Algorithm Theoretical Basis Document
AirMOSS Level 2/3 Root-Zone Soil Moisture (RZSM) Product

Mahta Moghaddam, Alireza Tabatabaeenejad, and Richard Chen
University of Southern California

Sassan Saatchi, Sermsak Jaruwatanadilok, Mariko Burgin, Xueyang Duan, and My-Linh Truong-Loï
Jet Propulsion Laboratory

Version 3.2
August 2016

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Mission Background and Science Objectives

North American ecosystems are critical components of the global carbon cycle, exchanging large amounts of carbon dioxide and other gases with the atmosphere. Net ecosystem exchange (NEE) quantifies these carbon fluxes, but current continental-scale estimates contain high levels of uncertainty. Root-zone soil moisture (RZSM) and its spatial and temporal heterogeneity influence NEE and contribute as much as 60–80% to the estimation uncertainty. The goal of the Airborne Microwave Observatory of Subcanopy and Subsurface (AirMOSS) investigation is to provide a new NEE estimate for North America that is constrained by RZSM measurements. The AirMOSS investigation will accomplish this by:

(1) Providing high-resolution radar backscatter observations used to calculate estimates of RZSM over regions representative of the major North American biomes;

(2) Estimating the impact of RZSM on regional carbon fluxes; and

(3) Integrating the measurement-constrained estimates of regional carbon fluxes to the continental scale of North America

The AirMOSS data products and expected science results are tailored to meet the need to reduce uncertainty in estimates of NEE through the development of methodologies to integrate remote sensing observations, in-ground soil sensors, and flux tower data into regional/continental flux models. Additionally, AirMOSS data provide a direct means for evaluating RZSM algorithms of the SMAP Decadal Survey mission and assessing the impact of fine-scale heterogeneities in its coarse-resolution products.

AirMOSS surveys were conducted over regions of approximately 100km × 25km surrounding FLUXNET tower sites within nine biomes (baseline investigation) representative of North American regimes primarily responsible for determining the North America NEE. Table 1 summarizes the locations and biome types. The surveys provided radar backscatter measurements at approximately 100m spatial resolution scales (or better) and at select sub-weekly, seasonal, and annual time scales.

The AirMOSS radar and associated RZSM benchmark datasets, a first of their kind, are a major breakthrough over current point-scale RZSM measurements and provide a critical input to carbon flux models. AirMOSS science data products include RZSM at 100m resolution (Level-2/3 RZSM), estimates of RZSM through hydrologic data assimilation and the use of land surface models (Level-4 RZSM), estimates of NEE at 1-km resolution through ecosystem modeling (Level-4A NEE), and integrated North American NEE estimates at 50 km resolution (Level-4B NEE). The project concludes with a new estimate of North American NEE and a quantitative assessment of the reduced uncertainty.

1.2 Measurement Approach

AirMOSS produces estimates of RZSM with data from the P-band SAR to capture the effects of gradients of soil, topography, and vegetation heterogeneity over an area of 100km × 25km at each of the 9 biomes listed in Table 1. AirMOSS has acquired this high resolution radar data at the temporal sampling frequency specified below, during an estimated 8.7 flight hours of SAR sampling during each campaign. AirMOSS conducted 21 campaigns during the growing seasons in 2012–2015.

TABLE 1. Summary of the AirMOSS baseline mission science study sites. The sites are selected to represent nine major biomes in North America.

Biome	Biome type International Geosphere-Biosphere Programme (IGBP) Vegetation class	Example site name and location
1	Boreal forest/evergreen needle-leaf, mixed forest, cropland	Boreal Ecosystem Research and Monitoring Sites (BERMS), Saskatchewan, Canada
2	Boreal transitional/mixed forest	Howland, Maine and Harvard, Massachusetts forests
3	Temperate forest/mixed forest, cropland	Duke forest, North Carolina
4	Temperate forest/evergreen needle-leaf	Metolius, Oregon
5	Temperate grasslands/crops	Marena Oklahoma in-Situ Sensor Testbed (MOISST), Marena, Oklahoma
6	Mediterranean forest/woody savanna	Tonzi Ranch, California
7	Desert and shrub/open shrubland and grassland	Walnut Gulch, Arizona
8	Subtropical dry forest/broadleaf deciduous, crops, woody savanna	Chamela, Mexico
9	Tropical moist forest/evergreen broadleaf, crops	La Selva, Costa Rica

In-ground soil parameters and ancillary data described in this document were acquired to calibrate and verify the science products required to reduce the NEE uncertainty. The in-ground sensors were installed before the first science deployment, and collected data through the end of the investigation in 2015 at seven of the nine flux tower sites. These sensors measure surface soil temperature, soil matric potential, soil moisture content at several different depths, and in addition, the precipitation amount was measured at each site.

The NEE estimate and accompanying uncertainty reduction estimates are achieved by combining the ecosystem demography simulations incorporating explicit sub-grid heterogeneity for each biome's coverage area with appropriate statistical weighting and interpolation values for all grid cells within the north American region. The appropriate statistical weighting and interpolation values for each grid cell are calculated via ecosystem demography model derived estimates of each grid cell's contribution to continental scale NEE, and the spatial and temporal correlations between NEE values within grid cells. The Level 4 products concerned with NEE computations are described in other documents.

With regards to the L2/3-RZSM products, the AirMOSS project has the following objectives:

a) Complete three 10–14 day campaigns during the growing season in a 12-month period for each of the biomes 1–5 in Table 1, one campaign during each of the dry and wet seasons for biomes 6 and 7 in Table 1, and one campaign for biomes 8 and 9 in Table 1.

b) Repeat the temporal and spatial sampling scenario above during 3 consecutive 12-month periods.

c) Execute aircraft science flights at each biome over an area of at least 2500 km^2 during each campaign.

d) Retrieve RZSM from radar backscatter measurements with a soil penetration depth of at least 25 centimeters in non-saturated conditions of less than 0.35 volumetric soil water content, and under less than 15 kg/m^2 of vegetation biomass, when averaged over all 10–14 day campaigns, all years and all sites.

e) Retrieve RZSM as described in d) with at most 0.05 (m^3/m^3) root-mean squared error, when averaged over all 10–14 day campaigns, all years and all sites.

1.3 Instrument Characteristics

Table 2 summarizes key radar parameters for the AirMOSS P-band radar (Chapin *et al.*, 2012). The AirMOSS SAR was used to make fully polarimetric imagery at ten sites (9 biomes) representative of the major North America biomes in the period 2012–2015. The SAR products have been used to retrieve RZSM and its spatial heterogeneity. To save cost, to reduce schedule, and to reduce risk, the AirMOSS radar reused many GeoSAR and the NASA/JPL L-band Unmanned Aerial Vehicle Synthetic Aperture Radar (UAVSAR) elements, leveraging the

heritage of the success of previous airborne SARs. Like the UAVSAR system, the AirMOSS radar system was fitted into a pod that mounted under a NASA Gulfstream III (G-III). The NASA G-III aircraft are equipped with a precision autopilot that facilitates repeat-pass interferometry, although repeat-pass interferometry is not planned as part of the AirMOSS project. The pod and radome are a copy of those used for UAVSAR. Both UAVSAR and AirMOSS have the same mechanical and electrical interfaces between the pod and the aircraft. The AirMOSS data were processed into synoptic imagery using synthetic aperture processing techniques. The absolute calibration accuracy of AirMOSS data is 0.6 dB.

