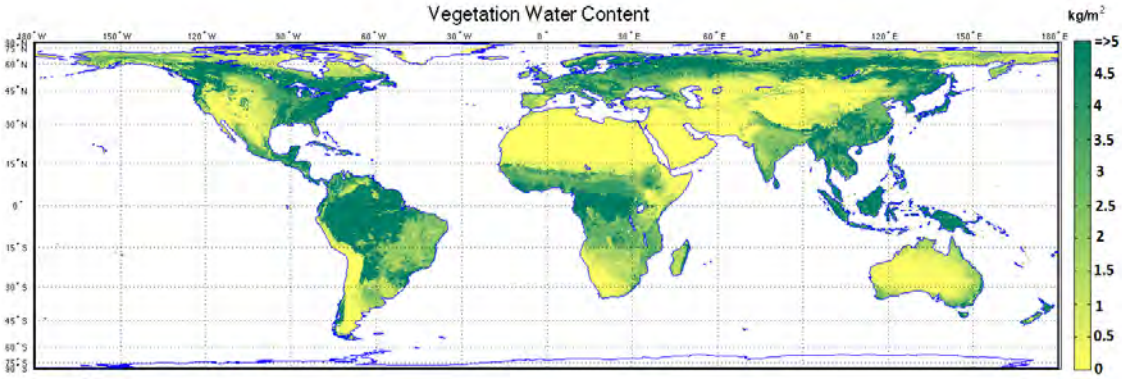


***Figure 28**: VWC at global extent at 9 km on one day in June.*

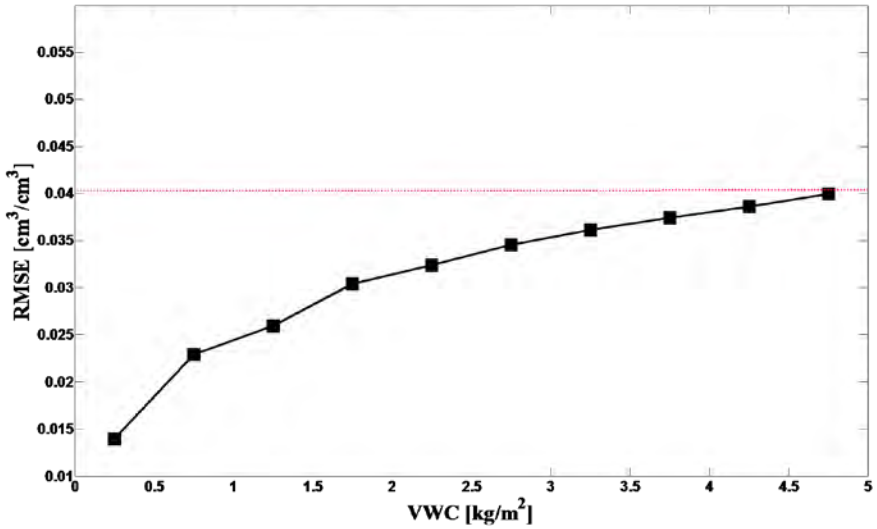


***Figure 29a**: Root-Mean-Square-Errors (RMSE) stratified by the mean VWC contained within 9 km from GloSim for six months period (Apr-Sep).*

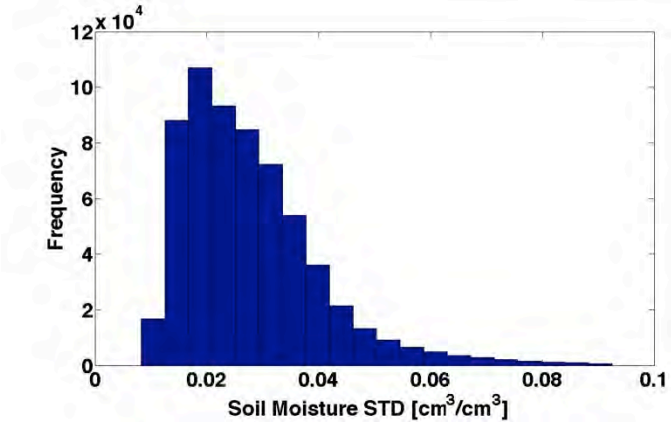


***Figure 29b**: Histogram of standard deviation (RMSE) in retrieved soil moisture within 9 km from GloSim for one-month period (June).*

4.2.2 **Monte-Carlo Simulation**

A Monte-Carlo error analysis is performed that provides further assessment of other important factors beyond the tests performed in Section 4.2. The analysis in this section focuses on two prominent issues that potentially influence the performance of L2_SM_AP baseline algorithm of the SMAP Mission:

1) Potential of cross-pol to characterize landcover spatial heterogeneity within a radiometer pixel, and

2) Sensitivity of algorithm to radar calibration and contamination uncertainties, noise in T_B and errors in ancillary data

To perform a conservative assessment of the L2_SM_AP baseline algorithm that addresses the above concerns, a dynamic heterogeneous zone with variable VWC from within the CONUS domain is extracted as shown in Figure 30.

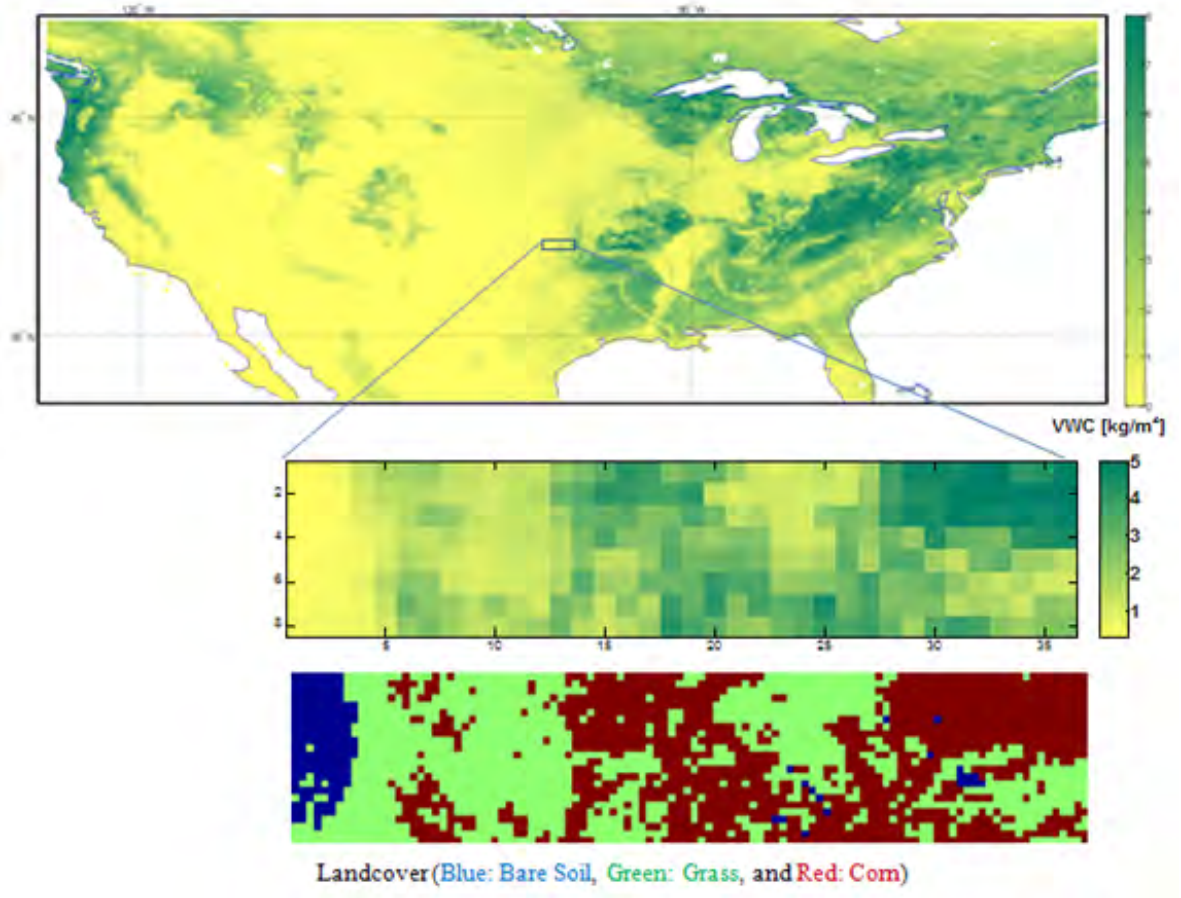


Figure 30*: Study region selected from the CONUS domain.*

The landcover of the study domain (Fig. 30) is assigned as follows: Bare soil when VWC is 0 – 0.5 kg/m^2, Grass when VWC is 0.5 – 1.5 kg/m^2 and Corn when VWC is 1.5 – 5 kg/m^2. This method of assigning landcover artificially is to create a landscape with even greater heterogeneity in

VWC (Fig. 30) to evaluate the performance of the L2_SM_AP baseline algorithm in a spatially diverse region. The choice of landcover is governed by the availability of lookup tables of radar measurements for grass and corn cover types. Forward models are used to calculate true brightness temperature (T_B) and backscatter (σ) from geophysical truths (e.g, soil moisture, and soil temperature) of the study domain. The geophysical truths are from a land surface model operated by Land Information System (LIS) at NASA GSFC. The tau-omega model is used in forward modeling of T_B and forward modeling of σ (co- and cross-polarization) is through the use of lookup tables provided by Dr. Leung Tsang (University of Washington).

A Monte Carlo study is conducted to evaluate the contribution of radar cross-pol (σ_{pq}) data to detect spatial heterogeneity, and how radar cross-pol (σ_{pq}) data improve the performance of the L2_SM_AP baseline algorithm. The results shown in Fig. 31 are from error-free (no noise in T_B, σ, and ancillary information) data when σ_{pq} is used in (17) as compared to the performance of no σ_{pq} as in (22). Significant improvement in soil moisture RMSE (i.e., reduction of 0.02 cm^3/cm^3 in RMSE) at M scale across the VWC range is observed by accounting for the heterogeneity in vegetation through σ_{pq} that adjusts σ_{pp} to reflect mostly the spatial fluctuations due to soil moisture within C.

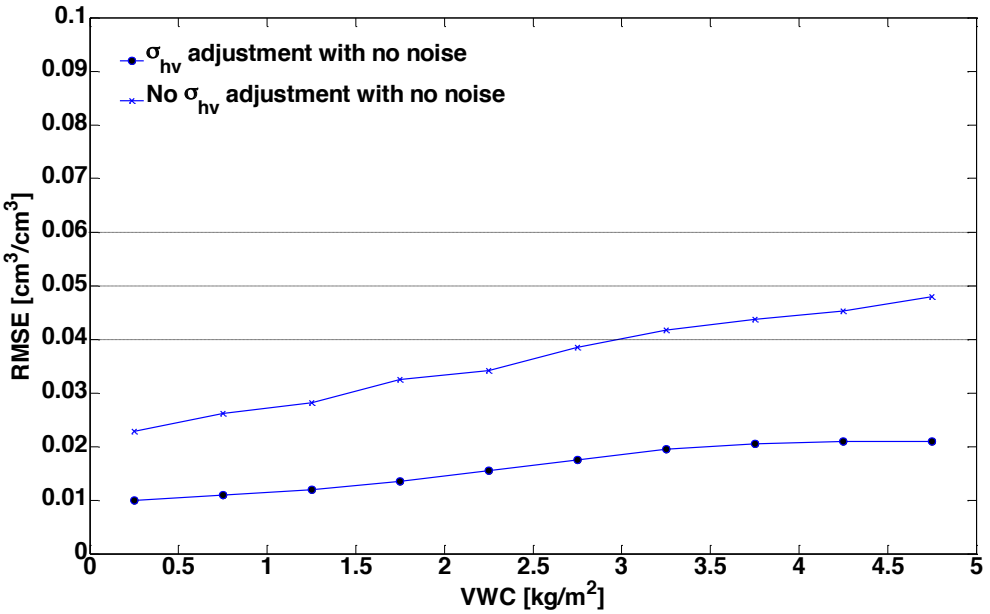


Figure 31: *Impact of σ_{pq} on the L2_SM_AP baseline algorithm.*

Another Monte Carlo simulation is conducted to assess the sensitivity of the algorithm to radar calibration and contamination uncertainties, noise in T_B, and errors in ancillary data. To obtain a conservative assessment, radar data with higher noise floor (σ^{NE} dB) level are used. The parameters of interest that affect the active-passive algorithm are σ^{NE}dB and number of looks (N). In the study, σ^{NE} and N are -33.5/-28.5 and 60/200 (include fore+aft) at inner/outer edge of the swath, respectively. Based on the $\sigma^{NE}$and N, three scenarios are used, with σ^{NE} equal to -33.5/-30.5/-28.5 and corresponding N equal to 60/130/200 that represent the inner-edge, middle and outer-edge of the swath. Noise due to calibration and contamination is also used, with calibration

error for co-pol in linear units of 0.0593, calibration error for cross-pol in linear units of 0.211, and contamination error for co- and cross-pol in linear units of 0.0682. These values are selected in consultation with the radar instrument team. The calibration and contamination noise is added at spatial scale of *M*. Simulations are conducted for all three scenarios (i.e., for different σ^{NE}). In the active-passive algorithm $\sigma(M)$ is obtained by spatially aggregating σ at 3 km. This step reduces the speckle noise (KpC) by a factor of 3 (i.e, $\sqrt{n}$). The aggregation of σ renders almost similar noise levels at *M* for all three above-mentioned scenarios. Since KpC at *M* benefits from increasing N with the degradation of σ^{NE}. Uncertainties in T_B and other parameters are introduced as shown in Table 4.

Table 4: *Uncertainty introduced in parameters for soil moisture retrieval .*

Parameter	Uncertainty associated (1 std.)
T_B	1.5 K
Vegetation Opacity (τ)	10%
Soil Temprature	2 K
Albedo (ω)	5%
Roughness (h)	10%
Sand fraction (sf)	10%
Clay fraction (cf)	10%

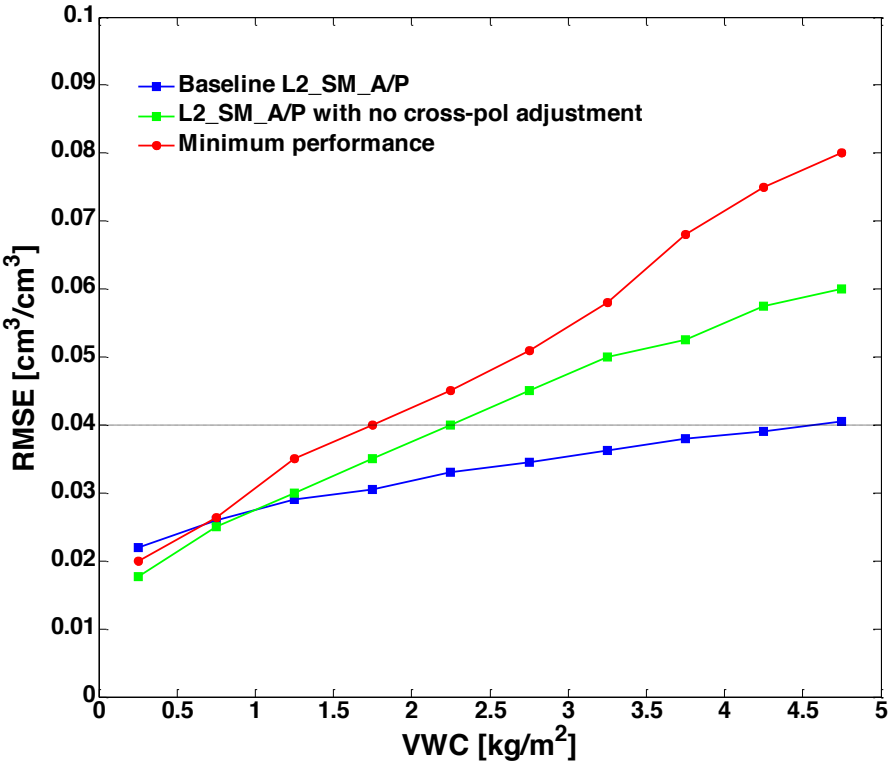


Figure 32*: Performance of the baseline algorithm from Monte Carlo simulation.*

Figure 32 shows the resulting RMSE values in the active-passive algorithm that incorporates noise in T_B, σ, and ancillary information. Results highlight that incorporating the cross-pol radar signature to deal with vegetation and surface roughness heterogeneity significantly improves the performance of the algorithm. The above results are based on a small region extracted from the CONUS domain, therefore a caveat is that the RMSE curve may be subject to change when the baseline algorithm is implemented to a larger spatial domain. However, we get a very favorable result from the GloSim L2_SM_AP outputs (Section 4.2.1) that include most of the expected geophysical conditions encountered during soil moisture retrievals.

4.3 Error Budget of Baseline Algorithm

The baseline algorithm (Section 3.2) error budget is estimated using the basic input uncertainty and parameter perturbations in a sensitivity study. The input data are the water body corrected brightness temperature and the radar backscatter cross-sections that are averaged to 9 km. The radar pixels that include water bodies are excluded as well.

Table 5 lists the various contributions to the disaggregated brightness temperature at 9 km resulting from the baseline algorithm. The first numbered row is the estimated error in the L1C_TB (36 km EASE grid) which is due to the instrument, geophysical contributions to Earth surface brightness temperature 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 SRTM. Dynamic water bodies are detected using the SMAP radar backscatter cross-sections (see L2_SM_A ATBD).

The water-body adjusted brightness temperature root-sum-of-squares (RSS) is reported in row three of Table 5. The baseline algorithm uses the radar backscatter cross-section and brightness temperature time-series to estimate a disaggregated 9 km brightness temperature. The contribution of radar backscatter cross-section calibration and contamination noise is estimated using the Monte Carlo method described in Section 4.2.2. The radar noise and uncertainty contributions to the disaggregated 9 km brightness temperature is estimated to be 1.65 K based on the baseline algorithm models. This uncertainty is shown in row four of Table 4 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 9 km brightness temperature, and the estimated value is 1.60 K (shown in row 5 of Table 4). The total 9 km disaggregated brightness temperature error of 2.73 K is shown as an RSS in the sixth row of Table 5.