TABLE 2. AirMOSS Radar Instrument Characteristics

Frequency range (MHz)	280–440
Nominal bandwidth (MHz)	20
Selectable bandwidths (MHz)	6, 20, 40, 80
Nominal slant range resolution (m)	7
Azimuth Resolution (m)	0.8
Incidence angle range (degree)	20–60
Altitude range (km)	3.5–12.5
Pulse length range (μsec)	5-50
Peak transmit power (kW)	2.0
Nominal spatial posting (m)	15

2 THE L2/3-RZSM DATA PRODUCT OVERVIEW

This AirMOSS L2/3 Algorithm Theoretical Basis Document (ATBD) describes retrieval of the AirMOSS L2/3 RZSM using polarimetric P-band radar backscattering coefficients. These products are generated using snapshots of such radar measurements, expected to be available several times during each campaign year as specified later in the document when temporal resolution is discussed. The L2/3-RZSM products are the only set of products directly derived from the AirMOSS radar measurements (L1-S0) and are unique to the AirMOSS investigation. The higher-level products (L4) of the AirMOSS investigation rely on the L2/3-RZSM products and various models to generate their outputs.

2.1 Historical Perspective

Radar remote sensing has long been recognized as a key component of an effective Earth observing system, due to the strong relationship between the radar backscattering coefficient (sometimes also called radar backscatter) and the vegetation geometric and bulk material properties. Recognizing the sensitivity of radar measurements to vegetation variables, many radar instruments have been flown on airborne and spaceborne platforms for synoptic observations of vegetation cover and soil moisture. A number of associated radar scattering models have been developed, which predict the backscattering cross sections for different

frequencies and polarizations for terrain covered with shrubs, forests, agricultural crops, or no vegetation at all. Using these scattering models, sensitivity analyses have been carried out to better understand SAR backscatter due to various terrain covers, and subsequently retrieve information about the areas imaged.

The information retrieval process, which is the goal of any remote sensing device, consists of two major analysis components for radars: the forward model and the inverse model. The forward model, or the scattering model, is as follows: given the description of a particular vegetated scene including the underlying ground, a forward scattering model predicts the radar measurements. The inverse model, on the other hand, relies on the data to estimate variables describing the scene and ideally uses no prior knowledge of the scattering scene parameters. The same physical relationships captured in the forward model are at work in the inverse model, and it follows that the inverse models have the forward model at their core.

Retrieval of vegetation canopy parameters and, simultaneously, subsurface/subcanopy soil moisture is a relatively new field, enabled by increasingly more sophisticated forward scattering models and inversion techniques in the past few years. The early retrievals of soil moisture (such as those by Dubois *et al.*, 1995; Oh *et al.*, 1992; Haddad *et al.*, 1996; many others) were based on empirical or semi-empirical relationships between radar scattering coefficients and soil moisture, and effectively ignored vegetation; they were not valid when the amount of vegetation increased beyond those of crops in the early-growth stages. They also were principally applied to L-band and higher frequencies, and were focused on surface soil moisture. The retrievals of vegetation properties such as biomass, on the other hand, have traditionally ignored the effect of ground properties (e.g., soil moisture) and have often been based on empirical and semi-empirical relationships between radar measurements and ground-based measurements of biomass (Dobson *et al.*, 1992; Dobson *et al.*, 1995; LeToan *et al.*, 1992; Imhoff *et al.*, 1995; Israelsson *et al.*, 1997; Ranson *et al.*, 1994; many others). Both classes of retrievals had limited validity with respect to geographic locations and the range of values of their parameters of interest.

The retrieval of surface soil moisture at L-band is still of great interest and has motivated missions such as the Soil Moisture Active and Passive (SMAP) mission. The retrieval algorithms have become more sophisticated and now take vegetation into account (SMAP ATBD—not available for public citation). This has been made possible by the development of relatively accurate numerical scattering models of vegetated surfaces, based on which soil moisture, surface roughness, and/or vegetation properties (such as vegetation water content (VWC) in the case of SMAP) can be estimated. If roughness or VWC is not estimated, then it must be known through ancillary information. The SMAP retrieval algorithms now recognize that the geometry of vegetation has an important role in determining the radar backscattering coefficients, and therefore the models and retrievals are tailored to specific species of vegetation.

As part of a comprehensive scattering model, ground surface roughness spectrum needs to be considered. A surface can be qualitatively designated as “smooth” (or slightly rough) if the

variations in height are less than $\lambda/25\cos\theta$ where θ is the incidence angle, and qualitatively designated as "rough" if the variations in height are greater than $\lambda/4\cos\theta$ (Van Zyl and Kim, 2011). Table 3 shows the smoothness and roughness criteria for the AirMOSS radar frequency range (280–440 MHz) and the AirMOSS swath (25–45 degrees). To the best of our knowledge, the RMS surface roughness at all AirMOSS sites is less than 5cm. Therefore, even at the highest operating frequency of AirMOSS the ground surface can be considered to be smooth (i.e., only slightly rough).

TABLE 3. Smoothness/roughness criteria within the AirMOSS frequency and incidence angle ranges.

Center Frequency (MHz)	Incidence Angle	Wavelength (m)	Smoothness Criteria $\lambda/25\cos\theta$	Roughness Criteria $\lambda/4\cos\theta$
280	25 degrees	1.07	5 cm	30 cm
280	45 degrees	1.07	6 cm	37 cm
360	25 degrees	0.83	3 cm	23 cm
360	45 degrees	0.83	4 cm	29 cm
440	25 degrees	0.68	4 cm	19cm
440	45 degrees	0.68	5 cm	24 cm

For AirMOSS, the presence of vegetation must be included in the radar scattering models since it is one of the most important factors in ensuring accurate retrievals of soil moisture. This is due to the double-bounce scattering mechanism (discussed later in the document) as well as the effects of vegetation canopy attenuation of radar signals. Furthermore, the long wavelength of P-band radar could carry important information about the subsurface and root-zone soil moisture, so that the radar scattering model and the retrieval algorithms based on the forward models should include the effects of the subsurface structure for most of the AirMOSS biomes. With the recent developments of comprehensive radar scattering models that include the effects of complex multispecies vegetation, subsurface soil moisture profiles, and multiple subsurface rough interfaces, the retrieval of root-zone soil moisture simultaneously with vegetation properties is possible. The challenge is to pose the retrieval (or estimation) problem in terms of as few unknowns as possible, because the number of measurements from the AirMOSS system is small and limited to the P-band polarimetric data set. To uniquely separate vegetation and subsurface moisture profile effects in the retrieval process, ideally many more measurements (e.g., multiple frequencies or multiple incidence angles) are needed. For the limited data set available from AirMOSS, ancillary data have been used to reduce the number of unknowns and especially the vegetation unknowns, so that the soil moisture profile can be uniquely determined.

2.2 The L2/3-RZSM Data Product Characteristics

This section provides a summary of the AirMOSS L2/3 RZSM product specifications.

Geophysical Parameters

The AirMOSS L2/3 RZSM baseline product includes root zone soil moisture (RZSM) for each 3-arcsec cell. Due to the wide diversity of vegetation covers present in the AirMOSS study sites (including the areas at the vicinity of flux towers), the retrieval algorithms and the specifics of the above parameter sets vary for different sites. In general, we categorize the sites into monospecies (including bare surfaces and grasslands) and mixed species. As such, Sites 1, 4, 5, 6, and 7 in Table 1 fall under the monospecies category, and Sites 2, 3, 8, and 9 fall under the mixed species category. Although, the soil moisture retrieval algorithm that produce L2/3-RZSM data are different, the products are similar. Soil moisture profile information is given in terms of second-order polynomial coefficient *a*, *b*, and *c* where $m_v(z) = az^2 + bz + c$, where *z* is the depth in meters

For the monospecies sites, Type A, species-specific allometric relations can be used to render the vegetation structure from a single kernel, for example, from diameter at breast height (DBH). The well-defined vegetation structure allows the use of the widely accepted discrete scatterer radar vegetation models. Along with scattering models of layered soil, it is then possible to calculate (or predict) the radar measurement corresponding to a given kernel and arbitrary soil moisture profile. The multichannel radar data set can then be used to estimate the unknowns in an iterative scheme based on the full scattering model. Extensive field campaigns are planned to collect ground truth data for vegetation structure throughout the AirMOSS radar swaths. To the extent possible, these data were used to parameterize the scattering models so that the vegetation unknown can be eliminated. This allowed more accurate retrieval of soil moisture profile.

For the mixed-species sites, Type B, since the vegetation cover is expected to be more complex than the monospecies sites and no longer describable by a single kernel, the retrieval burden is shifted towards an equivalent vegetation parameter such as biomass. Biomass is a compound quantity and does not preserve the information about vegetation geometry. For this reason, its relationship with the radar backscattering coefficient is non-unique unless constraints are placed on its value through the use of radar and ancillary data. Once use is made of these data to constrain biomass, the information effectively cannot be re-used to retrieve more unknowns within the soil. This, qualitatively, is why only one soil moisture parameter can be found from the type B algorithm. However, with some constraints and assumptions, as explained in more details in Section 4.2, the retrieval of soil moisture profile is possible.

Spatial Resolution, Posting, and Coverage

For each study site, the spatial resolution, posting and coverage of the L2/3-RZSM product was in accordance with the project-approved specifications as detailed in Shimada (2012). As such, the L2/3-RZSM products are posted at 3-arcsec spacing within each aggregate site coverage. The “aggregate site” is the union of all of the flight lines covering each study site and is approximately 25 km wide and 100 km long. Each aggregate site is typically built from 4 individual and parallel flight lines with regions of overlap between the adjacent parallel lines.

The overlap areas can be investigated for potential improvements in retrievals of RZSM due to the diversity of incidence angles. There is, however, only one soil moisture profile reported for each 3-arcsec pixel.

The L1-S0 product includes the radar backscattering cross section measurements on 0.5-arcsec and 3-arcsec ground-projected grids. The L2/3-RZSM product have been produced using the 3 arcsec L1-S0 data and reported on the same ground-projected grid. The AirMOSS radar has considerable flexibility in the center frequency and bandwidth with which it can operate. However, authorization was required from the National Telecommunications and Information Agency (NTIA) in order for the radar to transmit, and the NTIA limited the AirMOSS radar to operating between 420 and 440 MHz over the life of the AirMOSS program. The L2/3-RZSM products have been produced on a 3-arcsec grid.

Temporal Resolution and Sampling

The L2/3-RZSM product exists only for the overflight times of the AirMOSS radar. The timing of the flights was as follows, and was designed to best capture the inter-seasonal and inter-annual variations of root zone soil moisture:

- Complete three 10–14 day campaigns during the growing season in a 12-month period for each of the biomes 1–5 in Table 1, one campaign during each of the dry and wet seasons for biomes 6 and 7 in Table 1, and one campaign for biomes 8 and 9 in Table 1.
- Repeat the temporal and spatial sampling scenario above during three consecutive 12-month periods.

Latency

The L2/3-RZSM products were to be delivered within 90 days of the acquisition of the corresponding radar data. Since the latency of the L1-S0 product is 45 days after acquisition, this could have resulted in as few as 45 days of latency after the availability of the L1-S0 product.

Excluded Cells

RZSM was not retrieved for every pixel in the approximately 25 km by 100 km area imaged by the radar at each site. In particular, RZSM retrievals were not done over areas classified in the National Land Cover Database (NLCD) as: Open water, Perennial ice/snow, Developed—open space, Developed—low intensity, Developed—medium intensity, Developed—high intensity, Woody wetlands, and Herbaceous wetlands. In general, RZSM retrievals were done over areas classified as: Barren land, Deciduous forest, Evergreen forest, Mixed forest, Shrub/scrub, Grassland/herbaceous, Hay/pasture, and Cultivated crops. Similarly, RZSM retrievals were not done in areas with too high (greater than 2%) developed imperviousness in the NLCD. Fig. 1 shows a summary image of the NLCD and the corresponding legend.

For the Monospecies sites, the forward model is customized on a pixel by pixel basis to reflect that pixel *a priori* soil information, land cover classification, etc. Just as RZSM was not retrieved for cells with certain land cover types, RZSM was not retrieved for cells with certain local incidence angles, terrain slopes, and soil types. The details of excluded cells will be stated in Section 4.1.3.

For mixed-species sites, we performed soil moisture retrieval by treating data differently for forested pixels compared with bare soil pixels. The classification of forested and bare soil pixels was performed using NLCD with additional information from the radar data itself. In forested pixels, we used a scattering model which a representative of the forest behavior as a whole for that type of the forest. The bare soil pixels use retrieval derived from scattering from bare soil surface. Some details are given in Section 4.2. Similar to the monospecies sites, there are certain conditions under which we did not perform the retrieval.

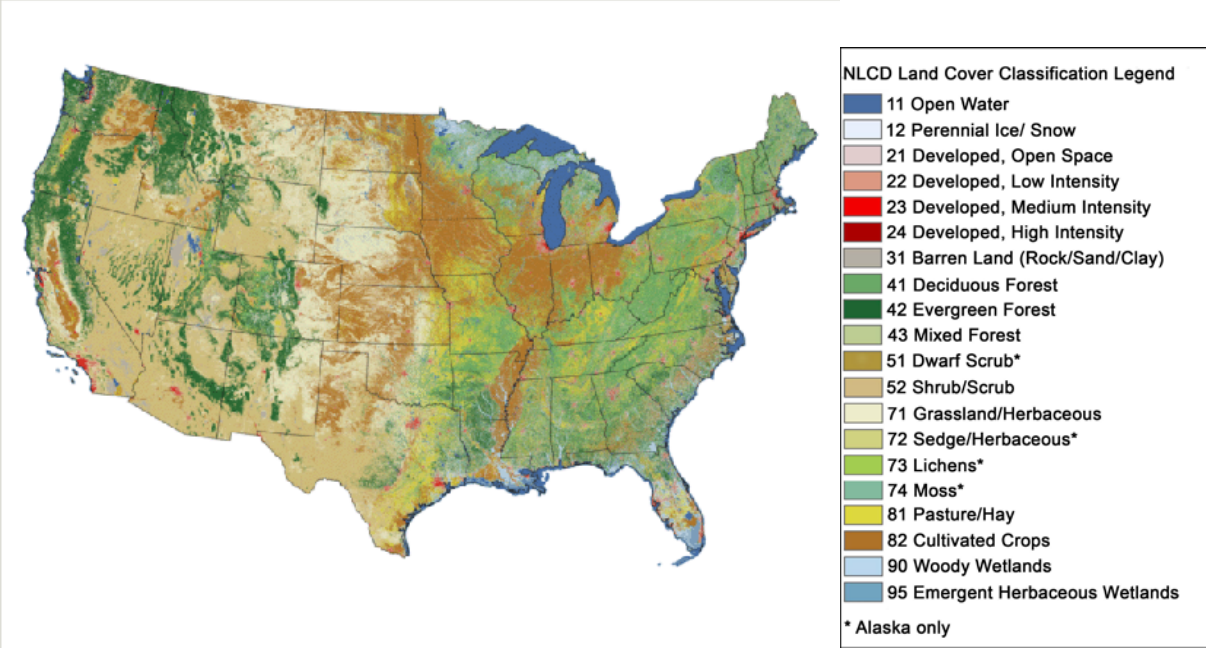

Fig. 1. National Land Cover Database (NLCD) summary image showing the land cover of the conterminous United States

For the mixed-species sites, a single forward model was developed for each site with the vegetation that is characteristic for that site. The retrieval for all pixels in a given approximately 25 km by 100 km site is based on that single, forward model independent of the NLCD classification of that pixel. Pixels that correspond to classifications other than the dominant classification for the site is marked in the data quality flag layer as having potentially lower accuracy.