Table 5: *Error budget in degrees Kelvin*

	Error Sources at 36 km Ease 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	0.70 K
3	**Adjusted Corrected T_B RSS**	**1.47 K**
4	Radar calibration and contamination error	1.65 K
5	Algorithm Parameter error	1.60 K
6	**Disaggregated T_B (9 km) estimation RSS**	**2.73 K**

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

Table 6 represents the same error budget but with more detail and in units of percent volumetric soil moisture cm^3/cm^3. As evident in row three, the correction for inland water bodies is critical because it can be a large contributor to the error budget. The emissivity of water and land are very different and any misspecification leads to large errors. Together with the radar backscatter uncertainty contributions in the baseline algorithm, the row one to four RSS total error of 9 km brightness temperature is shown in row six of Table 5. Tables 5 and 6 are different from this point onwards. Table 6 uses the brightness temperature single channel surface soil moisture retrieval algorithm to estimate the uncertainty contribution of ancillary data and retrieval model in percent volumetric soil moisture cm^3/cm^3. Also five levels of Vegetation Water Content (VWC) are used because vegetation opacity is considered to be a major factor in the retrieval accuracy. The errors due to 2.0 K land surface temperature, 10% uncertainty in 9 km VWC, 5% error in dielectric model percent sand and clay specification, and 5% error on major model parameters are shown in rows seven through ten of Table 6. The total retrieval uncertainty is shown in the last row of Table 6 for five vegetation classes. The nominal values used are indicated through underscore.

Table 6*: Error budget in volumetric soil moisture cm^3/cm^3*

	Error Sources	Estimated Error					Nominal Value
1	Radiometer precision and calibration stability, faraday rotation, atmospheric gases, non-precipitating clouds, and gridding	1.3 K					1.3 K
2	Waterbody fraction		5%				
			***0.70* K**				0.70 K
3	**Adjusted Corrected T_B RSS**		***1.47* K**				1.47 K
4	Radar S0(*pp*) and S0 (*pq*) errors*	***1.65* K**					1.65 K
5	Algorithm Parameter error	***1.60* K**					1.60 K
6	**Disaggregated T_B (9 km) RSS**	***2.73* K**					2.73 K
7	VWC** with 10% error	0-1 kg/m2	1-2 kg/m2	2-3 kg/m2	3-4 kg/m2	4-5 kg/m2	
		0.003 cm^3/cm^3	0.01 cm^3/cm^3	0.015 cm^3/cm^3	0.020 cm^3/cm^3	***0.025*** cm^3/cm^3	0.025 cm3/cm3
8	Soil temperature (2 K)	0.010 cm^3/cm^3	0.013 cm^3/cm^3	0.015 cm^3/cm^3	0.020 cm^3/cm^3	***0.025*** cm^3/cm^3	0.025 cm^3/cm^3
9	Soil texture (5% error in sand & clay fraction)	0.002 cm^3/cm^3	0.003 cm^3/cm^3	0.003 cm^3/cm^3	0.004 cm^3/cm^3	***0.010*** cm^3/cm^3	0.01 cm^3/cm^3
10	Parameters (h, ω, and b) 5% error each	0.003 cm^3/cm^3	0.004 cm^3/cm^3	0.004 cm^3/cm^3	0.006 cm^3/cm^3	***0.010*** cm^3/cm^3	0.010 cm^3/cm^3
11	**Soil moisture retrieval at 9 km**	**0.011** cm^3/cm^3	**0.017** cm^3/cm^3	**0.022** cm^3/cm^3	**0.029** cm^3/cm^3	***0.0375*** cm^3/cm^3	**0.0375** cm^3/cm^3

* Calibration and contamination errors
**Vegetation Water Content
Underlined values are used as nominal

The nominal value is based on the limit of 5 kg m^{-2} specified for SMAP Level 1 science requirements. As shown in Table 6 the baseline L2_SM_AP can meet the 0.04 cm^3/cm^3 requirement even under this vegetation cover. The above error budget table is developed based on Monte Carlo analysis of nominal set of conditions, e.g., mean VWC level, waterbody fraction, soil texture, soil moisture, etc. The error and uncertainty depend on these conditions and hence do not apply to every SMAP L2_SM_AP granule. In the next section we develop an analytical uncertainty analysis formulation that is based on every granule conditions. This uncertainty estimate accompanies everyL2_SM_AP retrievals in the science product data files.

4.4 Analytical Formulation for Disaggregated TB Error Estimates

SMAP L-band radiometer brightness temperature $T_{B_p}(C)$ at 36 km spatial resolution on EASE2 grid observations are subjected to instrument related noise and errors due to geophysical factors, processing and gridding. Similarly, SMAP L-band radar backscatters are subjected to instrument noise, SAR processing and gridded to 3 km spatial resolution on EASE2 grid. SAR backscatter comprises random/speckle noise, calibration and contamination noise. For the SMAP Active-Passive algorithm $\sigma_{pp}(M_j)$ and $\sigma_{pq}(M_j)$ at 9 km are obtained by spatially aggregating SAR

backscatter at 3 km. The parameters of interest that affect the SMAP Active-Passive algorithm are noise-equivalent (σ^{NE}[dB]) and number of looks (N) with respect to the swath position. For SMAP, σ^{NE} and N are -33.5/-28.5 and 60/200 (include fore+aft) at inner/outer edge of the swath, respectively. The speckle noise (Kpc) can be calculated using the information of σ^{NE} and N that leads to a varying value of KpC across the swath. Aggregation/averaging of σ_{pp} from 3 km to 9 km (M scale) reduces the speckle noise (KpC) by a factor of 3 (i.e, $\sqrt{n}$). However, uncertainties due to calibration and contamination present in the SAR backscatters are not reduced by averaging from 3 km to 9 km (M scale) because they are biases.

Contribution of errors due to SMAP Radiometer Brightness Temperature and SAR backscatters in Active-Passive Algorithm:

The errors in SMAP gridded 36 km brightness temperature $T_{B_p}(C)$ mostly come from radiometer precision and calibration stability, faraday rotation, atmospheric gases, non-precipitating clouds, and gridding. The SMAP $T_{B_p}(C)$ is expected to meet the requirements of 1.3 [K] for $T_{B_p}(C)$ for all portion of the 1000 km wide swath irrespective of the nadir region. This error source of 1.3 [K] is ingested and is an additive term in (17). Therefore when other errors are account for in Eq. 6, the resultant disaggregated brightness temperature $T_{B_p}(M_j)$ is always expected to have errors greater than 1.3 [K]. In case of SAR backscatters $\sigma_{pp}(M_j)$ and $\sigma_{pq}(M_j)$ ingestion in (17), the error isolation and accounting is not straight forward as $T_{B_p}(C)$. An analytical approach is formulated to quantify the error in $T_{B_p}(M_j)$ due to SMAP observations. The formulation is as follows:

Consider taking variance of (17), keeping all other parameters as constant except SMAP instrument observations.

$$var[T_{B_p}(M_j)] = var[T_{B_p}(C)] + \beta(C)^2\{var[\sigma_{pp}(M_j) - \sigma_{pp}(C)] + \Gamma^2.var[[\sigma_{pq}(C) - \sigma_{pq}(M_j)]]\} \quad (23)$$

Equation 23 can be written as,

$$var[T_{B_p}(M_j)] = var[T_{B_p}(C)] + \beta(C)^2\{var[\sigma_{pp}(M_j) - \langle\sigma_{pp}(M_j)\rangle] + \Gamma^2.var[\sigma_{pq}(M_j) - \langle\sigma_{pq}(M_j)\rangle]\} \quad (24)$$

where $\langle\sigma_{pq}(M_j)\rangle = \sigma_{pp}(C)$, simplifying further, substituting $var[\langle\sigma_{pp}(M_j)\rangle] = 0$, and since σ_{pp} is [dB] in (17), substituting in linear units as $10.1\log_{10}(\sigma_{pp}.(1+\omega.KpC))$, where KpC is for the speckle noise and ω is noise with zero mean and 1 std. For sake of simplicity σ_{pp}, σ_{pq} and β are mentioned instead of $\sigma_{pq}(M_j)$ and $\sigma_{pq}(M_j)$ and $\beta(C)$ from hereon, respectively.

$$var[T_{B_p}(M_j)] = var[T_{B_p}(C)] + \beta^2\{var[10.\log_{10}(\sigma_{pp}.(1+\omega.KpC_{pp}))] + \Gamma^2.\ var[10.log_{10}(\sigma_{pq}.(1+\omega.KpC_{pq}))]\} \quad (25)$$

Solving Eq. 25 further, leads to

$$var[T_{B_p}(M_j)] = var[T_{B_p}(C)] + \beta^2\{(10/\log_{10})^2\ .\ KpC_{pp}{}^2 + \Gamma^2.\ (10/\log_{10})^2\ .\ KpC_{pq}{}^2\}$$

$$var[T_{B_p}(M_j)] = var[T_{B_p}(C)] + \beta^2.\ (10/\log_{10})^2\ \{\ KpC_{pp}{}^2 + \Gamma^2.\ KpC_{pq}{}^2\} \quad (26)$$

Solving term $var[10.\log_{10}(\sigma_{pp}.(1+\omega.KpC_{pp}))]$ to $\{(10/\log_{10})^2.\ KpC_{pp}{}^2\}$ is elaborated in Appendix A1. Equation 26 is the analytical solution of the variance of (17) and quantifies contribution of errors in disaggregated brightness temperature $T_{B_p}(M_j)$ due to noise and errors in the SMAP observations.

Algorithm Parameters Uncertainties:

In the previous subsection, we dealt with errors in SMAP observations. However, parameters in (6) also have errors associated with them, because estimating them through observations is also not free of uncertainties. There are two basic sources for such uncertainties in determination of algorithm parameters (β, Γ), they are: i) errors in T_{B_p} and speckles in σ_{pp} and σ_{pq}; and ii) standard error due to linear regression. Consider in (17), keeping the observations error free and introducing uncertainties in parameters, Eq. 17 can be expressed as:

$T_{B_p}(M_j) = T_{B_p}(C) + \beta\{\delta_{pp} + \delta_{pq} \,.\, \Gamma\}$ (27)

$\delta_{pp} = [\sigma_{pp}(M_j) - \langle\sigma_{pp}(M_j)\rangle]$, deviation of $\sigma_{pp}(M_j)$ from mean i.e., $\langle\sigma_{pp}(M_j)$

$\delta_{pq} = [\sigma_{pq}(M_j) - \langle\sigma_{pq}(M_j)\rangle]$, deviation of $\sigma_{pq}(M_j)$ from mean i.e., $\langle\sigma_{pq}(M_j)$

Introducing errors/uncertainties in parameters

$T_{B_p}(M_j) = T_{B_p}(C) + (\bar{\beta} + \beta').\{\delta_{pp} + \delta_{pq} \,.\, (\bar{\Gamma} + \Gamma')\}$ (28)

Expanding (28),

$T_{B_p}(M_j) = T_{B_p}(C) + \bar{\beta}\,\delta_{pp} + \bar{\beta}\,\bar{\Gamma}\delta_{pq} + \bar{\beta}\,\Gamma'\delta_{pq} + \beta'\delta_{pp} + \beta'\,\bar{\Gamma}\delta_{pq} + \beta'\Gamma'\delta_{pq}$ (29)

Taking variance of (29),

$var[T_{B_p}(M_j)] = var[T_{B_p}(C)] + var[\bar{\beta}\,\delta_{pp}] + \text{var}[\bar{\beta}\,\bar{\Gamma}\delta_{pq}] + \text{var}[\bar{\beta}\,\Gamma'\delta_{pq}] + \text{var}[\beta'\delta_{pp}] + \text{var}[\beta'\,\bar{\Gamma}\delta_{pq}] + \ldots$
$\text{var}[\beta'\Gamma'\delta_{pq}]$ (30)

Many terms that are means in (30) are nullified because their variance is zero,

$var[T_{B_p}(C)] \approx 0$, $var[\bar{\beta}\,\delta_{pp}] \approx 0$, and $var[\bar{\beta}\,\bar{\Gamma}\delta_{pq}] \approx 0$

Another term in Eq. 30, $\text{var}[\beta'\Gamma'\delta_{pq}] \approx 0$, since β' and Γ' are assumed uncorrelated, therefore their covariance is zero. Based on these conditions, Eq. 30 can be expressed as,

$var[T_{B_p}(M_j)] = \text{var}[\bar{\beta}\,\Gamma'\delta_{pq}] + \text{var}[\beta'\delta_{pp}] + \text{var}[\beta'\,\bar{\Gamma}\delta_{pq}]$ (31)

$var[T_{B_p}(M_j)] = \bar{\beta}^2\delta_{pq}{}^2\,\text{var}[\Gamma'] + \delta_{pp}{}^2\,\text{var}[\beta'] + \bar{\Gamma}^2\,\delta_{pq}{}^2\,\text{var}[\beta']$ (32)

$var[T_{B_p}(M_j)] = \delta_{pp}{}^2\,\text{var}[\beta'] + \delta_{pq}{}^2\,\{\bar{\beta}^2\,\text{var}[\Gamma'] + \bar{\Gamma}^2\,\text{var}[\beta']\}$ (33)

Equation 18 describes the expected variance in $T_{B_p}(M_j)$ due to uncertainties in parameters β and Γ. The terms $\text{var}[\beta']$ and $\text{var}[\Gamma']$ in (33) can be analytically estimated individually because β and Γ are regression coefficients and their behavior are mostly determined by the time series of T_{B_p} and $\sigma_{pp}(M_j)$, and spatial evolution of $\sigma_{pp}(M_j)$ and $\sigma_{pq}(M_j)$, respectively.

Consider the regression (34),

$\hat{y} = \alpha + \beta\,\hat{x} + \varepsilon$ (34)

where for β estimate, $y = T_{B_p}(C)$, and $x = \sigma_{pp}(C)$ with an estimation time series window length of n. if $\hat{y}$ and $\hat{x}$ also include independent instruments noise, then

$\hat{y} = y + \omega_y$, and $\hat{x} = x + \omega_x$ (35)

From (34) and (35),

$y + \omega_y = \alpha + \beta x + \beta\omega_x + \varepsilon$ (36)

Taking average over time series window n and instruments, we get

$\bar{y} = \alpha + \beta\,\bar{x}$ (37)

Subtracting (37) from (36),

$y - \bar{y} + \omega_y = \beta(x - \bar{x}) + \beta\omega_x + \varepsilon$ (38)

Solving for ε,

$\varepsilon = (y - \bar{y}) - \beta(x - \bar{x}) - \beta\omega_x + \omega_y$ (39)

Squaring (39) and taking expectation

$E[\varepsilon^2] = E[(y - \bar{y})^2] + \beta^2 E[(x - \bar{x})^2] - \beta^2 E[(y - \bar{y})(x - \bar{x})] + \beta^2 E[\omega_x{}^2] + E[\omega_y{}^2]$ (40)

Note that the instruments noise ω_y and ω_x are independent, however, $(y - \bar{y})$ and $(x - \bar{x})$ are dependent.

$E[(y - \bar{y})^2] = S_y^2$ is the true physical variability of $T_{B_p}(C)$ over the estimation window n. Similarly, $E[(x - \bar{x})^2] = S_x^2$ is the true physical variability of $\sigma_{pp}(C)$ over the estimation window n. The term $E[(y - \bar{y})(x - \bar{x})]$ in (40) is the covariance between y and x. Therefore

$E[(y - \bar{y})(x - \bar{x})] = cov[x, y] = r(x, y)[S_y S_x]$ (41)

Substituting the above equations in Eq. 40, we

$E[\varepsilon^2] = var[\varepsilon] = S_y^2 + \beta^2 S_x^2 - \beta^2 r(x, y)[S_y S_x] + \beta^2 var[\omega_x] + var[\omega_y]$ (42)

where $var[\beta] = \frac{var[\varepsilon]}{S_x^2(n-1)}$, hence

$$var[\beta] = \frac{1}{S_x^2(n-1)}\{S_y^2 + \beta^2 S_x^2 - \beta^2 r(x,y)[S_y S_x] + \beta^2 var[\omega_x] + var[\omega_y]\} \quad (43)$$

Where S_y^2 is variance of $T_{B_p}(C)$ over estimation window n, S_x^2 is variance of $\sigma_{pp}(C)$ over estimation window n, $r(x,y)$ is the correlation between $T_{B_p}(C)$ and $\sigma_{pp}(C)$ over n, β is the nominal slope obtained through regression, $var[\omega_y]$ is the variance of radiometer measurement noise, and $var[\omega_x]$ is the variance in radar co-pol measurement noise.

Similarly $var[\Gamma]$ can be obtained using the approach mentioned above

$$var[\Gamma] = \frac{1}{S_x^2(n-1)}\{S_y^2 + \beta^2 S_x^2 - \beta^2 r(x,y)[S_y S_x] + \beta^2 var[\omega_x] + var[\omega_y]\} \quad (44)$$

In (44), Where S_y^2 is spatial variance of $\sigma_{pp}(M)$ within the spatial scale C, S_x^2 is spatial variance of $\sigma_{pq}(M)$ within the spatial scale C, $r(x,y)$ is the correlation between $\sigma_{pp}(M)$ and $\sigma_{pq}(M)$ within the spatial scale C, Γ is the nominal slope obtained through regression, $var[\omega_y]$ is the variance of radar co-pol measurement noise, and $var[\omega_x]$ is the variance in radar x-pol measurement noise. Incorporating (43) and (44) in (33) lead to estimation of variance in disaggregated $T_{B_p}(M)$ due to uncertainties in algorithm parameters.