Error Estimates

The baseline approach for assessing the retrieval errors of the L2/3-RZSM products is to compare them against the permanent in-situ soil moisture sensor profiles installed (or otherwise

available) at the AirMOSS flight sites. The error is calculated as a root mean square (RMS) measure, along with the calculation of a bias. The bias is removed for each biome on a yearly basis.

At each AirMOSS flux tower site, three permanent soil moisture sensor profiles have been installed. The area spanned by these stations is no more than 100m × 100m, and is therefore contained within one L2/3-RZSM product pixel. The total number of validation data points can be calculated as the product of the number of sites (9), the number of campaigns per year (1–3), the number of flights per campaign (3–4), and the number of years (3).

Stacking

Earlier in this section, the geophysical parameters for each 3-arcsec cell were listed. It should be pointed out that the RZSM and the soil moisture profile parameters, are unique for each "aggregate site" and the day that the radar imaged the site. Other input data layers to the RZSM retrieval, such as the soil surface roughness and the vegetation parameter, may not be unique day to day. For example, suppose that Tonzi Ranch is flown on November 1, that there is a light rain overnight at Tonzi Ranch, and that Tonzi Ranch is reflown on November 2. An assumption may be made that the soil surface roughness or the vegetation parameter did not change significantly between one day and the next. If applicable, these sorts of assumptions are very desirable since they reduce the number of parameters to estimate, increasing the fidelity of the RZSM retrieval for November 1 and November 2. Whether or not data can be stacked and the nature of the stacking would depend critically on the geophysical processes active between data collections such as the growth of leaves or heavy rain causing possible change in the soil roughness.

3 PHYSICS OF THE SCATTERING PROBLEM

3.1 Vegetation Model

For the purpose of calculating radar scattering coefficients, vegetated terrain with a single species of vegetation is typically modeled as a two-layer discrete random medium situated on a continuous random rough surface representing the ground. Fig. 2 depicts the typical traditional electromagnetic wave scattering geometry. The waves are incident at an arbitrary angle with respect to the vertical. The vegetation consists of two layers, one representing woody stems (tree trunks, for example) and modeled as nearly vertical cylinders, and another representing the crown layer and modeled as randomly oriented smaller cylinders, perhaps with a bi-modal size distribution, and containing small disks or needles to represent foliage. For short vegetation or crops, the depth of the woody stem layer can be set to zero. The crop or grass geometry can then be described in a similar fashion as with the crown layer, using appropriate random collections of cylindrical scatterers. For some crop types such as corn, the highly oriented stalks warrant the scattering medium to be represented as a uniaxial crystal, resulting in large differences between polarization responses.

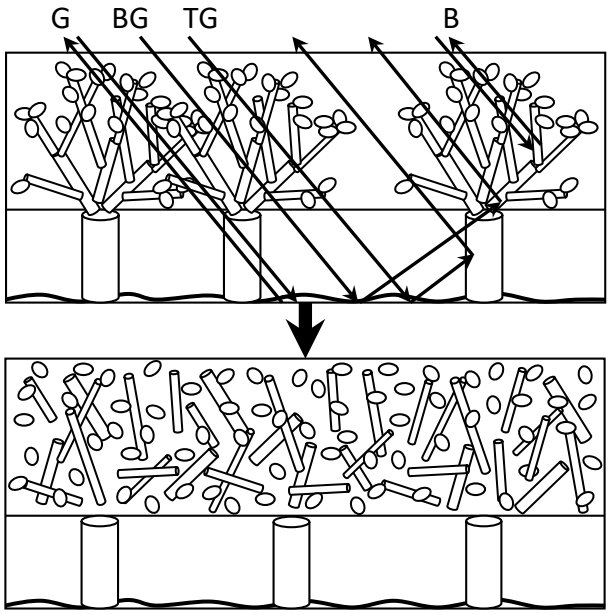

Fig. 2. Generalized geometry of electromagnetic scattering from vegetated terrain

The signal can follow a number of paths as it interacts with vegetation and ground and before it is received back at the radar. The figure shows the paths, which include (1) direct backscatter from the crown (branch) layer, (2) direct backscatter from the ground, (3) double-bounce scattering between branches and ground or the opposite path (B-G and G-B), and (4) double-bounce scattering between stems and ground or the opposite path G-T and T-G), In this context, ground refers to both the air/soil surface boundary and to the soil below it. In reality, there are also other paths through which the signal travels before returning to the radar receiver, which include multiple scattering within the crown layer, multiple scattering within the stem layer, and multiple scattering between the stem and crown layers. Higher-order contributions include such mechanisms as crown-ground-crown and stem-ground-stem, and so on, and are most evident in cross-polarized scattering. In general, however, interactions besides the main four listed above are usually quite small. The total backscattering coefficient is calculated by summing the contributions from each of these mechanisms. In both of the layers, the cylinders are of finite length, and could be of varying size scales compared to the wavelength. Each vegetation layer both scatters and attenuates the signals. Generally speaking, longer wavelengths are attenuated less than the shorter wavelengths, and may also be scattered to a smaller degree from the vegetation volume if they are significantly larger than vegetation components.

Traditionally, soil has been treated as a single half space. For lower frequency radars, subsurface structure and moisture of soil must be considered. All previously existing forest scattering models make this assumption (Durden *et al.*, 1989; Chauhan and Lang, 1991; Wang *et al.*, 1993; Ulaby *et al.*, 1990; Sun and Rasnon, 1995; many others).

There are two categories of approaches for calculating the radar backscatter from vegetation. Both approaches use essentially the same scattering geometry and vegetation model as shown in Fig. 2. A detailed comparison of the two approaches is given in Saatchi and McDonald (1997).

The first approach uses vector radiative transfer theory (RT) to formulate solutions for the radar backscattering coefficients (Ulaby *et al.*, 1990; Ishimaru, 1997; Tsang *et al.*, 1985). The RT theory formulates the transport of energy through a discrete random medium whose particles can scatter, absorb, or emit radiation. Since RT tracks the transport of energy and not the propagation of wave fields, it cannot properly account for coherent wave effects. Coherent effects are due to the superposition of numerous specular reflections, which contribute to the total scattered power. The RT-based models, therefore, tend to underestimate the total radar backscatter by an amount proportional to the coherent wave contribution. The coherent effects are prominent in the specular interactions of crown-ground and trunk-ground.