Uncertainty in Brightness Temperature due to error in Water Bodies Estimates:

The SMAP radiometer 3 dB elliptical footprint size is nearly ~40 km (36.5 km as minor axis and 47.2 km as major axis). In many parts of the world, it is possible to have fractional portion in form of waterbodies within such a big footprint size area. The brightness temperature of water and land are contrasting different, at least ~80 [K] to ~100 [K] greater at *h* and *v* polarization (L-band) observations over land. Hence, presence of water within the SMAP footprint greatly impacts and lowers the brightness temperature. For soil moisture retrieval, brightness temperature from land portion is desirable. Therefore the SMAP radiometer observed brightness temperature is subjected to corrections for the affects of waterbodies. If the fractional coverage of waterbodies is precisely known then it is fairly straightforward to compute the land portion of brightness temperature by (45).

$$T_{B_L} = \frac{T_{B_R} - f T_{B_W}}{1-f} \quad (45)$$

Where T_{B_L} is brightness temperature over land, T_{B_R} is brightness temperature observation of SMAP radiometer, T_{B_W} is the brightness temperature over water, and *f* is the fraction water cover within the SMAP 3 dB footprint. If *f* is zero, then $T_{B_L} = T_{B_R}$, and there is no error in T_{B_L} due to water body, however, if *f* is greater than zero and associated with some uncertainty, it is obvious that estimation of T_{B_L} is not perfect and some error is introduced. It is the case most of the time in real time correction of brightness temperature. Therefore it is essential to quantify error in T_{B_L} due to correction done using (45).

Expanding (45),

$$T_{B_L} = \left\{\frac{1}{1-\bar{f}} + \frac{(f-\bar{f})}{(1-\bar{f})^2}\right\}\{T_{B_R} - f T_{B_W}\} \quad (46)$$

$\frac{1}{1-\bar{f}} = \left\{\frac{1}{1-\bar{f}} + \frac{(f-\bar{f})}{(1-\bar{f})^2}\right\}$, obtained through Taylor series expansion around $\bar{f}$ (actual water fraction). This Taylor series expansion is mentioned in Appendix (A.2). Substituting $f = (\bar{f} + \acute{f})$ in (46), where $\acute{f}$ is error in waterbodies fraction within the radiometer footprint.

$$T_{B_L} = \left\{\frac{1}{1-\bar{f}} + \frac{\acute{f}}{(1-\bar{f})^2}\right\}\{T_{B_R} - (\bar{f} + \acute{f}) T_{B_W}\} \quad (47)$$

Expanding and simplifying Eq. 47, leads to (details provided in A.3),

$$T_{B_L} = \overline{T_{B_L}} - \frac{\acute{f}.T_{B_W}}{(1-\bar{f})} + \frac{\acute{f}.T_{B_R}}{(1-\bar{f})^2} - \frac{\bar{f}\acute{f}.T_{B_W}}{(1-\bar{f})^2} - \frac{\acute{f}^2 T_{B_W}}{(1-\bar{f})^2} \quad (48)$$

By further simplification, we get,

$$\left(T_{B_L} - \overline{T_{B_L}}\right) = \frac{-1}{(1-\bar{f})^2}\{T_{B_W}(\acute{f} - \acute{f}^2) - \acute{f} T_{B_R}\} \quad (49)$$

Squaring both side of (49) and taking expectation,

$$\left(T_{B_L} - \overline{T_{B_L}}\right)^2 = \frac{1}{(1-\bar{f})^4}\left\{{T_{B_W}}^2\left(\rho_f^2 + 3\rho_f^4\right) - 2T_{B_W}T_{B_R}\rho_f^2 + \rho_f^2\,T_{B_R}\right\} = MSE_f \quad (50)$$

Expectation of $\overline{\acute{f}} = 0$, $\overline{\acute{f}^2} = \rho_f^2$, $\overline{\acute{f}^3} = 0$, and $\overline{\acute{f}^4} = 3\rho_f^4$. MSE_f is the Mean-Square-Error in land only brightness temperature (T_{B_L}), and ρ_f is standard deviation in *f*. Solving (50) further get the final form (51) that is the analytical solution of getting errors in T_{B_L}, due to uncertainty in waterbody fraction *f*.

$$MSE_f = \frac{\rho_f^2}{(1-\bar{f})^4}\left\{3\,{T_{B_W}}^2\,\rho_f^2 + \left(T_{B_W} - T_{B_R}\right)\right\} \quad (51)$$

Analysis of Error Characteristics in the Disaggregated Brightness Temperature:

Above, we derived the analytical solutions to quantify errors in disaggregated T_B due to uncertainties in SMAP observations and the parameters used in (17). Sensitivity analysis using above derived equations (26, 33, 43, 44, 51) are conducted for each error source to verify the analytical solutions. The exercise also provided a path forward for establishing thresholds, determining limitation of algorithm (17), and potential for implementing flags in the SMAP Active-Passive product (L2_SM_AP).

Figure 33 illustrates the errors in disaggregated brightness temperature $T_{B_p}(M_j)$ due to instruments measurements noise. Nominal values of radar measurements noise in co-pol (*KpC* = 0.17) and x-pol (*KpC* = 0.26), and radiometer measurement noise (NEDT of 1.3 K) are used to create Fig. 33, with parameters β = -3.0 [K/dB] and Γ = 0.7 [dB/dB] having no uncertainty associated with them. The analytical solution closely agrees with the Monte Carlo outputs verifying the accuracy of (26). Similar very close concurrence is observed between the analytical solutions and the Monte Carlo outputs for valid ranges of β and Γ value. Another aspect not shown in Fig. 33 is that with the increase in absolute value of β the errors in $T_{B_p}(M_j)$ also increase due to multiplicative nature of radar measurement noise to the β values (17). From Fig. 33, a feature worth discussing is that the errors increase if all the nine 3 km grid cells are not used to create the aggregated 9 km grid cell. SMAP Active-Passive algorithm will come across such condition especially near water bodies, corrupt radar grid cells due to RFI, frozen ground and presence of snow and ice where 3 km radar grid cells over water are ignored during aggregation. Averaging from 3 km to 9 km spatial scale reduces the radar speckle noise (*KpC*) by a factor ($1/\sqrt{n}$), consequently decreasing the error contributions. Therefore when all the 3 km radar grid cells are available and then aggregated to 9km grid cell, significant reduction in noise of radar measurement takes place and subsequent reduction of error in disaggregated brightness temperature $T_{B_p}(M_j)$. Figure 33 also highlights the expected high errors, if the SMAP Active-Passive algorithm is implemented at 3 km (i.e., *n* = 1), where no reduction of radar speckle/noise takes place because of averaging.

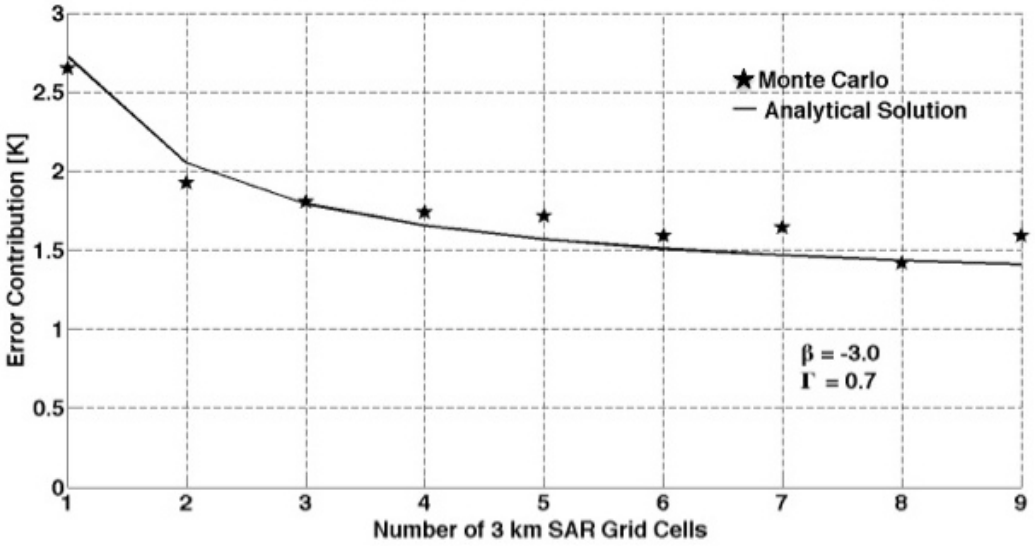


Figure 33: Contribution of error in the disaggregated brightness temperature $T_{B_p}(M_j)$ due to SMAP radiometer NEDT and radar co-pol and x-pol speckle noise.

Other major sources of error in $T_{B_p}(M_j)$ are imparted through the uncertainties of the algorithm parameters β and Γ. The parameters β and Γ vary in space and time due change in vegetation characteristics and landcover. Uncertainty especially, in parameter β may arises for following reasons: a) if the surface soil moisture condition does not change enough within the temporal window that leads to low dynamic range in brightness temperature and radar backscatter, hence high standard error in regression estimate, b) suboptimal temporal window size for estimating regression coefficient i.e., β, and c) presence of heterogeneous landcover conditions within the spatial scale (C) of estimation. Leroux et al., 2014 conducted a study that used airborne remote sensing data from Passive and Active L-band System (PALS) to implement the SMAP Active-Passive algorithm. They found that β estimate improve considerably if temporal window of estimation is optimized hence reducing uncertainty in β estimate. Based on the above discussion, it is obvious that extent of uncertainty in β may come from variety of reasons. However, in the SMAP mission, the β parameter will be unique for every 36 km grid cell (C), and is expected to improve with time over many fairly stable landcovers (barren areas, deserts, rain forests, and shrublands) because of large temporal window of estimation that helps obtain good dynamic range of radiometer brightness temperature and radar backscatter. For other landcovers that show seasonality i.e., varying condition of vegetation (for example agricultural dominated region), SMAP data from same season from different years will help reduce uncertainty in the β parameter. However, there will be always some level of uncertainty associated with the β parameter that warrants sensitivity analysis to fully understand the impact and contribution of errors in $T_{B_p}(M_j)$.

In contrast to β parameter, confidence in estimation of Γ parameter is higher because it does not require any temporal window optimization and many data pairs of co-pol and x-pol backscatters at fine resolution (3 km) are available within the C-scale (36 km) imparting statistically significant results. Studies have shown that the Γ parameter mostly ranges between 0.6 and 0.8 with a mode of nearly 0.7. However, to perform sensitivity analysis we injected various level of uncertainty in the β and Γ parameters to evaluate their impact on $T_{B_p}(M_j)$. Figures 34a-34b illustrates error in $T_{B_p}(M_j)$ for range of uncertainties β with 10% and 20% uncertainties in Γ parameters, respectively.

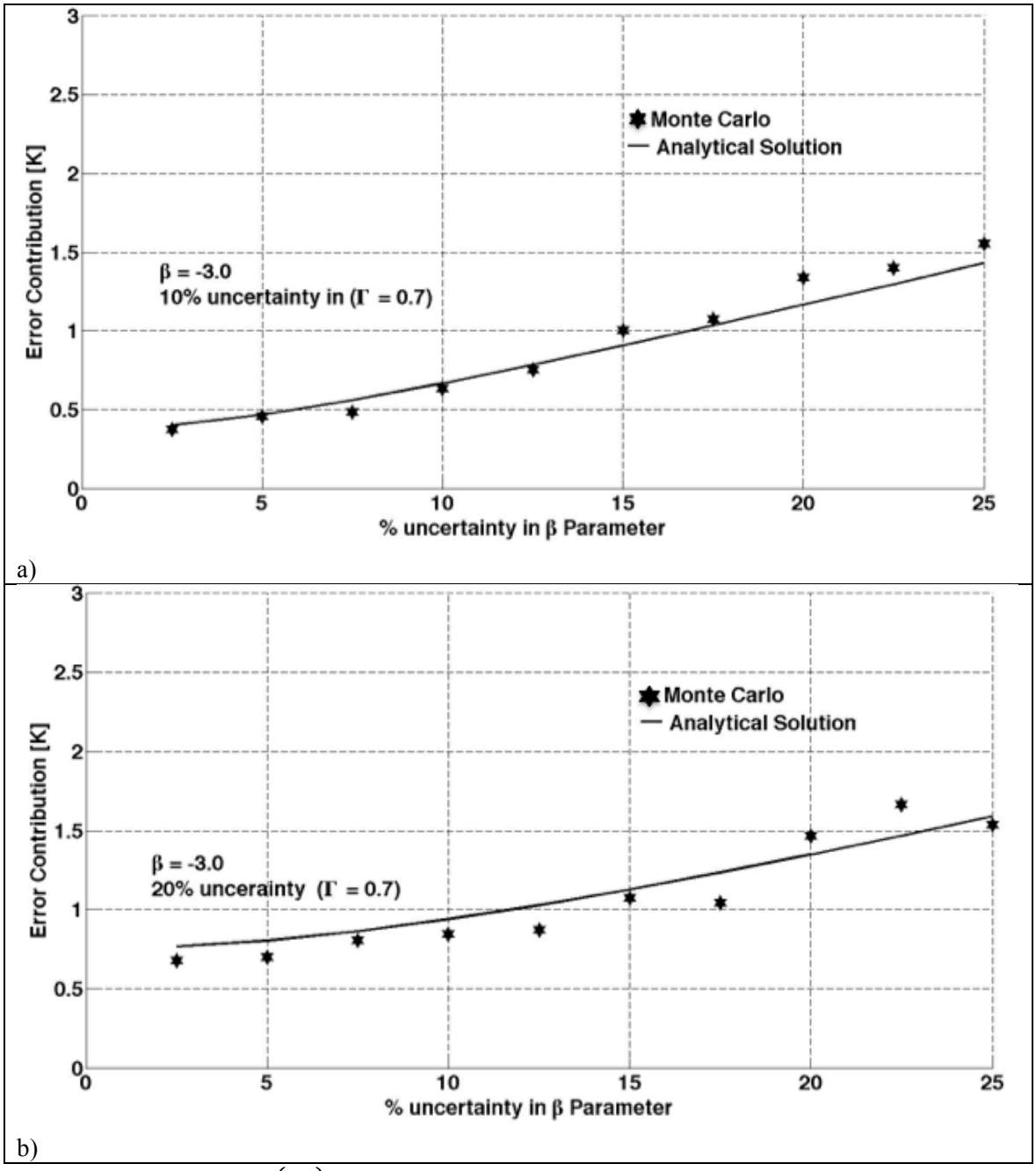


Figure 34: Errors in $T_{B_p}(M_j)$ due to uncertainties in β, a) with 10% error in Γ parameter; and b) with 20% error in Γ parameter.

Figures 34a-34b validate the analytical solution presented in (33) by closely following the Monte Carlo estimates for errors in $T_{B_p}(M_j)$. Two aspects are very obvious from Figures 34a-34b, they are: i) uncertainty in β has greater impact on $T_{B_p}(M_j)$ due to its magnitude and multiplication to deviation of co-pol and x-pol measurements; and ii) the level of uncertainty in Γ parameter impacts errors in $T_{B_p}(M_j)$ mostly when uncertainty in β paremeter is low as obvious in Figs. 34a-34b. Therefore these aspects suggest that is imperative to better understand the nature of uncertainty present in β parameter.

With a nominal values -3.0 and 0.7 for β and Γ parameters, respectively, the errors in $T_{B_p}(M_j)$ is ~1.5 K even for conservative estimate of 25% uncertainty in the parameter β (Figs. 34a-34b). The plots of Fig. 34 also exhibit that less than 15% uncertainty in β and Γ parameters leads less than ~1.0 K error in $T_{B_p}(M_j)$ and will have minimal impact on disaggregation algorithm (Eq. 17). However, with the dynamics of SMAP observations of brightness temperature and backscatters over various landcovers in different seasons, climates and many regions of the world there is enough probability that in many occasions the uncertainty in the algorithm parameters may go above 15% and therefore it is essential to exactly know the level of uncertainty in the parameters and ultimately their contribution towards error in $T_{B_p}(M_j)$.

Analytical solution output for estimating uncertainty in algorithm parameters β is performed using Eq. 43 is shown in Figs. 35a-35b. They illustrate errors in β parameter for two probable conditions. Figure 35a shows the errors in β parameter when temporal standard deviation of 2.5 [K] and 1.5 [dB] in $T_{B_p}(C)$ and $\sigma_{pp}(C)$ are low, respectively, and the number of SMAP overpasses are less (between 10 and 30 resembling ~1 month to ~3 months, respectively) for estimation of β parameter. Such ground conditions may exist over regions like agricultural domain where geophysical variables (for example, size of crop and vegetation water content) change much faster than regions like bare surface, shrublands and dense-forest. It is essential to limit the window size on such regions to capture seasonality and optimize β parameter. The uncertainty is quiet high ~30-35% of β parameter estimate if only 10 overpasses (1 month) are used (Fig. 4a), and leading to ~2.0 [K] error contribution in $T_{B_p}(M_j)$ due to parameters. In contrast, errors in Γ parameter remain quite stable because of snapshot determination approach, and error contribution in $T_{B_p}(M_j)$ due to Γ parameters is very low. A diametrically opposite scenario (Fig. 35b) is observed over bare surface and mildly vegetated landcovers that experience wetting and drying over a period of time with almost no changes on ground conditions. In such scenarios, the standard deviations of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ are relatively high due to range of soil moisture conditions, resulting in robust estimation of β parameter with less uncertainty. Figure 35b is created using standard deviation of 7.0 [K] and 3.0 [dB] in $T_{B_p}(C)$ and $\sigma_{pp}(C)$, respectively, and having high correlation. In this scenario, the magnitude of uncertainty in β parameter is expected to be < 20%. Such regions will have high fidelity in β parameter estimation. However, will require longer time series information in tune of 60 to 120 overpasses (6 months to 1 year time frame) to achieve optimal β parameter that will consequently result in ~1.0 [K] error contribution in $T_{B_p}(M_j)$. However, Figs. 35a-35b represent two typical scenarios but there can be many other conditions, for instance pertaining to times series of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ over dense forest. The standard deviations of $T_{B_p}(C)$ and $\sigma_{pp}(C)$ are less because of isometric radiometric characteristics of dense forest canopy, less variation of physical temperature over temporal scale, and homogeneity of vegetation-water-content (VWC). This leads to low absolute value of β parameter over such dense forest landcover. Therefore the magnitude of uncertainty in β parameter is also small. However, the percentage uncertainty in β parameter is relatively high, typically more than 50%.