The second approach starts from Maxwell's equations and proceeds to derive the scattering matrix for each of the random media depicted in Fig. 3. First, the scattering matrix of a single finite cylinder of arbitrary orientation is derived based on vector scattering theory. The scattering from the ensemble is then calculated by integrating over an arbitrary, but realistic, probability density function for the size and orientation of scatterers within a given volume of vegetation. The interactions between vegetation layers can be accounted for by cascading the scattering matrices of the relevant layers. Additionally, attenuation due to each layer is calculated from the

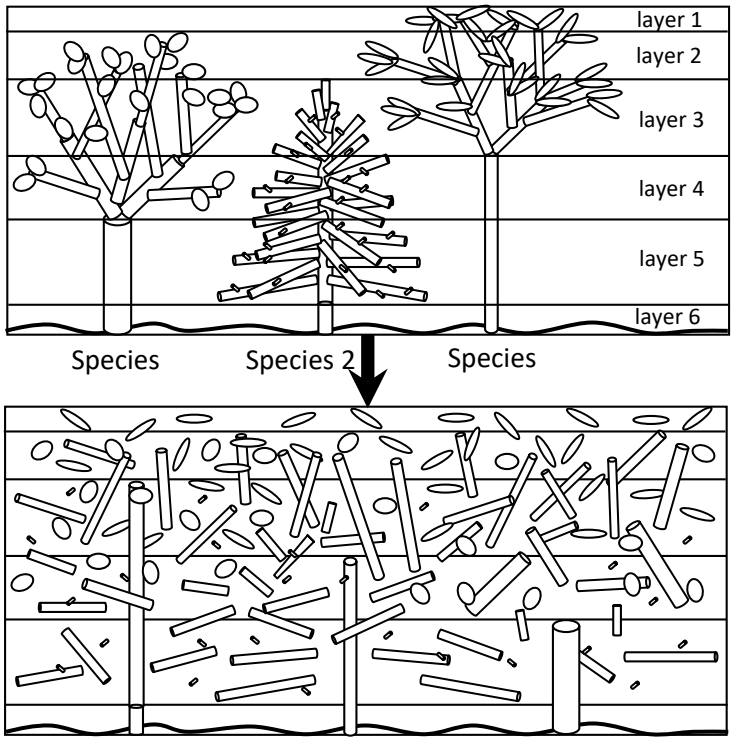

Fig. 3. Multilayer vegetation model geometry used in Burgin *et al.* (2011)

forward scattering matrix, and is included in the calculation of the backscattered signal from the layer below it. The inclusion of the layer attenuation is also referred to as the distorted Born approximation, since it modifies the single-scattering (Born approximation) solution by allowing the waves to travel in an equivalent, lossy medium, which is more realistic than traveling in free pace. To make the computations more efficient and feasible, the scattering matrix of each layer is used to first calculate the scattered power by forming the Mueller matrix, then summing the Mueller matrices of the layers and accounting for propagation through the preceding layers.

Among these models are those by Durden *et al.* (1989), Chauhan and Lang (1991), Wang *et al.* (1993), Lin and Sarabandi (1999), Israelsson *et al.* (2000), Picard *et al.* (2003), and Thirion *et al.* (2006). Although the two-layer vegetation is the most widely used model, most real landscapes have vegetation that is of either mixed species or the same species at different stages of growth. Therefore, there is a need to model multilayer vegetation. Recently, two such models have been developed, one which uses the vector RT formulation (Liang *et al.*, 2005) and the more recent one based on wave theory (Burgin *et al.*, 2011). The scattering geometry is shown in Fig. 3, where it is seen that an arbitrarily large number of layers can be considered, each with a distinct distribution function for the scatterers. Here we focus on the model based on wave theory, since it is expected to be more accurate in its prediction of coherent wave contributions. Comparison with actual SAR data from the JPL airborne SAR (AIRSAR) at a forested site in Australia, with abundance of ground reference data, suggests that this more realistic forest model can predict the radar backscatter more accurately than the traditional two-layer model.

3.2 Ground Surface Scattering Model

An integral part of a forest scattering model is the model for the underlying ground. As seen from Fig. 2, there are several ways in which the scattering from ground contributes to the total radar backscatter: direct ground, crown-ground, and stem-ground. A random rough surface can be defined in terms of its root-mean-square (RMS) height, the correlation length, and its permittivity. Various statistical representations have been suggested for typical, naturally occurring rough surfaces, which include Gaussian and exponential distributions. If there are multiple crown layers, each interacts with the ground and adds to the total backscattered signal.

Since exact solutions for scattering from general random rough surfaces do not exist, approximate analytical solutions are often used to account for the direct ground scattering term. These solutions are typically derived for the small roughness and/or small slope regimes, and include the small perturbation method (SPM), the Kirchhoff approximation (KA), and the geometric optics approximation, which is closely related to the KA. These techniques have been well developed and are treated in depth in several textbooks (Kong, 2000; Tsang *et al.*, 1985; Ishimaru, 1997). For the crown-ground and the trunk-ground double-bounce terms, the roughness effect of ground is captured via an exponential correction term to the Fresnel specular reflection coefficient (Ishimaru, 1997). To the best of our knowledge, the RMS surface roughness at AirMOSS sites fall within the limits of small roughness (or “smoothness”, Table 3) for P-band frequencies.

There are several competing factors for vegetated surfaces: For higher radar frequencies and in the presence of substantial vegetation, the contribution of the ground surface (direct or double-bounce) is not as marked as that for the lower frequencies, since the canopy attenuation masks the backscattered waves. For sparse or no vegetation, on the other hand, ground scattering becomes quite important at higher frequencies where the same surfaces may not be considered only slightly rough. At lower frequencies, where canopy attenuation is much smaller even if the vegetation is quite dense, ground double-bounce scattering is an important (often dominant) effect. On the other hand, direct ground scattering is less important due to the lower backscattering coefficients at lower frequencies if the ground is assumed to be only slightly rough (See Table 3).

When the roughness of the underlying ground surface cannot be regarded as only slightly rough, analytical methods listed previously no longer hold, and semi-analytical or numerical techniques have to be enlisted. These techniques rely on direct solutions of Maxwell's equations. A recent popular method has been the Integral Equation Method (IEM), initially proposed by Fung (1992) and later extended and improved in several ways (e.g., Shi *et al.*, 1997). Another, more accurate (but also more computationally complex) method is where the rough surface is randomly and numerically generated, then discretized into fine samples, and the total solution is found by Monte Carlo simulations of the scattered waves for a large number of surface realizations. Method of Moments (MoM) is the most widely used frequency domain technique for this purpose (Tsang *et al.*, 1994; Kapp and Brown, 1996; many others), along with various acceleration techniques used to reduce the computational complexity of MoM (e.g., El-Shenawee *et al.*, 2001; Moss *et al.*, 2006). In both of these classes of approaches, a major concern is error induced by truncating the numerical simulation grid. As the numerical computation domain is increased, the computational costs increase. Therefore, there is an inherent tradeoff between computational costs and accuracy for these solutions.

3.3 Subsurface Scattering Model

Due to the potentially large penetration depth of P-band signals, the waves can travel well inside the vegetation layer and into the ground surface, then scatter from subsurface layers, even for dense forests. For AirMOSS, it is therefore necessary that the radar scattering model also capture these effects. There are two scattering contributions to consider: (1) direct ground return, and (2) double-bounce scattering between trunks or crown and ground.

The direct backscatter from ground with subsurface layers can be treated with both analytical and numerical techniques, much like the single-layer ground. Naturally, the complexity of the solution is increased for the multilayer ground. For this reason and due to the novelty of the problem, there are far fewer published techniques available on this topic. On the analytical side, limiting assumptions have been made such as a single slightly rough interface on top of or embedded in a layered medium (Fuks and Voronovich, 2000). Two ways of simplifying the analysis have been by ignoring multiple scattering between the rough boundaries (Nghiem *et al.*,

1995) and by using the reduced Rayleigh equations to eliminate scattered fields inside the layered medium in the case of two independent rough boundaries (Soubret *et al.*, 2001).