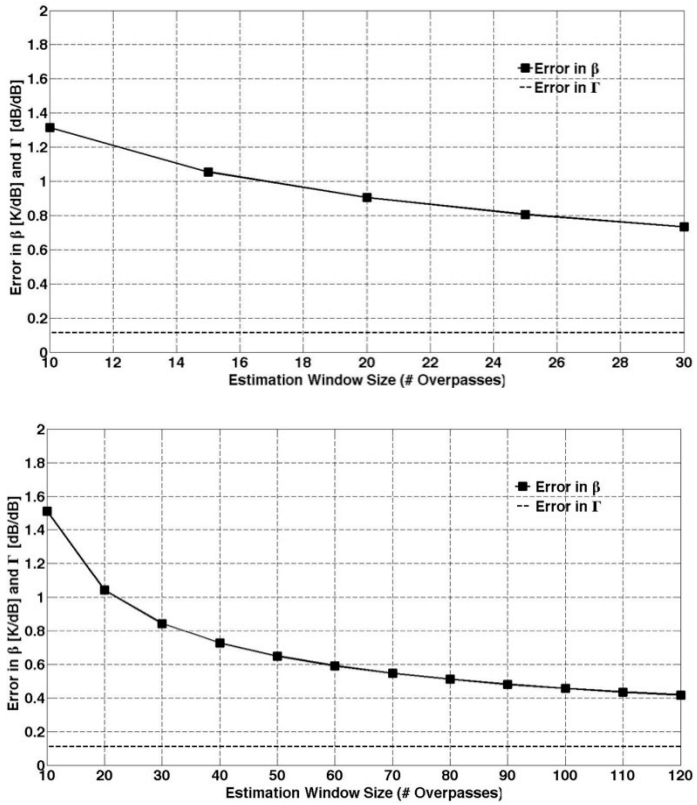


Figure 35: Expected error in algorithm parameters β and Γ: a) when having moderate correlation between $T_{B_p}(C)$ and $\sigma_{pp}(C)$, and max window size is 30 with temporal std of 2.5 [K] for $T_{B_p}(C)$ and 1.5 [dB] temporal std for $\sigma_{pp}(C)$; and b) when having higher correlation between $T_{B_p}(C)$ and $\sigma_{pp}(C)$, and max window size is 120 with temporal std of 7.0 [K] for $T_{B_p}(C)$ and 3 [dB] temporal std for $\sigma_{pp}(C)$.

Another major source of error in $T_{B_p}(M_j)$ is due to correction made to $T_{B_p}(C)$ for presence of waterbodies. Knowledge of exact fraction of waterbody within the *C*-scale is questionable. It is always expected that water fraction information will have certain amount of uncertainty associated with it; therefore we conducted a sensitivity analysis that is shown in Fig. 36. The analysis is conducted for various water fractions ranging from 5 to 50%. We limit the correction of $T_{B_p}(C)$ for waterbody up to 50% because beyond that no retrievals are performed due to excessive influence of water on $T_{B_p}(C)$. Figure 36 shows results of error contributions due to correction made on 5%, 10% and 20% water fractions having 10% uncertainty. Figure 5 also illustrates comparison between the Monte Carlo and analytical solutions (36), and the agreement is extremely good validating the analytical solution.

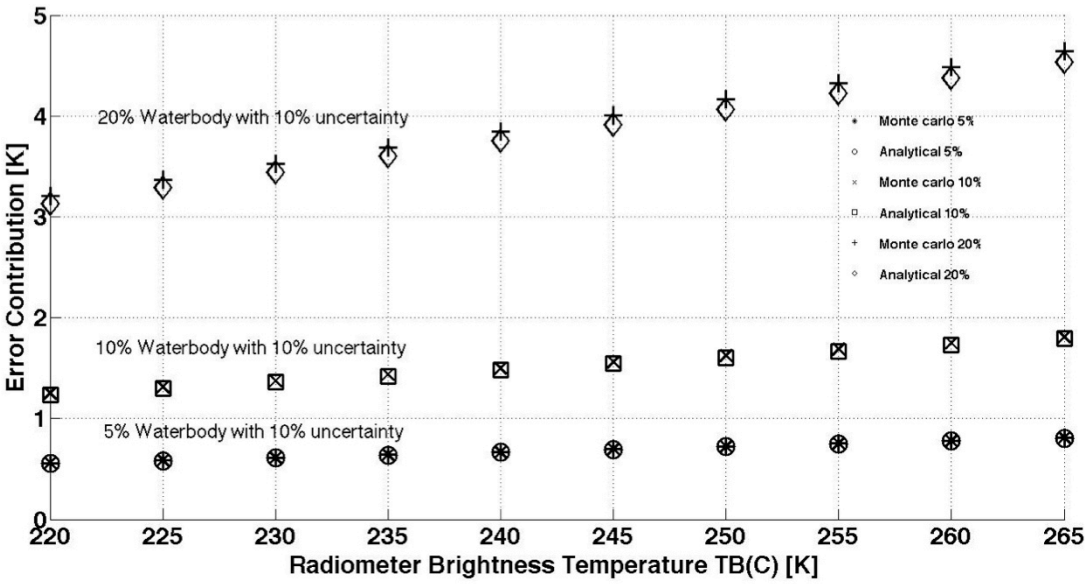


Figure 36: *Error in $T_{B_p}(C)$ due to errors in waterbody fraction within gridded radiometer data at (C) scale.*

Two distinct features are obvious in Fig. 36, they are: i) as expected the error contribution in $T_{B_p}(M_j)$ rises with increase in water fraction with certain uncertainty associated with it; and ii) the error also increases gradually with increase in radiometer brightness temperature. Based on the analysis as shown in Fig. 36, and $T_{B_p}(C)$ having 5% water fraction (with 10% uncertainty) within *C*-scale (36 km EASE2 grid) will have less than 0.7 [K] error that satisfy the error budget, and is considered a nominal value to be used for quality control.

The analytical formulations developed in equations (26, 33, 43, 44, 51) are implemented and coded in the science software of the SMAP L2_SM_AP to provide estimate of expected standard deviation (1-sigma) in the most of the $T_{B_p}(M_j)$ grid cells. Figure 37 illustrates the output of standard deviation in brightness temperature (STD_Tb) extracted from the SMAP Mission global simulation. These STD_Tb values are obtained by taking the Root-Sum-Square (RSS) of standard deviation values obtained from (26), (33), and (51).

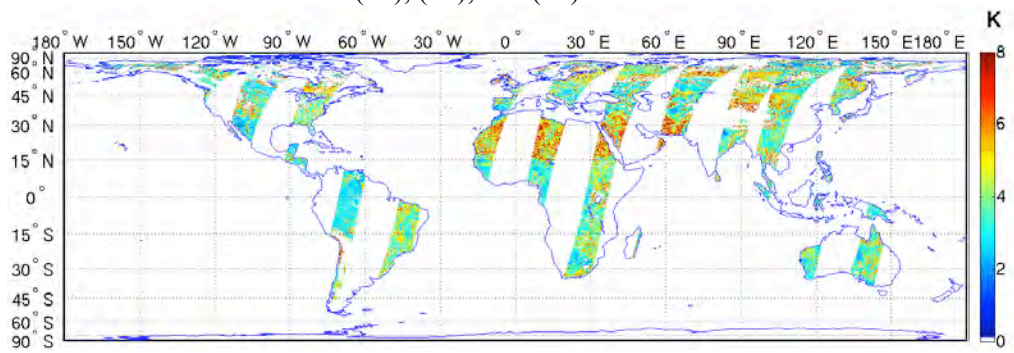


Figure 37: *Standard deviation in brightness temperature $T_{B_p}(M_j)$ at 9 km obtained from global simulation from one month of global simulation.*

Certain features are very obvious and expected, for example, low values of STD_Tb over forested regions and high values of STD_Tb over bare surface (deserts). This is due to the fact that

over forested regions magnitude of β parameters are smaller and the standard deviations in $T_{B_p}(C)$ are also less leading to low values of STD_Tb. The high value of STD_Tb over deserts is primarily due of high magnitude of β parameters and high standard deviations in $T_{B_p}(C)$, and the β parameters over deserts are also not optimized during the simulation. Even with high STD_Tb over deserts, the soil moisture retrievals are within the mission requirement of 0.04 cm^3/cm^3 RMSE. Figure 38 obtained from one month of STD_Tb statistics shows a different perspective of Fig. 37, the spread of STD_Tb categoried with respect to VWC, and also proves that the mean STD_Tb reduces with the increase of VWC. Another aspect that is consistent in Fig. 37 is the high STD_Tb values for regions (northeast Canada) with greater fraction of water bodies within a grid cell. Some of the features that are expected but not apparent in Fig. 37 are high STD_Tb values due more noise (high KpC values) in the co-pol (σ_{pp}) and x-pol (σ_{pq}) radar backscatters at 3 km resolution in the nadir region of the swaths.

Over most the regions that fall under the mission Level-1 thresholds (areas with VWC < 5.0 Kg/m^2, having no snow/ice cover, with no presence of mountainous terrain, and no considerable presence of water bodies) the mean STD_Tb is nearly ~2.7 K, except for the deserted regions. The mean STD_Tb in $T_{B_p}(M_j)$ conforms to the number presented in the Error Budget (Table 5). A comparison of Monte Carlo error estimations and analytical solutions are tabulated in Table 7. The analytical approach to compute error estimates from each contributing source is nearly similar to the error values obtained from Monte Carlo experiments; hence demonstrate the stability and validity of the analytical solutions.

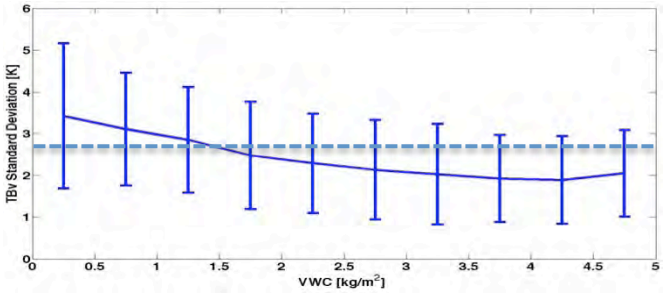


Figure 38: *Spread of standard deviation of $T_{B_p}(M_j)$ with respect to VWC, and the global mean in ~2.7 [K]*

Table 7: Comparison of Monte Carlo error estimations against the analytical solutions.

	Error Source (1 std)	Estimated Error (Monte Carlo)	Estimated Error (Analytical)
1	Radiometer precision and calibration stability, faraday rotation, atmospheric gases, non-precipitating clouds, and gridding	*1.3 K*	*1.3 K*
2	Waterbody fraction	5%	5%
		0.70 K	*0.70 K*
3	Adjusted Corrected T_{B_p} (36 km) RSS	*1.47 K*	*1.47 K*
4	SAR σ_{pp} and σ_{pq} errors	*1.65 K*	*1.55 K*
5	Algorithm Parameter errors	*1.60 K*	*1.50 K*
6	Disaggregated T_{B_p} (9 km) RSS	**2.73 K**	**2.62 K**

5 PRACTICAL CONSIDERATIONS

5.1 Algorithm Tests and Baseline Selection

Prelaunch activities:
I) Development of active-passive retrieval algorithms in accordance with the SMAP configuration and measurements.
II) Modification and tuning of active-passive retrieval algorithms to comply with the SMAP measurements.
III) Performance assessment of the candidate retrieval algorithms on a common geophysical database and prelaunch cal/val data
IV) Downselection of the baseline algorithm based on performance, better understanding of errors, lower bias and value to applications.

Current and future work on SMAP testbed is designed accordingly to fulfill these prelaunch conditions stated above. The work on the testbed will also facilitate quantification of error budget for L2_SM_AP product based on inputs and ancillary data.

The selection of the current baseline algorithm (Section 3.2) is based on the following criteria:

- The primary inputs to the baseline algorithm are direct measurements of SMAP, and the expected uncertainties and errors associated with these measurements are better known than the retrieved/derived SMAP products that are used in the option algorithm (Section 3.3). This facilitates understanding the inner mechanism of the baseline algorithm and better assessment of RMSE in output soil moisture product (~9 km)
- Measurements made using airborne instruments (PALS) during cal/val campaigns directly contribute to the development of the baseline algorithm and help understand the dynamics of the relationship that exists between the inputs to the current baseline algorithm for different conditions and landcovers.
- Provide more confidence in bias removal from the disaggregated brightness temperature and ultimately for the output soil moisture product (~9 km); the mean of the 9 km disaggregated brightness temperature is set to be equal to the SMAP radiometer measurement. In this way biases are removed between the radiometer and disaggregated fields. The same would not be possible with soil moisture retrieval fields from different algorithms.
- In the heritage algorithm the RMSE in the output soil moisture product depends on the RMSE of input soil moisture at coarser resolution. Therefore the RMSE of option algorithm is always greater than the input. However, this limitation is alleviated in the baseline algorithm.
- The baseline algorithm produces a high resolution (9 km) disaggregated brightness temperature which could be of added science value to the suites of products e.g., direct radiance assimilation in weather forecast models, land surface hydrology models or even L4_SM.

Postlaunch activities:
During the postlaunch period most of the activities would be to tune the L2_SM_AP baseline algorithm parameters using the real SMAP observations/products and assess the algorithms using the postlaunch cal/val data. The algorithm assessment will be a comparison of retrievals at 9 km with ground-based observations that have been verified as providing a spatial average of soil

moisture at this scale. However, other types of observations from soil moisture measurement networks (e.g., SCAN network) or products from hydrologic modeling may contribute to post-launch assessment of the active-passive algorithms.

5.2 Calibration and Validation (Cal/Val)

The cal/val plan of L2_SM_AP product is distinctly divided into two parts: 1) Prelaunch, and 2) Postlaunch.

5.2.1 **Prelaunch Cal/Val**

The prelaunch cal/val activities should focus on algorithm development needs. The L2_SM_AP is a merged product of the SMAP radiometer and radar observations. L2_SM_AP baseline and optional algorithms use a time series of fine resolution radar measurements that are contained within a coarse resolution radiometer observation. The behavior and dynamics of the time series used in the algorithm is not fully understood for changing VWC, soil moisture regimes and for different landuse classifications. For prelaunch cal/val of the L2_SM_AP algorithm, a system such as the airborne PALS (Passive and Active L-band System) instrument is the most suitable platform because it provides concurrent and coincident L-band radar and radiometer measurements.

So far, all of the available PALS data from previous soil moisture field campaigns are optimally used for the L2_SM_AP algorithm development efforts. From the L2_SM_AP perspective, the most useful of all PALS data are from the SMEX02 field experiment because the PALS observations are obtained over a dynamic range of soil moisture and vegetation conditions. The current baseline L2_SM_AP algorithm is successfully implemented using the PALS data from the SMEX02 field campaign. The PALS data from all the above mentioned four campaigns are also used to build a preliminary database of algorithm parameters. However, PALS data from these experiments do not have long and consistent time series and therefore impedes objective evaluation of the baseline and option algorithms.

Total area observed by PALS flights during 4 field campaigns (SGP99, SMEX02, CLASIC07, and SMAPVEX08) is nearly 1900 km^2, and the fraction for different landcover types are: Mixed forest = 26%; Grassland = 27%; and Cropland = 47%. The analysis highlights that the observations are predominantly over croplands. The following shows the % of different crop types within the PALS coverage: Bare soil = 13%; Corn = 36%; Soy = 41%; and Wheat = 10%. Based on the above analysis of landcover data, it is obvious that the PALS coverage of different landcover types is limited.

The baseline landcover classification for the SMAP mission science algorithms is the MODIS derived IGBP dataset (refer to the Memo of Landcover for Ancillary Dataset) that comprises 17 classes of landcover types out of which 14 classes are relevant for L2_SM_AP algorithm. Table 8 elaborates the coverage of the different landcover types in percentage (%) of the global domain obtained from the IGBP dataset at 1 km spatial resolution. From Table 8, it is obvious that besides many other crop types, a few important landcover types such as shrubland, savanna and forest (Evergreen Broadleaf) are not yet covered by PALS flights. Shrubland and savanna type landcovers are of more relevance to the L2_SM_AP algorithm because of their extensive coverage over the global landmass (Table 8), and due to their VWC that is typically below 5 kg/m^2 (which is a condition for satisfying the retrieval accuracy requirement in the SMAP requirements). For the time being the Evergreen Broadleaf landcover type can be ignored because the VWC of those forests is generally higher than 5 kg/m^2. Therefore it is imperative to fly PALS especially over regions having shrubland and savanna type landcovers for fine tuning and further development of the baseline L2_SM_AP algorithm.