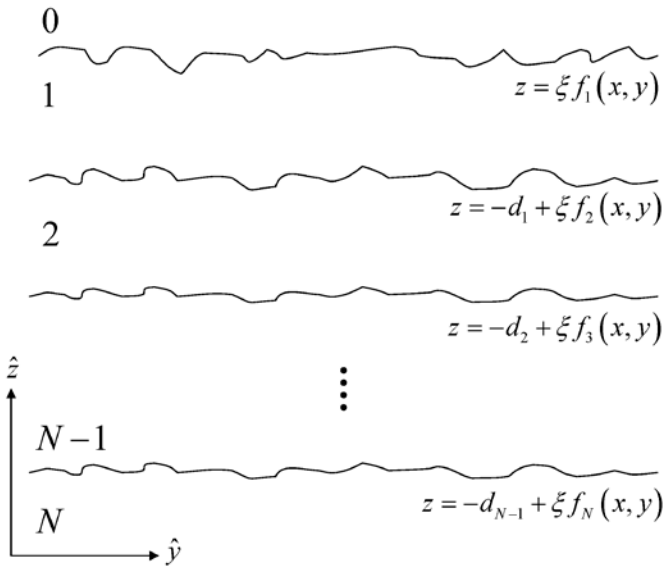

Fig. 4. Geometry of scattering from multilayer ground with multiple rough interfaces

A recent solution based on SPM offers a desirable mix of accuracy and computational efficiency for low-frequency radars (Tabatabaeenejad and Moghaddam, 2006). Fields in each region are represented as the summation of up-going and down-going waves, with their amplitudes found by simultaneously matching the boundary conditions. The boundary conditions are imposed up to the first order, as in any other first-order SPM. However, the resulting equations are solved in the far field without any further approximations. Consequently, this method intrinsically takes into account multiple scattering processes between the boundaries, all of which are considered rough simultaneously. This is a distinguishing factor from the previous methods in that it can be extended to higher orders, because it does not rely on the assumption that each rough boundary contributes to the solution independently of the other boundaries, which is only true in the first-order solution. This technique has been applied to an arbitrary number of rough layers, and can be used as well when each of the layers contains an arbitrary soil moisture profile. Fig. 4 shows a general *N*-layer geometry assumed for the SPM simulations in AirMOSS retrievals.

To account for coherent interactions between the ground and either the trunk or the branch layers (Fig. 5), a model has been recently developed to calculate the specular scattering from a two-layer rough surface within the Kirchhoff (small slope) regime (Tabatabaeenejad and Moghaddam, 2013). The model results in a convenient and compact expression for this scattering mechanism, which resembles the Fresnel refection coefficient of a layered structure, but with several terms modifying it to account for the roughness of interfaces. The coherent interaction between vegetation and layered ground can be shown to be the dominant scattering mechanism in many cases, especially when tall trees are present. The extensions of this model to

include a larger number of layers as well as smooth profiles within each layer are also needed and will be developed in the near future.

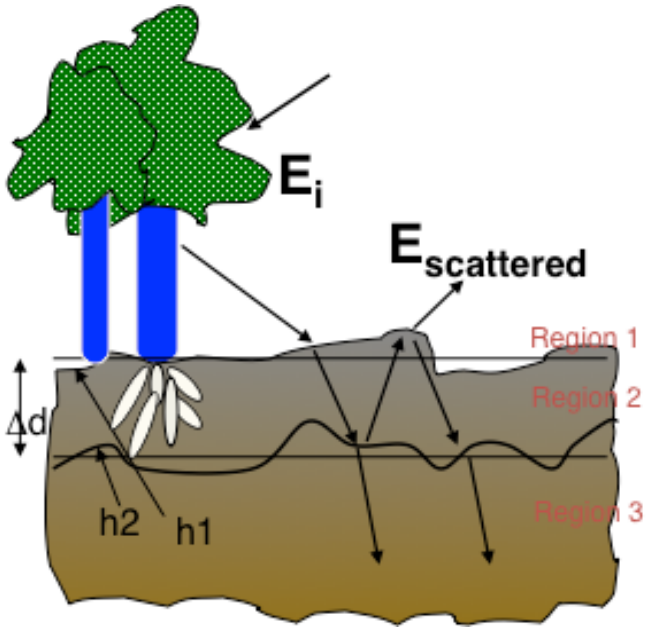

Fig. 5. Coherent scattering mechanism between vegetation and layered rough ground

Both the SPM-based and KA-based models are analytical and can readily be calculated. Their negligible computational cost is essential in the inversion process, because the forward model can be calculated thousands of times in the inversion algorithm.

The numerical models for treating subsurface ground layers are also few. Their advantage over the analytical models is that they are not, in principle, restricted to the small roughness or small slope regimes, though computational complexity and accuracy (due to limited computational domain) may ultimately limit their utility. Moss *et al.* (2006) and Demir and Johnson (2012) have developed MoM-based techniques to solve the two-layer rough surface problem using tapered wave illumination. Due to the added edge effects from the second layer and the interaction between the layers, the length of the simulated surface for the two-layer problem has to be larger than that of the single surface, on the order of 50λ. Another more recent solution is based on the extended boundary condition method (EBCM) and scattering matrix technique (Kuo and Moghaddam, 2007), and the much improved stabilized 3D EBCM (Duan and Moghaddam, 2012; Duan and Moghaddam, 2013), which can be used to solve an arbitrary number of rough layers and permittivity (moisture) profiles between each two rough interfaces. The reflection and transmission matrices of rough interfaces are constructed using EBCM. The permittivity profiles are modeled as stacks of thin dielectric layers. The interactions between the rough interfaces and stratified dielectric profiles are taken into account by applying the generalized scattering matrix technique. This method could be considered advantageous to MoM due to its numerical efficiency and that it does not rely on the tapered illumination assumption.

3.4 Validation of Models

The components of the baseline scattering model to be used for the AirMOSS L2/3-RZSM retrievals for ***mono-species vegetation*** are listed below. These components were chosen based on their accuracy and computational efficiency. The accuracy of each model component has been verified through (1) comparison with various other theoretical and computational models, and/or (2) comparison with measured radar data both in the absence and in the presence of vegetation. For each model component, references are cited that include extensive examples to demonstrate the validity of the chosen models and the choice of model parameters from field data and other ancillary information.

Vegetation: Wave-theory based discrete random medium model (Durden *et al.*, 1989, Burgin *et al.*, 2011); even though the model of Burgin *et al.* (2011) is a general multispecies model, we only used it assuming a single-species pixel for reasons that are explained in the next section on retrieval algorithms.

Ground Surface and Subsurface: Multilayer SPM (Tabatabaeenejad and Moghaddam, 2006) was used for direct scattering contribution from ground, and the layered rough surface Kirchhoff model (Tabatabaeenejad and Moghaddam, 2013) was used for the coherent double-bounce contribution between vegetation and ground.

The baseline scattering model to be used for the AirMOSS L2/3-RZSM retrievals for ***mixed-species vegetation*** is explained in 4.2 below. The details of the distorted Born model, which is used to construct the simplified model is given in Saatchi *et al.* (1997).

4 L2/3-RZSM RETRIEVAL ALGORITHM

4.1 Baseline Algorithm for Monospecies Vegetation

4.1.1 Algorithm Overview

The retrieval algorithm can be summarized as finding the RZSM and any other relevant soil and vegetation parameters that minimize a cost function– or a misfit function – representing the difference between measured radar data and the scattering model. To achieve accurate retrievals, we must have (1) well calibrated radar data, (2) accurate radar scattering models including accurate parameterization from ancillary data, (3) a powerful inversion algorithm and (4) proper choice of unknowns. Item (1) is discussed in the AirMOSS Calibration and Validation Plan. Item (2) was discussed above in Section 3. We focus on items (3) and (4) below.