***Table 8**: Global coverage in % from IGBP dataset.*

MODIS IGBP class	Definition of class	Coverage %
Evergreen needle	Tree canopy cover > 60%	4.0
Evergreen broadleaf	Tree canopy cover > 60%	10.0
Deciduous needle	Tree canopy cover > 60%	0.6
Deciduous broadleaf	Tree canopy cover > 60%	1.6
Mixed forest	Tree canopy cover > 60% but no type exceed 60%	4.7
Closed shrub	Shrub cover > 60%	0.5
Open shrub	60% > shrub cover > 10%	18.3
Woody savanna	Herbaceous system & 60% > forest > 30%	7.5
Savanna	Herbaceous system & 30% > forest > 30%	7.0
Grassland	Herbaceous system & 10% > forest	9.2
Wetland	Water + herbaceous system + tree	0.2
Cropland	Temporary crops	9.0
Crop/pasture mix	No class exceed 60%	2.1
Bare	Vegetation cover < 10%	13.7

The SMAP project has conducted a major prelaunch cal/val campaign SMAPVEX'12. The ground and airborne data acquisition phase of SMAPVEX12 took place over a period of approximately six weeks from June 6 to July 19, 2012 in an agricultural region south of Winnipeg, Manitoba (Canada). The study domain of SMAPVEX12 comprises agricultural fields (corn, soybean, canola and wheat), grasslands and forest sites. The site is about 15 km x 70 km within the larger Red River Watershed. Airborne measurements using active and passive instruments including NASA's L-band Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) flown in a Gulfstream III, and the Passive/Active L-band Sensor (PALS) on board the Twin Otter International. The experiment also provides ground measurements of soil moisture, soil temperature, and others surface characteristics (vegetation, roughness, and soil density, etc.) at a time close to satellite and airborne acquisitions. The data from SMAPVEX12 campaign is are processed and is available for L2_SM_AP algorithm development and testing. The L2_SM_AP ATBD includes results rom data of SMAPVEX12.

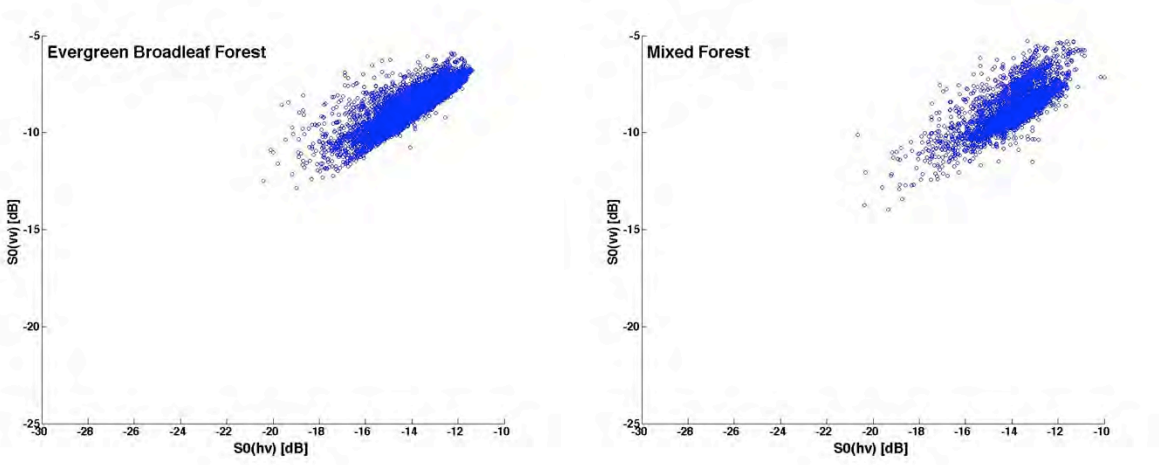
Measurements from the NASA GSFC ComRAD (Combined Radar/Radiometer System) truck-based instrument [22] are also planned for the prelaunch cal/val activities. The upgraded system will be tested in field conditions. After the performance validation in field conditions the instrument is deployed in Maryland at the OPE3 study site for the 2012 growing season. The observations include at least two crop types at the site. Long deployments of ComRAD are recommended over different landcovers in the study site. Longer time series resulting from long deployment of ComRAD will assist to develop robust algorithm parameters. Longer time series data of T_B and σ from ComRad will also help understand the sensitivity of algorithm parameters to the dynamic nature of vegetation and surface conditions.

Use of Aquarius Data for L2_SM_AP Algorithm Parameters

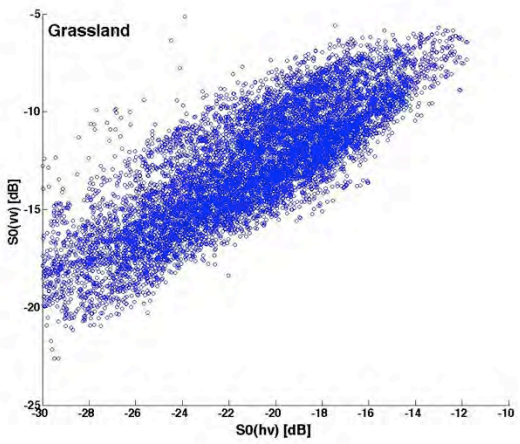
Aquarius was launched on June 10, 2011, and almost three years worth of Aquarius mission data is now available. Aquarius provides L-band 1.2 GHz and 1.413 GHz radar and radiometer measurements, respectively, at three beam positions with incidence angle of 28.7, 37.8, and 45.6 degrees. The spatial resolutions of three beams are slightly different, but all are nearly ~100 km and together provide 390km swath width. With this swath width it takes Aquarius seven days to cover

the whole Earth surface. The satellite is in a polar sun-synchronous orbit crossing the equator at 6pm (ascending) and 6am (descending) local time. The Aquarius radar and radiometer data are quite significant for SMAP active-passive (L2_SM_AP) algorithm because they are co-registered and have similar overpass time on Equator as proposed for the SMAP mission. Though Aquarius radar and radiometer data do not satisfy SMAP mission requirements, however, its L-band radar and radiometer provide a unique opportunity for prelaunch calibration and develop a prior database of parameters for the SMAP active-passive (L2_SM_AP) baseline algorithm.

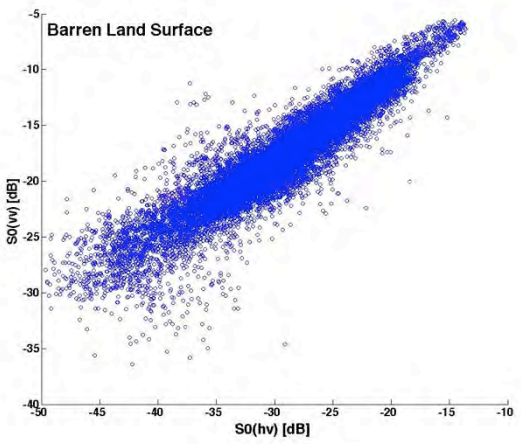
To make the Aquarius data suitable for the SMAP L2_SM_AP baseline algorithm prelaunch calibration analysis, the Aquarius data are preprocessed because of different antenna configuration and coarse spatial resolution. Firstly, the Aquarius radar and radiometer swath data in each beam are separately converted into the 36 km EASE grid by interpolating the data using the inverse-distance-weighted technique. Secondly, the overlapping areas of three beams are selected, which means that for any overlapping area the selected grid cell has at least two beam data. The incidence angle dependence slopes ([dB] or [K] per degree) are statistically calculated for 16 IGBP landcover classes. After obtaining the slopes for each landcover class, the three beam data are scaled to 40-degree incidence angle and binned at 36 km EASE grid projection for 7 days (seven days composite) to get the global coverage of L-band radar and radiometer data. The RFI was eliminated using the quality flag available in the Aquarius data. One year (Aug'11 to July'12) of Aquarius dataset are processed to get 30 global extent dataset of radar and radiometer measurements. The processing step may introduce artifacts in the gridded radar and radiometer data; however, such errors are tolerable because the processed data are used for initial evaluation of parameters (for L2_SM_AP baseline algorithm) by regression.



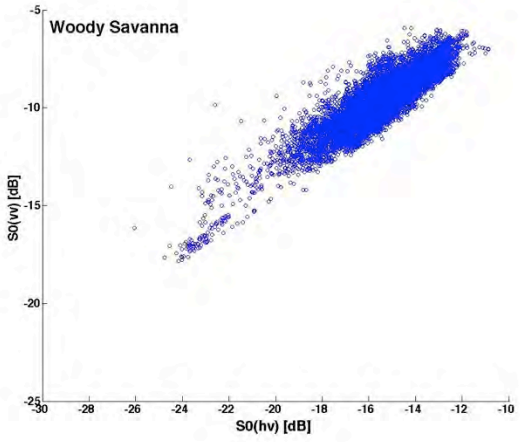
a) b)



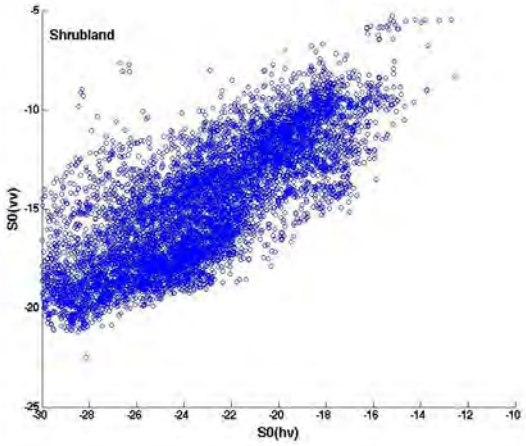
c)



d)



e)



f)

Figure 39*: Scatter plots of σ_{hv} and σ_{vv} for different landcover classes obtained from processed Aquarius data.*

The processed Aquarius data (seven days composite) provides preliminary assessment of the algorithm parameters for SMAP L2_SM_AP baseline algorithm. Figure 39a-f illustrates the correlation between σ_{hv} and σ_{vv} for different landcover classes using one-week period (during August, 2011) data at a global extent. As expected, significant correlations and interesting radar backscatter features are observed for all the relevant landcover classes. The scatter plots (Fig. 39a-f) clearly show that the dynamic range of σ_{hv} and σ_{vv} increases with decreased volume scattering. High correlation between σ_{hv} and σ_{vv} leads to robust estimation of Γ parameter (17). For most of the IGBP landcovers Γ is ~0.7, however, obtained at coarse resolution from Aquarius data.

Analyses are also conducted to obtain initial assessment of β parameter (12) at a global extent. Especially, the Aquarius data provides a new insight of L-band radar and radiometer measurements correlation dynamics over different landcover classes at various parts of the world. Aquarius data over homogeneous regions (e.g., Sahara desert, Australian shrublands, and Amazon forest) are very valuable because the airborne measurements from field campaigns (SMEX02, SGP99, CLASIC, SMAPVEX08, and SMAPVEX12) do not include major biomes and landcovers (shrubland, savanna, desert, rainforest, tundra) that exist on the Earth surface. Figures 40-43

illustrate the value of β parameter and relevant statistics that influence its magnitude. Inappropriate values of β are eliminated from the global map because of lack of sufficient dynamic ranges in σ and T_B data, and freeze-thaw effects that alter the regime of backscatter and brightness temperature. It is obvious from Fig. 41 that reasonably high R^2 values are observed where enough dynamic ranges in σ and T_B time series data exist. Further down selection of β parameter from Aquarius data is conducted to develop a prior database. β parameters are categorized according to landcover classes and screened based on high R^2 (>= 0.75) thresholding, to ensure fidelity of the parameters. Clear trends are observed in β parameters when plotted verses σ_{hv} (Fig. 44) for different landcover classes that indicate increase in magnitude of β with increasing cross-pol backscatter. The physical explanation to such phenomenon is the considerable decrease of dynamic range in σ_{vv} due to increased volume scattering as compared to the changes in range of T_B values. Volume scattering in radar backscatter is influenced by presence of vegetation and surface roughness, and is captured by σ_{hv} measurement, which is directly proportional to volume scattering.

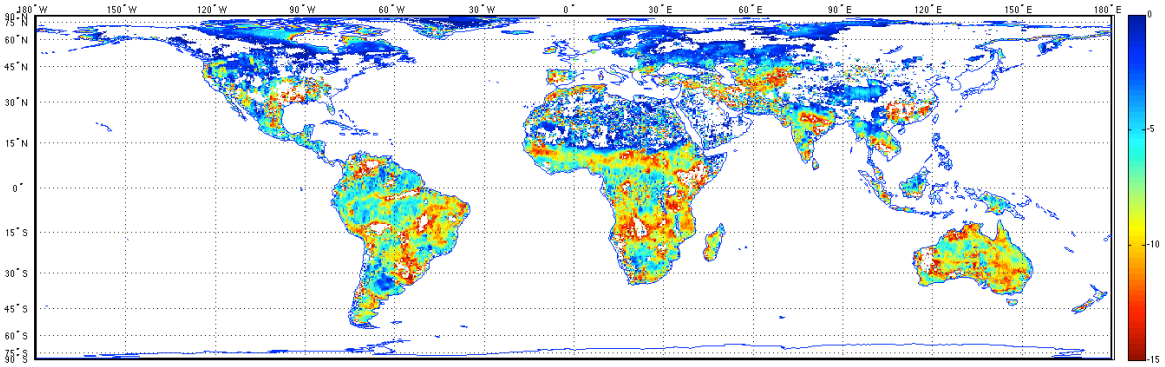


Figure 40: β [K/dB] parameters obtained at global extent using processed Aquarius data for 8 months period (Aug'2011 to Mar'2012).

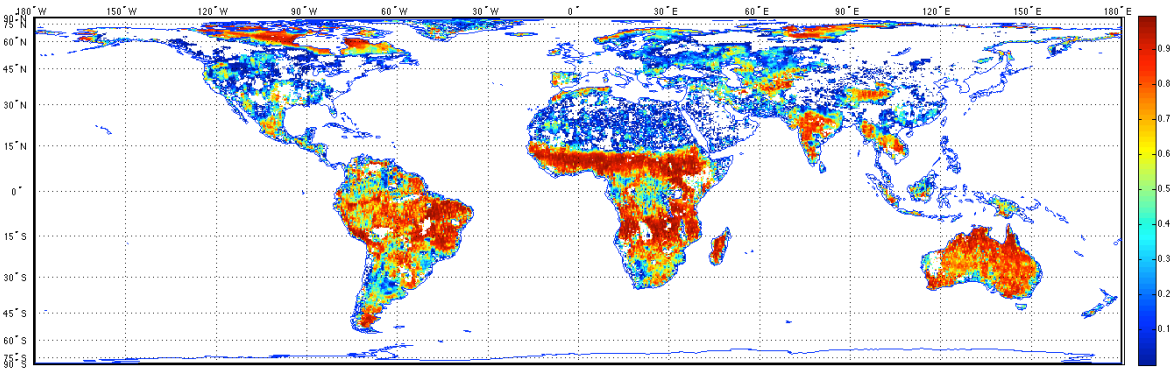


Figure 41: R^2 statistics of correlation between T_{B_v} [K] and σ_{vv} [dB] obtained from the processed Aquarius data for 8 months period (Aug'2011 to Mar'2012).

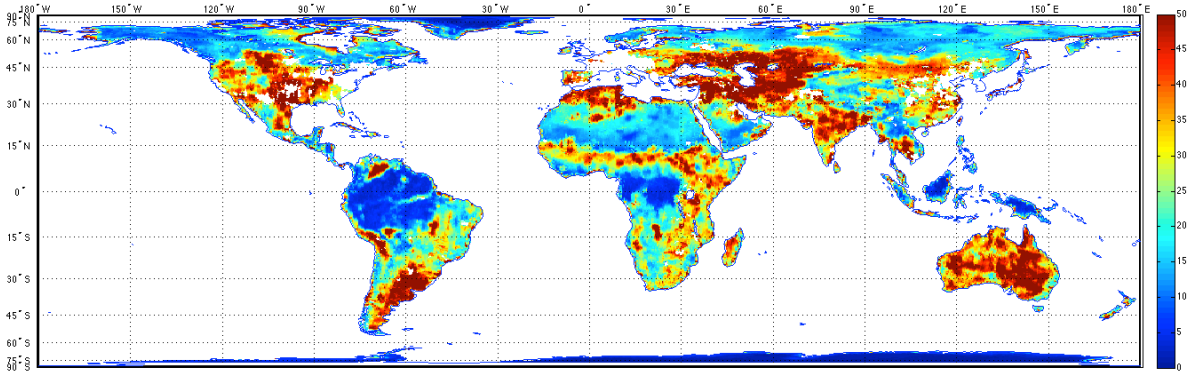


Figure 42: Range in T_{B_v} [K] obtained from the processed Aquarius data for 8 months period (Aug'2011 to Mar'2012).

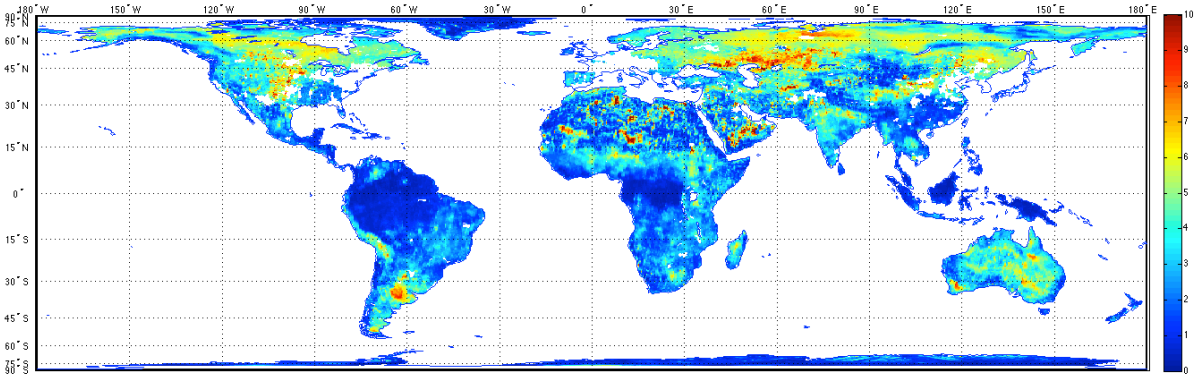


Figure 43: Range in σ_{vv} [dB] obtained from the processed Aquarius data for 8 months period (Aug'2011 to Mar'2012).

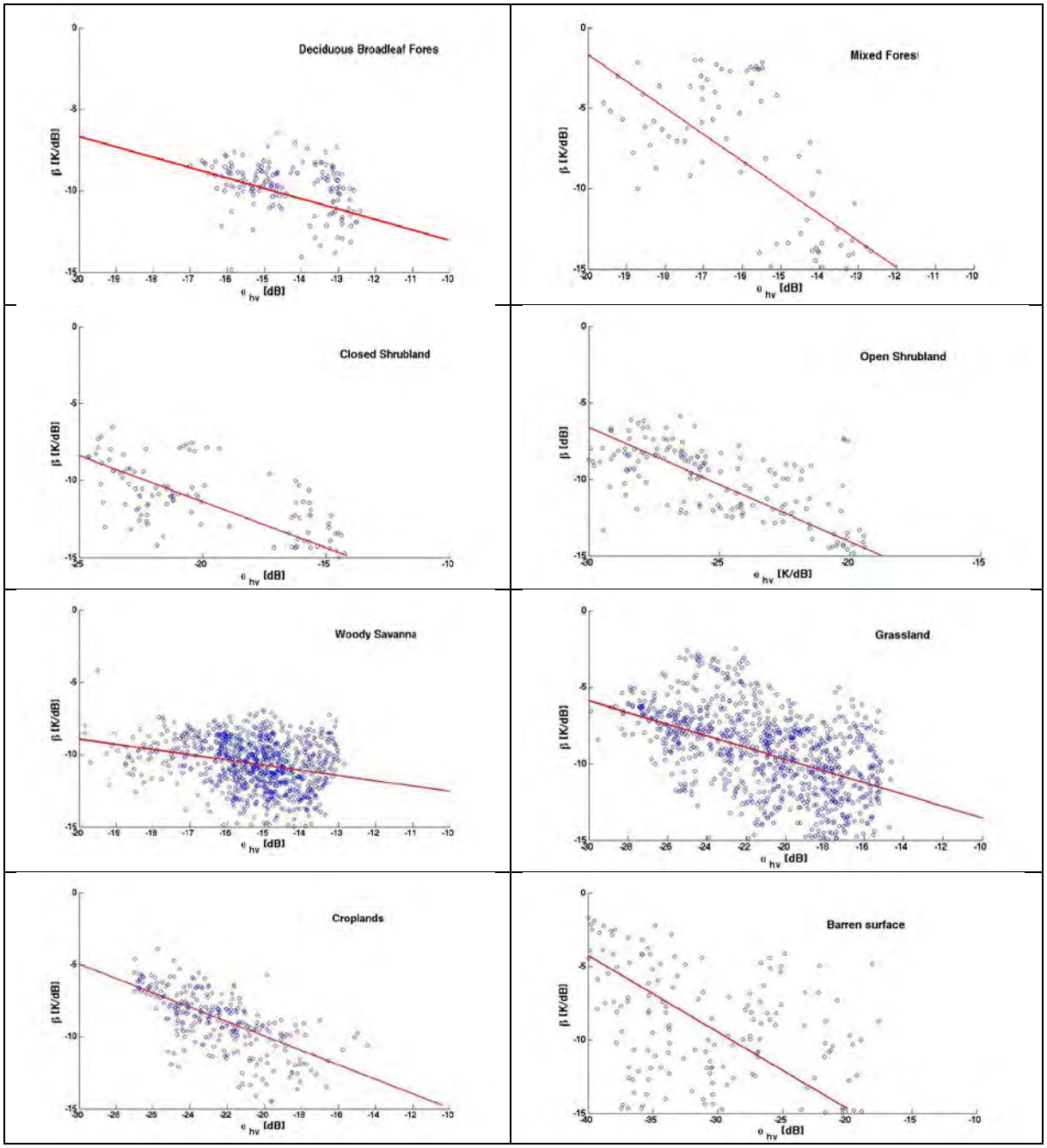


Figure 44: Trend in β parameters with respect to σ_{hv} for various landcovers.

This Aquarius data analysis provides an excellent insight on prior estimates of Γ and β parameters obtained at global extent for the baseline active-passive algorithm (17). The GloSim retrieval results used these prior parameters estimate. In future, prior to the launch of SMAP, the data obtained from Aquarius measurements will be regularly processed to get longer time series data of σ and T_B to make the parameters estimates more robust.

5.2.2 **Postlaunch Cal/Val**

The postlaunch cal/val objective is to gather necessary scientific data to evaluate the data product to verify compliance with the SMAP mission requirement i.e., to produce an estimate of soil moisture in the 0-5 cm layer to an RMSE of 0.04 cm^3/cm^3 of absolute volumetric soil moisture for those areas of the global land area excluding regions of snow and ice, mountainous topography, open water, and vegetation with total water content greater than 5 kg/m^2 at a grid resolution of 9 km. Postlaunch cal/val focus is also to fine tune the L2_SM_AP algorithms based on collected scientific data. The baseline validation will be a comparison of L2_SM_AP at 9 km with ground-based observations that have compatible spatial average of soil moisture. Other form of observations, measurements, and modeling outputs may also contribute to post-launch validation. Described below is a methodology matrix (Table 9) to achieve post-launch cal/val for L2_SM_AP.

***Table 9**: Methodology prioritization matrix for L2_SM_AP cal/val.*

Methodology	Data Required	Importance	Metrics
Core Validation Sites	Grid cell average for all overpass	Primary	RMSE, Bias, Correlation
Sparse Network	Spatially scaled grid cell value for each overpass	Secondary: Pending results of scaling analyses	RMSE, Bias, Correlation
Satellite Product	Orbit based matchups	Secondary	Pattern matching, Correlation
Model Product	Orbit based matchups	Secondary	Pattern matching, Correlation
Field Campaign	Detail estimates for limited set of conditions	Primary	RMSE, Bias, Correlation

Core Validation Sites: As shown in the cal/val methodology prioritization matrix, the Core Validation Sites (CVS) are of primary importance. The CVS will provide spatial average of soil moisture at 9 km spatial resolution with adequate replications, with minimal latency and with well known error quantification (verified against gravimetric measurements). NASA has established agreements with cal/val partners to provide CVS data. There are nearly ~30 sites, out of these, some of them will be selected for CVS based on their configurations of ins situ network and capability to, provide spatial average of soil moisture at 9 km. Selection of CVS for L2_SM_AP primary validation are being conducted, the selection process will also ensure global distribution from various landcovers and biomes.

Sparse Network: Sparse soil moisture measurement networks available in the United States and and other part of the world. The limitation of such networks is low measurement density that results in one measurement site within a 9 km grid cell. It is challenging and difficult to use measurements from one *in situ* site within a 9 km grid cell for validation. Therefore, applied research to identify strategic measurement sites that are temporally stable (least affected and manipulated over a longer time period) and represent the grid average are critical for the success of extensive validation of the

L2_SM_AP 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 Illinois State Water Survey. Gaining access to sparse measurement networks located outside the United States should be considered. 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_AP has to deal with data latency, verification of calibration and spatial scaling. At present the prioritization of sparse network is secondary, however, if the spatial scaling technique using triple collocation method yields reasonable comparison then the sparse networks will be used more effectively for L2_SM_AP cal/val.

Besides the sparse networks there are new technologies (COSMOS and GPS) being evaluated that could potentially provide soil moisture information at much different spatial resolution. The details of these are still being developed. SMAP should participate in the site selection and evaluation of these networks in order to establish both the verification of the relevant depth and scaling to SMAP footprints.

Satellite Products: During the SMAP mission lifetime, it is possible to get overlapping SMOS and GCOM-W soil moisture products. However, the spatial resolutions of these satellites (SMOS and GCOM-W) are coarser then SMAP L2_SM_AP resolution but the soil moisture data could be used for proving synoptic perspective and for matching spatial and temporal trends.

Model Products: Soil moisture data/maps generated through hydrologic modeling at basin-wide and continental scales using assimilated data independent of SMAP products could be used for evaluation across large space and time domains. Most of the models (ECMWF, NCEP, and NASA/GMAO) produce data at coarser resolution then L2_SM_AP. Therefore soil moisture fields from models compared with SMAP soil moisture will be useful for evaluating spatial and temporal trends. There are many caveats while using the model products because of inherent uncertainty in any model-based soil moisture product since this is not one of the National Weather Prediction (NWP) system primary variables. The NWP models typically simulate a thicker surface soil layer than the layer that dominates the satellite measurement. Therefore, while these model products are useful, they must be used very carefully. As a result, they are considered to be of secondary importance for validating L2_SM_AP soil moisture product.

Field Campaign: Soil moisture field campaigns specifically designed for the SMAP mission will take place after the launch. *In situ* surface and profile soil moisture measurements obtained during these experiments will play a vital role in a rigorous validation of the L2_SM_AP product against L1 requirements. Based on a schedule of a 2014 launch, at least one cal/val campaign is needed in the summer of 2015. The focus should be on the core validation sites such as Oklahoma and sites that include different hydroclimatic domains. Intensive and well-distributed *in situ* soil moisture sampling within the validation sites that falls within the Earth fixed grid of 9 km will provide the best spatially averaged soil moisture measurements. Long duration of these field experiments are also critical because it will capture the full dynamic range of soil moisture that will help understand bias in the product and compare the trends of soil moisture evolution.

Using ancillary information: Independent sources of ancillary information like precipitation data sets from GPCP are useful for qualitative comparison of spatial patterns in soil moisture in the L2_SM_AP product.

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

5.3 Ancillary Data

The data required other than SMAP mission data to process and retrieve near surface soil moisture are termed as ancillary data. The baseline algorithm retrieves soil moisture estimates from the disaggregated (downscaled) T_B. Therefore the required ancillary data needed to process L2_SM_AP are similar to L2_SM_P requirements, except for the spatial resolution. The L2_SM_AP 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_AP process needs permanent masks and geophysical parameters as static ancillary data. Permanent masks of water bodies, mountainous region, forest cover (VWC > 5 kg/m2), urban areas, and land area at high resolution (1 – 3 km) resampled in Earth fixed grid is desirable for L2_SM_AP processing and quality control. Sand fraction and clay fraction data at 9 km spatial resolution are examples of static ancillary data essential to retrieve soil moisture. The L2_SM_AP process will use static ancillary data archived in SDS for the mission. Dynamic ancillary data poses more challenges because they need frequent updates (daily, biweekly, monthly, and seasonally). The dynamic data required for L2_SM_AP process are effective surface soil temperature, VWC (for vegetation opacity τ), surface roughness, and landuse-landcover.

Ancillary data from various resources are analyzed, and are selected as baseline ancillary data. SMAP ancillary data reports are being written for the individual ancillary data listed in Table 10. These reports document the rationale for the choice of the primary source of the ancillary data, and will be made available to the public. Table 10 lists most of the static and dynamic ancillary data, data source, updating frequency and desired spatial resolution required for the L2_SM_AP retrieval. The amount and type of ancillary data needed are dependent to some extent on the choice of the specific retrieval algorithms. Some examples of the ancillary data used in L2_SM_AP retrieval are shown in Fig. 45-47.

***Table 10**: Ancillary data required to produce the SMAP L2_SM_AP product.*

Parameter	Updating Frequency	Desired Grid Resolution	Data Type	Data Source
%Sand and %Clay	Done once	9 km	Static	Composite of soil databases (HWSD, FAO, ASRIS, STATSGO, NSDC)
Soil Texture	Done once	9 km	Static	Composite of soil databases (HWSD, FAO, ASRIS, STATSGO, NSDC)
Vegetation (b and ω)	1-2 weeks	9 km	Static	Values obtained from L2_SM_P ATBD look-up table
Vegetation Opacity (τ)	Daily/Weekly	9 km	Dynamic	MODIS 1 km NDVI converted to VWC and then to $\tau = (b*\text{VWC})$
Roughness (h)	Monthly	9 km	Static	L2_SM_P ATBD look-up table
Effective soil temperature	Daily	9 km	Dynamic	ECMWF / MERRA (TBC)

Waterbodies (permanent)	Yearly	9 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	9 km	Static	SRTM and GTOPO30 derived elevation, slope, range and variance
Freeze/Thaw flag	Daily	3 km	Dynamic	Obtained from L2_SM_A
Urban area fraction	Done one	9 km	Static	GRUMP data

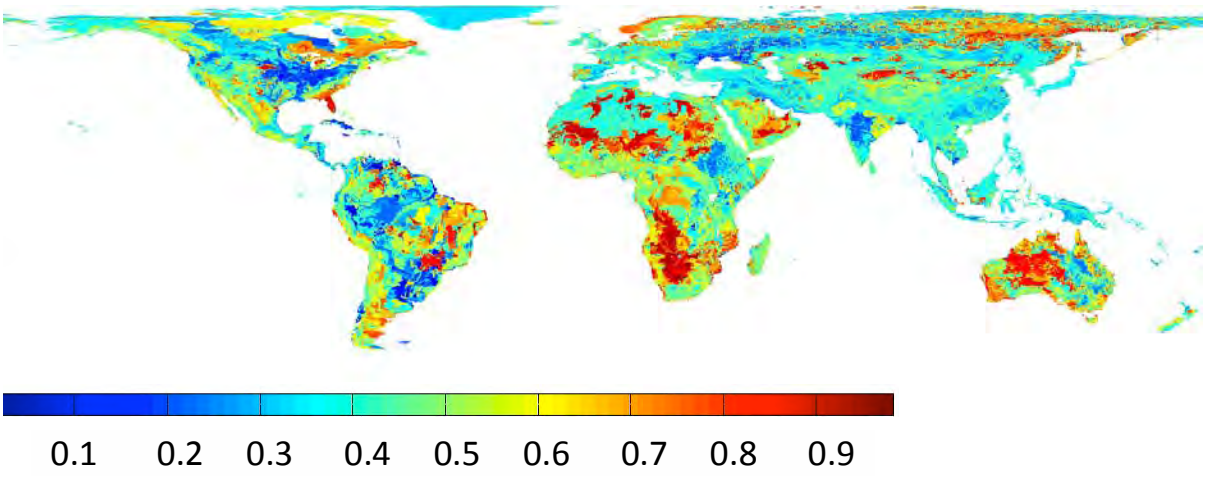


Figure 45: Global sand fraction of top soil at 9 km EASE grid projection.

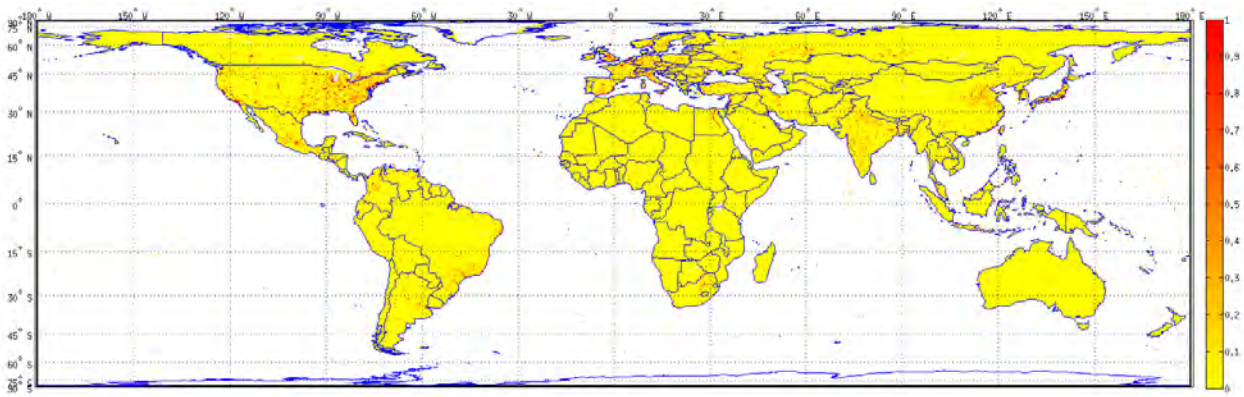


Figure 46: Urban extent fraction gridded at 9 km EASE grid projection.

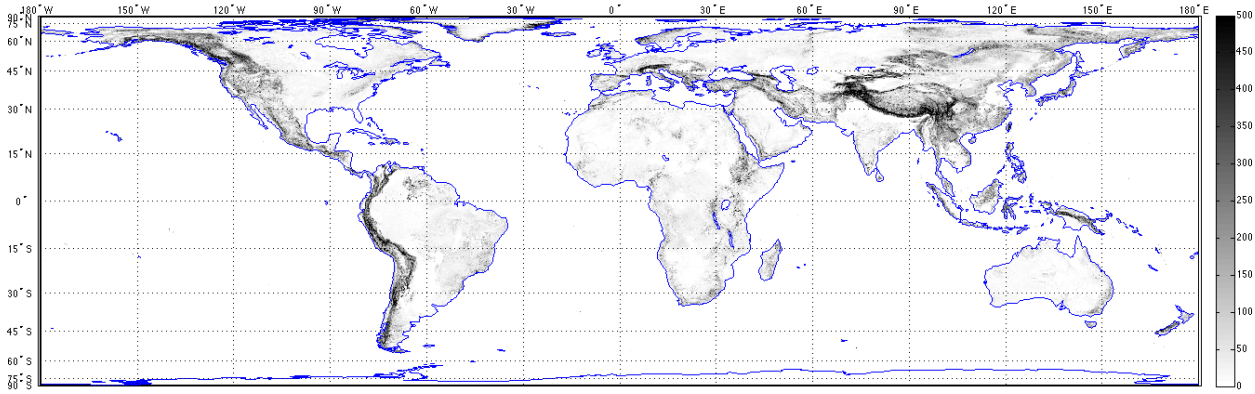


Figure 47: Standard deviation in DEM gridded at 9 km EASE grid projection.

Table 10 also lists the parameters (b, ω, and h) essential for soil moisture retrieval using downscaled T_B (9 km). These parameters are dependent on the landcover types. Details of these parameters are available in L2_SM_P ATBD and the associated ancillary data report.

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_AP algorithms and ultimately the accuracy of the output product. This highlights the fact that the baseline and option algorithms should be robust to accommodate the expected inherent errors in the ancillary data. Results mentioned in Section 4.2.2 from Monte Carlo study conducted for the baseline algorithm show the effects and robustness due to errors in ancillary data. The issue of latency involved is more prominent for the dynamic ancillary data that requires daily update because they influence the stated latency requirements of the L2_SM_AP product availability to the public within 24 hours of data acquisition by the satellite. With the current L2_SM_AP baseline algorithm, the soil effective temperature is one such dynamic ancillary data parameter. However, the option algorithm is also affected indirectly by the latency of soil effective temperature. Current assessment indicates that there are no major obstacles based on the ancillary data latency that would prevent the L2_SM_AP product from meeting its requirements. Figure 48 show another use of ancillary data used as masks of water bodies, mountainous region, forest cover/VWC maps, urban areas, and land area at high resolution (1 – 3 km) resampled in Earth fixed grid of 9 km used to create a map where SMAP is expected to L1 requirement.