Radars are primarily sensitive to the strength and geometric distribution of dielectric properties of the targets in a scene. Dielectric constants are strong functions of their water content. The relationship between dielectric constant and soil moisture, for example, has been studied extensively and determined for various frequencies and soil textures (e.g., Dobson *et al.*, 1985; Peplinski *et al.*, 1995a and 1995b). Likewise, the dielectric constant of vegetation has been shown to be a strong function of its moisture content (El-Rayes and Ulaby, 1987). The relation

between microwave measurements and dielectric constant and soil and vegetation geometry can be studied via the radiative transfer or wave scattering models discussed above. In the forward mode, these models predict the radar backscatter measurements given the scene properties, and in the inverse mode (or retrieval mode), they are used to estimate the scene properties from radar scattering measurements.

Many retrieval studies have been carried out using non-unique vegetation parameters, such as the physical parameters of a layer of random spheroids such as small needle-like or disk-like particles (Xu and Jin, 2006), vegetation water content (VWC) (Notarnicola and Posa, 2007), or biomass (Saatchi and Moghaddam, 2000), to characterize the vegetation cover of a forested area. These parameters are non-unique in the sense that a given set of these parameters could correspond to a large range of vegetation structural characteristics, which can produce vastly different radar scattering coefficients. For monospecies vegetation cover, it is possible to remove this ambiguity and still keep only a small number of vegetation unknowns through the use of allometric relations. Allometric relations describe the relationship between different vegetation parameters, such as between trunk diameter at breast height (DBH) and tree height or crown size or branch density. As such, it may be possible to describe a tree only through one vegetation parameters, which we will refer to as the vegetation "kernel." Another parameter for vegetation is the density of stems, which can be assigned either as an independent unknown, or known from ground observations or the so-called self-thinning relationships that relate density of trees trunks to their size. Since we only plan to use at most three independent measurements from the AirMOSS radar (HH, VV, HV scattering coefficients), we opt not to book-keep density as an independent unknown. Instead, the assumption is that it is known from ground observations or other ancillary information.

For subsurface soil moisture, several options may be examined to determine the most appropriate unknowns. For monospecies vegetation as well as non-vegetated ground, our simulations for P-band have shown that simplifying the ground geometry to a single half space, whereas in reality there might be a variable depth profile, leads to highly erroneous retrieval results: The half-space retrieval does not represent the average soil moisture within the profile, and its correlation to a particular weighted average of the profile is also not obvious. The results become increasingly more accurate with the inclusion of a larger number of layers in the subsurface. We have shown that, when applicable, representing a moisture profile (of a forested area) with a polynomial and inverting the (unknown) coefficients of the polynomial reduces the number of unknowns, which would otherwise be the moisture contents of all the layers included in the inversion (Tabatabaeenejad *et al.*, 2012b). For example, Fig. 6 shows the measured moisture profiles and their second- and third-order polynomial fits at locations close to the flux towers in (a) BOREAS Old Jack Pine from July 29, 1994 and (b) Tonzi Ranch from January 8, 2012. The RMSEs show that a second-order fit meets the AirMOSS error criterion of an absolute RMSE of less than 0.05 m^3/m^3.

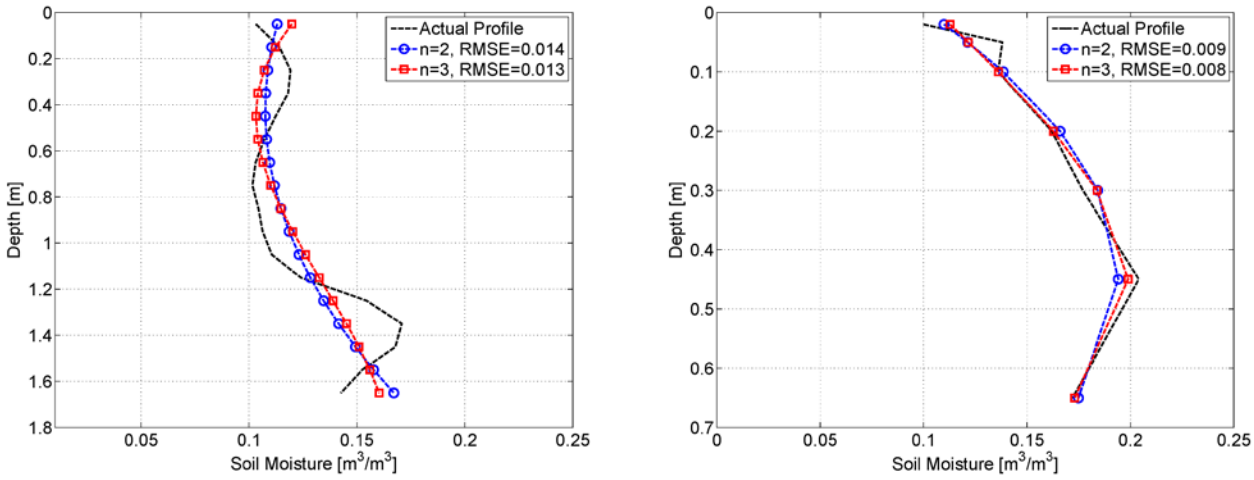

Fig. 6: Measured moisture profiles and their second- and third-order polynomial fits at locations close to the flux towers in BOREAS Old Jack Pine from July 29, 1994 (Left) and Tonzi Ranch from January 8, 2012 (Right).

AirMOSS used a 20 MHz bandwidth of 420–440 MHz and only 2 co-polarized channels. With this single-band assumption, the retrieval process cannot produce accurate results if the number of unknowns is more than 3. This number can be allocated to both vegetation and soil. For example, it can be allocated to one vegetation unknown (the kernel) and two soil unknowns (surface roughness and one half-space effective moisture content). It can also be allocated to three soil unknowns (surface roughness and coefficients of soil profile function), assuming the vegetation properties are known through land cover maps and/or ground observations by the AirMOSS team. The latter is the approach we took in the implementation of the baseline algorithm. Fig. 7 depicts the basic geometry of the radar scattering model used for monospecies biomes of AirMOSS.

4.1.2 Mathematical Description of the Algorithm

We pose the inversion as an optimization problem and use simulated annealing to minimize a cost function that is based on the difference between measured and calculated backscattering coefficients at the incidence angle and measurement frequency. The cost function (or the misfit function) to minimize for a given cell is given as

$$L(\mathbf{X}) = \sum_{pq} \left(\frac{10 \log \sigma^{o}_{pq}(\mathbf{X}; f, \theta) - 10 \log d_{pq}(f, \theta)}{10 \log d_{pq}(f, \theta)} \right)^2$$

where σ^{o}_{pq} and d_{pq} are, respectively, the calculated and measured backscattering coefficient of the forested area at the frequency f and observation angle θ for pq polarization. The quantities are expressed in decibels to make L more sensitive to changes in the model parameters during inversion. The vector, X, denotes the vector of unknown model parameters.

The baseline optimization algorithm is based on simulated annealing: a small randomly generated perturbation is applied to the current model parameters. The new parameters are then used to calculate a new estimate of the backscattering coefficients, hence a new value for the cost function. If the cost function decreases, i.e., $\Delta L < 0$, the new state is accepted, otherwise it is accepted with probability $\exp[-\Delta L/T]$, where T is an inversion parameter referred to as temperature. This rule, referred to as the Metropolis criterion, is used at a sequence of decreasing temperatures and each sequence of states at a constant temperature is referred to as one iteration. The simulated annealing algorithm we implement is based on the work by Corana *et al.* (Corana *et al.*, 1987).