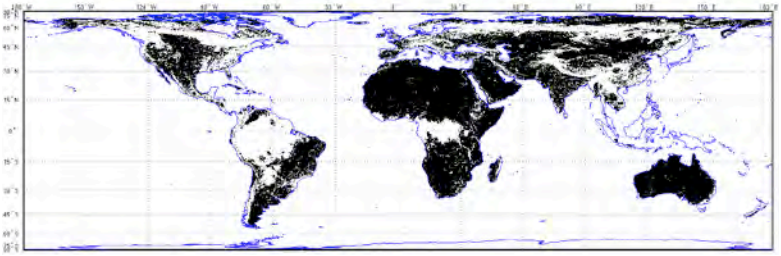


Figure 48: *Retrievable mask (black colored regions) that is expected to meet L1 requirements prepared with following specifications: with VWC ≤ 5 [kg/m^2], with urban fraction ≤ 25%, with water fraction ≤ 10%, and with DEM slope standard deviation ≤ 3 [deg].*

5.4 Quality Control and Diagnostics

Quality control (QC) is an integral part of the L2_SM_AP processing. The QC steps of L2_SM_AP processing are based on the flags that are provided with the input data streams (L2_SM_P, and L2_SM_A), different types of masks, flags, and fractional coverage of other variables provided by ancillary data. The L2_SM_AP will process all data that have favorable conditions for soil moisture retrieval (VWC <= 5 kg/m^2, no rain, no snow cover, no frozen ground, no RFI, sufficient distance from open water). However, soil moisture retrieval will also be conducted for regions with VWC > 5, rain, RFI repaired data, and places closer to water bodies, but appropriate flags will be added to these data points indicating their susceptibility to potentially high errors. The product specification table provided in Section 6 elaborates the fields for QC bit flags. A flow diagram in Fig. 48 illustrates the decision tree to perform L2_SM_AP retrieval.

As shown in Fig. 48, the L2_SM_AP processing involves merging of two data stream i.e., L2_SM_P and L2_SM_A. Therefore, the QC of L2_SM_AP output is influenced by these input data streams. In other words, the QC flags of L2_SM_AP output are the union of QC flags from L2_SM_P and L2_SM_A data streams. However, due to differences in spatial resolution of the inputs (L2_SM_P and L2_SM_A) and output (L2_SM_AP), the assignment of QC flags in L2_SM_AP may differ from the flags associated with the inputs. The thresholds of ancillary data that initiate flagging in the L2_SM_AP product are still TBD/TBC. For example, T_{B_p} data in L2_SM_P are corrected for the presence of water bodies. Studies are being 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_AP processing. The water body fraction is reported for all land-based 9 km grid cells in L2_SM_AP 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_AP retrieval is performed at all the grid cells irrespective of VWC but QC flag set only for grid cell having VWC > 5 kg/m^2. No retrievals are performed for L2_SM_AP grid cells that are associated with RFI, water body fraction above a particular threshold, frozen ground, snow, and urban fraction above threshold.

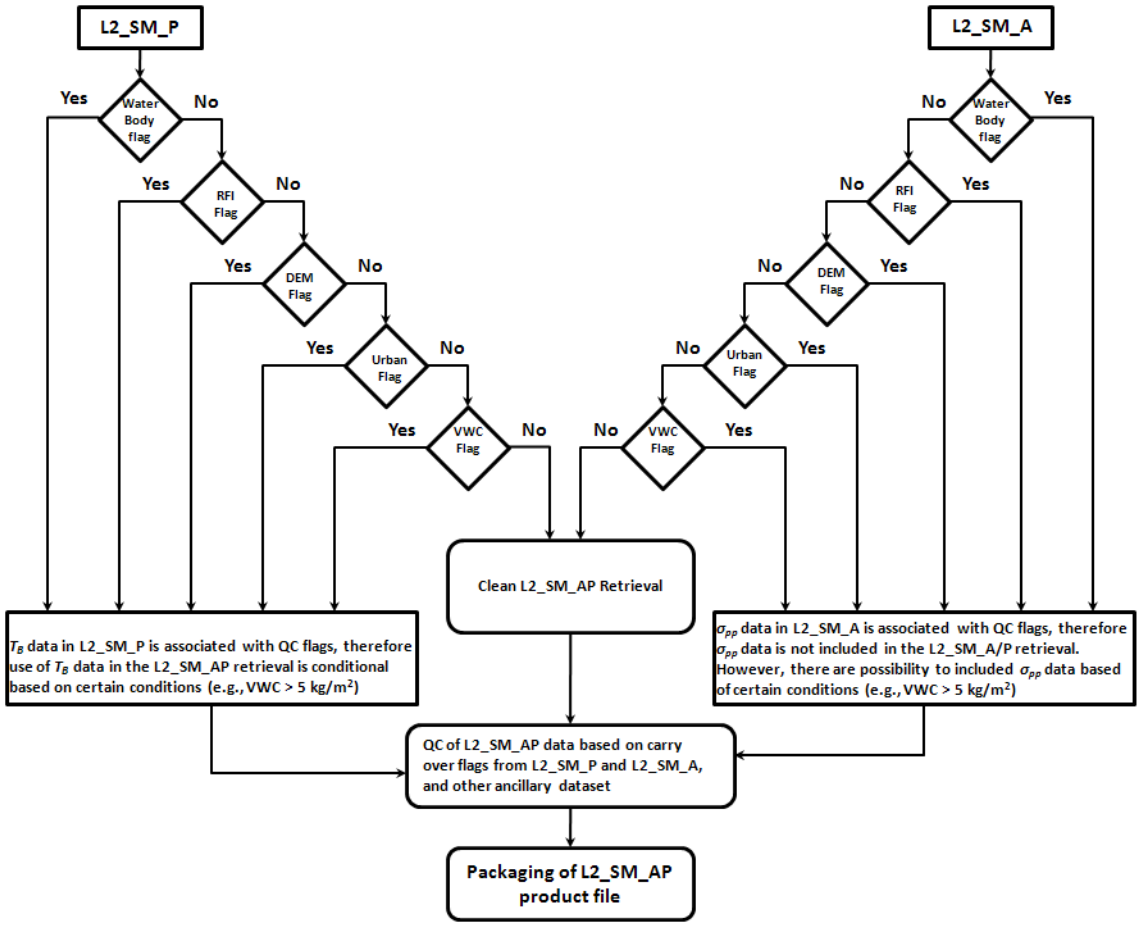


***Figure 48**: Decision Tree of QC for L2_SM_AP product.*

Thresholds from masks that will initiate flags and operational decisions to process L2_SM_AP product are used in GloSim and will be available in L2_SM_AP product. Thresholds that initiate the flagging operation is mentioned as follows:

Open water body flag: The open water fraction will be produced by the SMAP L2_SM_A product coupled that will be coupled with prior information on permanent water bodies from the MODIS (MOD44W) database. This information will be used to flag grid cells during soil moisture retrieval processing in the following way:

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.

RFI flag: Presence of RFI in the SMAP T_B and σ data adversely affects the L2_SM_AP algorithm. Therefore specific logics are inbuilt in the L2_SM_AP processor to initiate flag during soil moisture retrievals. The RFI flag is initiates as follows:

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

RFI detected in σ and repaired: Flag and retrieve soil moisture
RFI detected in T_B and not repaired: Flag and do not retrieve soil moisture.
RFI detected in σ and not repaired: Flag and do not retrieve soil moisture.

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

Snow data indicates no snow: Retrieve soil moisture, do not flag.
Snow data indicates snow: Flag, do not retrieve soil moisture.

Precipitation flag: Presence of heavy rainfall during SMAP data acquisition may adversely affect the T_B and σ measurements. The precipitation data from GPM or from the model nowcast/forecast will be used to flag the concern 9 km grid cell. L2_SM_AP retrievals will be performed irrespective of rainfall; however, the grid cell will be flagged in case of the presence of precipitation.

VWC flag: L2_SM_AP retrievals are conducted for all the locations irrespective of VWC level. The grid cells are flagged for VWC greater than 5 kg/m^2.

Urban area flag: Presence of urban area adversely affects the L-band radiometric measurements. The presence of urban area within the SMAP measurement is likely to bias soil moisture retrievals. Currently the L2_SM_AP processor flag the regions having urban area as follows:

Urban fraction is 0.00 – 0.25: Retrieve soil moisture, do not flag.
Urban fraction is 0.25 – 1.00: Flag and retrieve soil moisture.

Mountain area flag: Statistics of mountainous regions that initiate flags and operational decisions during L2_SM_AP processing is standard deviation of slope. There are other options and thresholds 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_AP 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_AP processor flags the region where standard deviation of slope is more than 3 degrees, however, the retrieval is performed for all locations.

5.5 Numerical Computation and Storage Consideration

The expected computational requirements of L2_SM_AP activities are moderate. With the present technology in computing and archiving of electronic files, it can easily accommodate the L2_SM_AP processing. The maximum computational demand (processor time and cache memory) is anticipated during ingestion of L2_SM_A data into the algorithm. With the current infrastructure at Science Data System (SDS), no bottlenecks are expected while running the L2_SM_AP software PGE for the SMAP product. Preliminary studies calculated a data volume of 89 MB for the half orbit granule of L2_SM_AP product. This make a yearly volume of ~600 GB based on ~14.5/day half orbit granules. This is only 0.25% of total storage requirements of SMAP science related data and products.

5.6 Programming Consideration

Processing and operational codes for L2_SM_AP algorithm is written in Fortran to make it consistent with other algorithms. This facilitates inter-algorithm functioning of switches, data and information transactions. Programming of L2_SM_AP algorithm adheres to standard coding specifications to ensure the consistent, maintainable, and readable code deliveries within the SMAP SDS. Fortran programming of L2_SM_AP 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 simplify the Algorithm-to-PGE development process. The L2_SM_AP Fortran codes contain comments and version control information to track the changes and streamline the development of software. A software specification document will be developed for documentation of all the source codes.

5.7 Exception Handling

To obtain the L2_SM_AP product from SMAP 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_AP algorithm on the SDS testbed. The development of L2_SM_AP software also includes identification of expected exceptions. However, the formulation and computer coding of L2_SM_AP software will be made robust to withstand the expected exceptions and exit normally with messages in case of any exception.

5.8 Interface Assumptions

The L2_SM_AP baseline and option algorithms will generate data based on the input of L2_SM_P (water bodies corrected brightness temperature fields) and the L2_SM_A (radar backscatter cross-section) products. Masks and flags contained in the foundation products will be propagated in the L2_SM_AP. 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_AP product will be requested to be included in the parent SMAP data product.

5.9 Compositing L2_SM_AP to L3_SM_AP

On a daily basis there will be ~14.5 files of L2_SM_AP product that corresponds to descending pass (half orbit granule/swath) of SMAP. L3_SM_AP is a daily global composite of these 14.5 L2_SM_AP product files. Global compositing of L2_SM_AP files have two aspects; 1) with no overlap of swaths, and 2) with overlap of swaths.

With no overlap of swaths: Latitudes between 60˚N and 60˚S will have this characteristic. The compositing process is just mapping the L2_SM_AP file values to the global 9 km Earth fixed grid based on the latitude-longitude or grid-id information. It is also obvious that in such situation there will be only one value per cell.

With overlap of swaths: This situation will arise in polar latitudes from 60˚N to 90˚N and 60˚S to 90˚S. For each L2_SM_AP product file the compositing process is mapping the values to the global 9 km Earth fixed grid based on the lat-lon or grid-id. In this scenario, in a particular 9 km grid cell there are possibilities of having more than one value. The selection of final soil moisture value from this array is TBD. Various options like average value, first value, last value, and the value closest to 6:00 AM local time are being explored. However, based on the preliminary consideration, the value closest to 6:00 AM local time is used currently to produce L3_SM_AP product file.

5.10 Latency in L2_SM_AP Product

The L2_SM_AP product has to meet the latency requirement of 24 hours from the time of data acquisition by the satellite. The latency of the L2_SM_AP product is mainly governed by the latencies involved in the input data streams (L2_SM_A and L2_SM_P) and the dynamic ancillary data (surface temperature). To ensure that the L2_SM_AP product meets the latency requirement the following criteria are required in operational processing at the SMAP SDS.

a) Processing of L2_SM_A by ~22 hours from the time of data acquisition by the satellite.
b) Processing of L2_SM_P by ~22 hours from the time of data acquisition by the satellite.
c) Availability of dynamic ancillary data by at most 15 hours from the time of data acquisition by the satellite.

With the current computing capability at the SMAP SDS and using only one node, the operational processing of the L2_SM_AP algorithm to product an output file is accomplished in less than 3 minutes, provided all the necessary input data and ancillary data are available.

6 PRODUCTS SPECIFICATIONS

The L2_SM_AP product contains relevant outputs from the current baseline algorithm (Section 3.2) and option algorithms, dynamic ancillary data used in soil moisture retrieval, and metadata information. The L2_SM_AP product also contains retrievals at 3 km EASE2 grid spatial resolution. Table 9 provides the fields contained in the L2_SM_AP product file. The primary data field of the L2_SM_AP product is at 9 km resolution (EASE grid), and the 3 km resolution product is experimental. For completeness, disaggregated T_B from the current baseline and option algorithms, averaged radar backscatter at 9 km, dynamic ancillary inputs and algorithm parameters are included. Geolocation information (Latitude, Longitude, and EASE grid-id) and UTC time of the spacecraft overpass are also included for all the grid cells at 9 km and 3 km. The L2_SM_AP product file comprises two distinct data groups, one for 9 km EASE2 grid and second one for 3 km EASE2 grid. The L2_SM_AP product file contains only those EASE2 grid cells of the descending pass where relevant data are present; specifically the spacecraft observations over open water are ignored. With the current design of the L2_SM_AP product file, the contents can be easily mapped on to the 9 km and 3 km fixed EASE grid cells using the geolocation information present in the product file. Complete details about the contents are provided in the L2_SM_AP product specification document. The metadata and the populated product fields will be packaged in HDF5 file format and will be made available to public from the NASA DAAC at National Snow and Ice Data Center (NSIDC).

***Table 9**: L2_SM_AP product file data fields (mostly in alphabetical order) at 9 km and 3 km EASE2 grid resolutions that contain geolocation information, disaggregated T_B, aggregated σ, geophysical data fields, QC data, and dynamic ancillary data.*

Data Group at 9 km		
Data Elements		
Latitude	real	degrees
Longitude	real	degrees
EASE_row_index	integer	(-)
EASE_column_index	integer	(-)
Albedo	real	(-)
Alpha_tbh_hh	real	Kelvins
Alpha_tbv_vv	real	Kelvins
Bare_soil_roughness_retrieved	real	
Beta_tbh_hh	real	Kelvins /dB
Beta_tbv_vv	real	Kelvins /dB
disaggregated_tb_h_qual_flag_option1	bit flag	(-)
disaggregated_tb_h_qual_flag_option2	bit flag	(-)
disaggregated_tb_v_qual_flag_option1	bit flag	(-)
disaggregated_tb_v_qual_flag_option2	bit flag	(-)
distance_from_nadir	real	meters
freeze_thaw_fraction	real	(-)
gamma_hh_xpol	real	dB/dB
gamma_vv_xpol	real	dB/dB
landcover_class	enum	(-)
radar_vegetation_index	real	(-)
retrieval_param_flag	bit flag	(-)
retrieval_qual_flag	bit flag	(-)
retrieval_qual_flag_option2	bit flag	(-)
sigma0_hh_aggregated	real	dB
sigma0_vv_aggregated	real	dB
sigma0_xpol_aggregated	real	dB
soil_moisture	real	cm^3/cm^3
soil_moisture_h_option1	real	cm^3/cm^3
soil_moisture_h_option2	real	cm^3/cm^3
soil_moisture_h_option3	real	cm^3/cm^3
soil_moisture_h_std_option1	real	cm^3/cm^3
soil_moisture_h_std_option2	real	cm^3/cm^3

soil_moisture_h_std_option3	real	cm^3/cm^3
soil_moisture_option2	real	cm^3/cm^3
soil_moisture_option3	real	cm^3/cm^3
soil_moisture_std_dev	real	cm^3/cm^3
soil_moisture_v_option1	real	cm^3/cm^3
soil_moisture_v_option2	real	cm^3/cm^3
soil_moisture_v_option3	real	cm^3/cm^3
soil_moisture_v_std_option1	real	cm^3/cm^3
soil_moisture_v_std_option2	real	cm^3/cm^3
soil_moisture_v_std_option3	real	cm^3/cm^3
spacecraft_overpass_time_seconds	real	seconds
spacecraft_overpass_time_utc	string	(-)
Surface_flag	bit flag	(-)
surface_temperature	real	˚C
tb_h_disaggregated	real	Kelvins
tb_h_disaggregated_option2	real	Kelvins
tb_h_disaggregated_std_option1	enum	Kelvins
tb_h_disaggregated_std_option2	real	Kelvins
tb_v_disaggregated	real	Kelvins
tb_v_disaggregated_option2	real	Kelvins
tb_v_disaggregated_std_option1	real	Kelvins
tb_v_disaggregated_std_option2	real	Kelvins
vegetation opacity	real	(-)
vegetation_water_content	real	m^2/m^2
water_body_fraction	real	(-)

Data Group at 3 km		
Data Elements		
Latitude_3km	real	degrees
Longitude_3km	real	degrees
EASE_row_index_3km	integer	(-)
EASE_column_index_3km	integer	(-)
Albedo_3km	real	(-)
Bare_soil_roughness_retrieved_3km	real	
disaggregated_tb_h_qual_flag_3km	bit flag	(-)
disaggregated_tb_v_qual_flag_3km	bit flag	(-)
distance_from_nadir_3km	real	meters
landcover_class_3km	enum	(-)
radar_vegetation_index_3km	real	(-)

retrieval_qual_flag_3km	bit flag	(-)
sigma0_hh_3km	real	dB
sigma0_vv_3km	real	dB
sigma0_xpol_3km	real	dB
soil_moisture_3km	real	cm^3/cm^3
soil_moisture_h_3km	real	cm^3/cm^3
soil_moisture_h_std_3km	real	cm^3/cm^3
soil_moisture_v_3km	real	cm^3/cm^3
soil_moisture_v_std_3km	real	cm^3/cm^3
spacecraft_overpass_time_seconds_3km	real	seconds
Surface_flag_3km	bit flag	(-)
surface_temperature_3km	real	˚C
tb_h_disaggregated_3km	real	Kelvins
tb_h_disaggregated_std_3km	enum	Kelvins
tb_v_disaggregated_3km	real	Kelvins
tb_v_disaggregated_std_3km	real	Kelvins
vegetation_opacity_3km	real	(-)
vegetation_water_content_3km	real	m^2/m^2
water_body_fraction_3km	real	(-)

There are total of 56 data elements at 9 km EASE2 grid resolution and 30 data elements at 3 km EASE2 grid resolution. Some of the data elements are experimental and may be removed after the SMAP Cal/Val phase.

Description of data elements (Table 9) in L2_SM_AP product file:

Data Group at 9 km

1) Latitude: Latitude of the center of the EASE based grid cell.
2) Longitude: Longitude of the center of the EASE based grid cell.
3) EASE_row_index: The row index of the 9 km EASE grid cell that contains the associated data.
4) EASE_column_index: The column index of the 9 km EASE grid cell that contains the associated data.
5) Albedo: Vegetation albedo used in Tau-Omega model for soil moisture retrieval.
6) Alpha_tbh_hh: Alpha parameter derived from T_{Bh} and σ_{hh} time series.
7) Alpha_tbv_vv: Alpha parameter derived from T_{Bv} and σ_{vv} time series.
8) Bare_soil_roughness: Retrieved soil roughness provided by the active soil moisture algorithm.
9) Beta_tbh_hh: Beta parameter derived from T_{Bh} and σ_{hh} time series used in the Active/Passive retrieval algorithm for the corresponding EASE grid cell.

10) Beta_tbv_vv: Beta parameter derived from T_{Bv} and σ_{vv} time series used in the Active/Passive retrieval algorithm for the corresponding EASE grid cell.
11) Disaggregated_tb_h_qual_flag_option1: Bit flags that record the conditions and the quality of the disaggregated T_{Bh} for baseline algorithm (option 1).
12) Disaggregated_tb_h_qual_flag_option2: Bit flags that record the conditions and the quality of the disaggregated T_{Bh} for option algorithm 2.
13) Disaggregated_tb_v_qual_flag_option1: Bit flags that record the conditions and the quality of the disaggregated T_{Bv} for baseline algorithm (option 1).
14) Disaggregated_tb_v_qual_flag_option2: Bit flags that record the conditions and the quality of the disaggregated T_{Bv} for for option algorithm 2..
15) Distance_from_nadir: The distance from the center of the 9 km EASE grid cell to the spacecraft's sub-nadir track on the Earth's surface.
16) Freeze_thaw_fraction: Fraction of the 9 km grid cell that is denoted as frozen. Based on binary flag that specifies freeze thaw conditions in each of the component 3 km grid cells.
17) Gamma_hh_xpol: Heterogeneity parameter of the Active-Passive algorithm used to obtain disaggregated T_{Bh}.
18) Gamma_vv_xpol: Heterogeneity parameter of the Active-Passive algorithm used to obtain disaggregated T_{Bv}.
19) Landcover_class: Enumerated data that classify MODIS derived dominant landcover at 9 km.
20) Radar_vegetation_index: Vegetation index derived from radar backscatter.
21) Retrieval_param_flag: Bit flag that inform about the parameters, tentative, may be removed from data product.
22) Retrieval_qual_flag: Bit flags that record the conditions and the quality of the retrieval algorithms that generate soil moisture for the baseline (option 1) active-passive algorithm.
23) Retrieval_qual_flag_option2: Bit flags that record the conditions and the quality of the retrieval algorithms that generate soil moisture for the option 2 active-passive algorithm.
24) Sigma0_hh_aggregated: The outcome of aggregating a set of 3 km horizontal polarization radar backscatter measurements into a 9 km EASE grid cell.
25) Sigma0_vv_aggregated: The outcome of aggregating a set of 3 km vertical polarization radar backscatter measurements into a 9 km EASE grid cell.
26) Sigma0_xpol_aggregated: The outcome of aggregating a set of 3 km cross-polarized radar backscatter measurements into a 9 km EASE grid cell.
27) Soil_moisture: Representative soil moisture measurement for the 9 km Earth based grid cell (this data field is populated with the most confident retrieval) of baseline algorithm.
28) Soil_moisture_h_option1: Soil moisture retrieved using disaggregated T_{Bh} for the 9 km Earth based grid cell for baseline algorithm (option 1).
29) Soil_moisture_h_option2: Soil moisture retrieved using disaggregated T_{Bh} for the 9 km Earth based grid cell for option 2 active-passive algorithm.
30) Soil_moisture_h_option3: Soil moisture retrieved using disaggregated T_{Bh} for the 9 km Earth based grid cell for option 3 active-passive algorithm.
31) Soil_moisture_h_std_option1: Soil moisture standard deviation retrieved using disaggregated T_{Bh} for the 9 km Earth based grid cell for baseline algorithm (option 1).
32) Soil_moisture_h_std_option2: Soil moisture standard deviation retrieved using disaggregated T_{Bh} for the 9 km Earth based grid cell for option 2 active-passive algorithm.
33) Soil_moisture_h_std_option2: Soil moisture standard deviation retrieved using disaggregated T_{Bh} for the 9 km Earth based grid cell for option 3 active-passive algorithm.
34) Soil_moisture_option2: Representative soil moisture measurement for the 9 km Earth based grid cell (this data field is populated with the most confident retrieval) of option2 active-passive algorithm.

35) Soil_moisture_option3: Representative soil moisture measurement for the 9 km Earth based grid cell (this data field is populated with the most confident retrieval) of of option3 active-passive algorithm.
36) Soil_moisture_v_option1: Soil moisture retrieved using disaggregated T_{Bv} for the 9 km Earth based grid cell for baseline algorithm (option 1).
37) Soil_moisture_v_option2: Soil moisture retrieved using disaggregated T_{Bv} for the 9 km Earth based grid cell for option 2 active-passive algorithm.
38) Soil_moisture_v_option3: Soil moisture retrieved using disaggregated T_{Bv} for the 9 km Earth based grid cell for option 3 active-passive algorithm.
39) Soil_moisture_v_std_option1: Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 9 km Earth based grid cell for baseline algorithm (option 1).
40) Soil_moisture_v_std_option2: Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 9 km Earth based grid cell for option 2 active-passive algorithm.
41) Soil_moisture_v_std_option2: Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 9 km Earth based grid cell for option 3 active-passive algorithm.
42) Spacecraft_overpass_time_utc: Time of spacecraft overpass in UTC. The spacecraft time is relative to the 36 km EASE grid cell that contains each 9 km EASE grid cell represented in this data product.
43) Spacecraft_overpass_time_seconds: Number of seconds since a specified epoch that represents the spacecraft overpass relative to the 36 km EASE grid cell that contains each 9 km EASE grid cell represented in this data product.
44) Surface_flag: Bit flags that record ambient surface conditions for the grid cell.
45) Surface_temperature: Temperature at land surface based on a geophysical model (e.g., GMAO).
46) Tb_h_disaggregated: Horizontal polarization brightness temperature adjusted for the presence of water bodies and disaggregated from the 36 km EASE grid cells into 9 km EASE grid cells.
47) Tb_h_disaggregated_option2: Horizontal polarization brightness temperature adjusted for the presence of water bodies and disaggregated from the 36 km EASE grid cells into 9 km EASE grid cells obtained using option2 active-passive algorithm.
48) Tb_h_disaggregated_std_option1: Horizontal polarization brightness temperature standard deviation at 9 km EASE grid cells for baseline algorithm (option1).
49) Tb_h_disaggregated_std_option2: Horizontal polarization brightness temperature standard deviation at 9 km EASE grid cells for option2 active-passive algorithm.
50) Tb_v_disaggregated: Vertical polarization brightness temperature adjusted for the presence of water bodies and disaggregated from the 36 km EASE grid cells into 9 km EASE grid cells.
51) Tb_v_disaggregated_option2: Vertical polarization brightness temperature adjusted for the presence of water bodies and disaggregated from the 36 km EASE grid cells into 9 km EASE grid cells obtained using option2 active-passive algorithm.
52) Tb_v_disaggregated_std_option1: Vertical polarization brightness temperature standard deviation at 9 km EASE grid cells for baseline algorithm (option1).
53) Tb_v_disaggregated_std_option2: Vertical polarization brightness temperature standard deviation at 9 km EASE grid cells for option2 active-passive algorithm.
54) Vegetation_opacity: The measured opacity of the vegetation in the 9 km grid cell.
55) Vegetation_Water_Content: Water content of vegetation in the 9 km grid cell.
56) Water_body_fraction: Fraction of the area of 9 km grid cell that is a permanent or transient water body. Derived from the DEM and radar processing.

Data Group at 3 km

1) Latitude_3km: Latitude of the center of the EASE based grid cell at 3 km.
2) Longitude_3km: Longitude of the center of the EASE based grid cell at 3 km.
3) EASE_row_index_3km: The row index of the 3 km EASE grid cell that contains the associated data.
4) EASE_column_index_3km: The column index of the 3 km EASE grid cell that contains the associated data.
5) Albedo_3km: Vegetation albedo used in Tau-Omega model for soil moisture retrieval at 3 km.
6) Bare_soil_roughness_3km: Retrieved soil roughness provided by the active soil moisture algorithm at 3 km.
7) Disaggregated_tb_h_qual_flag_3km: Bit flags that record the conditions and the quality of the disaggregated T_{Bh} for baseline algorithm (option 1) at 3 km.
8) Disaggregated_tb_v_qual_flag_3km: Bit flags that record the conditions and the quality of the disaggregated T_{Bv} for baseline algorithm (option 1) at 3 km.
9) Distance_from_nadir_3km: The distance from the center of the 3 km EASE grid cell to the spacecraft's sub-nadir track on the Earth's surface.
10) Landcover_class_3km: Enumerated data that classify MODIS derived dominant landcover at 3 km.
11) Radar_vegetation_index_3km: Vegetation index derived from radar backscatter at 3 km.
12) Retrieval_qual_flag_3km: Bit flags that record the conditions and the quality of the retrieval algorithms that generate soil moisture for the baseline (option 1) active-passive algorithm at 3 km.
13) Sigma0_hh_3km: The outcome of aggregating a set of 3 km horizontal polarization radar backscatter measurements into a 3 km EASE grid cell.
14) Sigma0_vv_3km: The outcome of aggregating a set of 3 km vertical polarization radar backscatter measurements into a 3 km EASE grid cell.
15) Sigma0_xpol_3km: The outcome of aggregating a set of 3 km cross-polarized radar backscatter measurements into a 3 km EASE grid cell.
16) Soil_moisture_3km: Representative soil moisture measurement for the 3 km Earth based grid cell (this data field is populated with the most confident retrieval) of baseline algorithm.
17) Soil_moisture_h__3km: Soil moisture retrieved using disaggregated T_{Bh} for the 3 km Earth based grid cell for baseline algorithm (option 1).
18) Soil_moisture_h_std__3km: Soil moisture standard deviation retrieved using disaggregated T_{Bh} for the 3 km Earth based grid cell for baseline algorithm (option 1).
19) Soil_moisture_v_3km: Soil moisture retrieved using disaggregated T_{Bv} for the 3 km Earth based grid cell for baseline algorithm (option 1).
20) Soil_moisture_v_std_3km: Soil moisture standard deviation retrieved using disaggregated T_{Bv} for the 3 km Earth based grid cell for baseline algorithm (option 1).
21) Spacecraft_overpass_time_seconds_3km: Number of seconds since a specified epoch that represents the spacecraft overpass relative to the 36 km EASE grid cell that contains each 3 km EASE grid cell represented in this data product.
22) Surface_flag_3km: Bit flags that record ambient surface conditions for the grid cell at 3 km.
23) Surface_temperature_3km: Temperature at land surface based on a geophysical model (e.g., GMAO).
24) Tb_h_disaggregated_3km: Horizontal polarization brightness temperature adjusted for the presence of water bodies and disaggregated from the 36 km EASE grid cells into 3 km EASE grid cells.
25) Tb_h_disaggregated_std__3km: Horizontal polarization brightness temperature standard deviation at 3 km EASE grid cells for baseline algorithm (option1).

26) Tb_v_disaggregated_3km: Vertical polarization brightness temperature adjusted for the presence of water bodies and disaggregated from the 36 km EASE grid cells into 3 km EASE grid cells.
27) Tb_v_disaggregated_std_3km: Vertical polarization brightness temperature standard deviation at 3 km EASE grid cells for baseline algorithm (option1).
28) Vegetation_opacity_3km: The measured opacity of the vegetation in the 3 km grid cell.
29) Vegetation_water_content_3km: Water content of vegetation in the 3 km grid cell.
30) Water_body_fraction_3km: Fraction of the area of 3 km grid cell that is a permanent or transient water body. Derived from the DEM and radar processing.

The L3_SM_AP product is a subset of L2_SM_AP data fields. The L3_SM_AP is organized in EASE2 grid arrays of 9 km resolution. No 3km experimental data fields are included in the L3_SM_AP product. The data fields of L3_SM_AP are summarized in the following table.

***Table 11**: L3_SM_AP product file data fields (mostly in alphabetical order) at 9 km and 3 km EASE grid resolutions that contain geolocation information, disaggregated T_B, aggregated σ, geophysical data fields, QC data, and dynamic ancillary data.*

Data Elements		
Latitude	real	degrees
Longitude	real	degrees
EASE_row_index	integer	(-)
EASE_column_index	integer	(-)
Albedo	real	(-)
Alpha_tbh_hh	real	Kelvins
Alpha_tbv_vv	real	Kelvins
Bare_soil_roughness_retrieved	real	
Beta_tbh_hh	real	Kelvins /dB
Beta_tbv_vv	real	Kelvins /dB
disaggregated_tb_h_qual_flag	bit flag	(-)
disaggregated_tb_v_qual_flag	bit flag	(-)
distance_from_nadir	real	meters
freeze_thaw_fraction	real	(-)
gamma_hh_xpol	real	dB/dB
gamma_vv_xpol	real	dB/dB
landcover_class	enum	(-)
radar_vegetation_index	real	(-)
retrieval_param_flag	bit flag	(-)
retrieval_qual_flag	bit flag	(-)
retrieval_qual_flag_option2	bit flag	(-)
sigma0_hh_aggregated	real	dB

sigma0_vv_aggregated	real	dB
sigma0_xpol_aggregated	real	dB
soil_moisture	real	cm^3/cm^3
soil_moisture_std_dev	real	cm^3/cm^3
spacecraft_overpass_time_seconds	real	seconds
spacecraft_overpass_time_utc	string	(-)
Surface_flag	bit flag	(-)
surface_temperature	real	˚C
tb_h_disaggregated	real	Kelvins
tb_h_disaggregated_std	enum	Kelvins
tb_v_disaggregated	real	Kelvins
tb_v_disaggregated_std	real	Kelvins
vegetation opacity	real	(-)
vegetation_water_content	real	m^2/m^2
water_body_fraction	real	(-)

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

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

[8] 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.

[9] 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.

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

[11] 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.

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

[13] U. Narayan, V. Lakshmi, and T. J. Jackson, "High resoluiton 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Ulaby, F. T., R. K. Moore and A. K. Fung, Microwave Remote Sensing: Active and Passive, 3, Volume Scattering and Emission Theory, Advanced Systems and Applications, Reading, MA, Addison-Wesley, 1986.

[20] Dobson, C.M. and F. T. Ulaby, "Preliminary evaluation of the SIR-B response to soil moisture, surface roughness, and crop canopy cover", *IEEE Trans. Geosci. Remote Sens*., GE-24(4), 517-526, 1986.

[21] D. J. Leroux, N. N. Das, D. Entekhabi, A. Colliander, E. G. Njoku, T. J. Jackson, and S. Yueh, "Disaggregation of brightness temperatures and soil moisture retrievals during the SMAPVEX12 campaign", *IEEE Trans. Geosci. Remote Sens*., Submitted (under review).

[22] P. E. O'Neill, R. H. Lang, M. Kurum, K. R. Carver, and C. Utku, "Multisensor microwave remote sensing of NASA's combined radar/radiometer (ComRAD) system," in Proc. MicroRad'06, San Juan, Puerto Rico, pp. 50–54, Feb. 2006.

8 GLOSSARY

(Under development)

ALGORITHM. (1) Software delivered by a science investigator to be used as the primary tool in the generation of science products. The term includes executable code, source code, job control scripts, as well as documentation. (2) A prescription for the calculation of a quantity; used to derive geophysical properties from observations and to facilitate calculation of state variables in models.

ANCILLARY DATA. Data other than instrument data required to perform an instrument's data processing. They include orbit data, attitude data, time information, spacecraft engineering data, calibration data, data quality information, data from other instruments (spaceborne, airborne, ground-based) and models.

CALIBRATION. (1) The activities involved in adjusting an instrument to be intrinsically accurate, either before or after launch (i.e., "instrument calibration"). (2) The process of collecting instrument characterization information (scale, offset, nonlinearity, operational, and environmental effects), using either laboratory standards, field standards, or modeling, which is used to interpret instrument measurements (i.e., "data calibration").