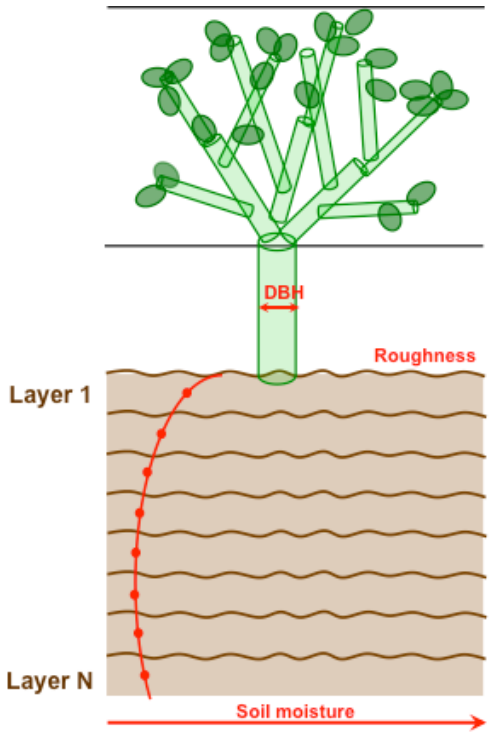

Fig. 7: Basic geometry implemented in the radar scattering model used for AirMOSS sites with monospecies vegetation

The algorithm starts from an initial guess X, an initial temperature, and an initial step length vector. A random move is generated sequentially along each coordinate direction. The trial point is either accepted or rejected according to the Metropolis criterion. This set of sequential perturbations is repeated N_s times. The step length is then adjusted according to the Corana*'s* algorithm step-length adjustment rule and the iteration continues until the number of step length adjustments reaches a pre-set number N_T. The temperature is reduced at this point by $T_{new} = R_T T_{old}$ and the iteration continues at the new temperature starting from the current optimal point. This procedure is illustrated in Fig. 8, where the current state at each step length adjustment is denoted by X_2, X_3..., X_{NT+1}. Note that the first step length adjustment in each chain happens at X_2 and the last one happens at X_{NT+1}, which is also the last accepted point of the corresponding

chain. The inversion process stops when the cost function value becomes smaller than a pre-set value, when the number of forward function evaluations reaches a certain number, or when the algorithm converges to local minima for a certain number of times. More details can be found in Tabatabaeenejad and Moghaddam (2009).

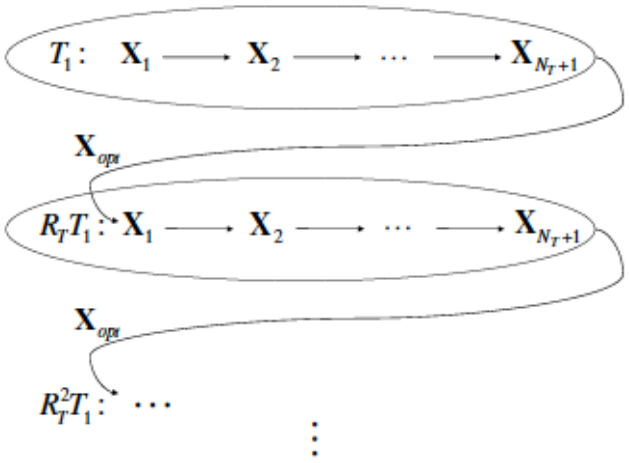

Fig. 8: Depiction of the steps involved in the simulated annealing method. The step length is adjusted NT times before the temperature is reduced and algorithm proceeds to the next chain.

Fig. 9 shows a sample result of the simulated annealing algorithm when applied to a two-layer rough surface structure with no vegetation on top. The soil properties are represented with the real part of the relative dielectric constants and the conductivity of the layers. The plots show the relative errors in the current value and optimal value—best value encountered so far in term of cost function value—of each unknown after each iteration of the simulated annealing algorithm.

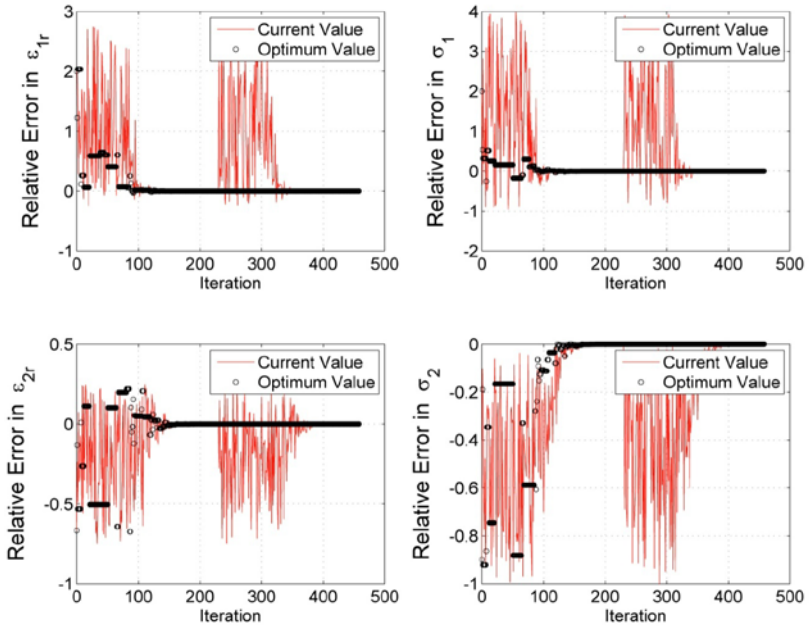

Fig. 9. Sample output of the simulated annealing algorithm. The plots show the relative errors in the current value and optimal value of the geometrical parameters of a two-layer rough surface structure as a function of number of iteration.

4.1.3 Input Data Parameters

Table 4 lists the required input data layers to the L2/3-RZSM retrieval algorithm for monospecies vegetation. These include the radar backscatter measurements and the ancillary data that describe the physical structures for each 3-arcsec pixel. The treatment and the required processing for each data layer are as follows.

a) Radar Backscatter
   The AirMOSS L1-S0 gridded radar backscatter data are the primary input to the L2/3-RZSM retrieval algorithm. In particular, only the HH and VV channels gridded on a 3 arcsec ground projected grid are used in the retrievals. The other data layers are also gridded on the same grid for the ease of subsequent processing and inversion processes.

b) Incidence Angle
   Due to the calibration accuracy of AirMOSS radar antenna radiation pattern, pixels with incidence angle outside the 25 to 50-degree range are excluded in the retrievals. Also, the RZSM are not retrieved for pixels for which the radar did not image well. No retrieval is performed for pixels for which radar data are absent due to shadow or layover

c) Terrain Slope
   Pixels that have excessive slopes are not included in RZSM retrievals. The criterion for excessive slope is the ones that are larger than 5 degrees. This estimation is done using the SRTM DEM, from which a co-registered slope data layer is produced as part of AirMOSS L1-S0 processing.

d) Land Cover Class
   The NLCD 2011 land cover class are only available for the conterminous United States, Alaska, Hawaii, and Puerto Rico. Consequently, they are not available at the BERMS site (Saskatchewan, Canada), and GlobCover 2009 data from the ESA GlobCover project are used instead. Land cover maps are first projected into 1-arcsec latitude/longitude grids from their original projections, and then upscaled to 3-arcsec grids based on the most dominant land cover class among the pixels being upscaled.

e) Vegetation Parameters
   Vegetation parameters such as vegetation structure and composition for a given land cover class are determined from the United States Department of Agriculture (USDA) Forest Inventory and Analysis (FIA) datasets, allometric relations, and filed measurements conducted during the course of AirMOSS flights. These parameters are assumed to be the same for the pixels with the same land cover class within a given site. Vegetation parameters of the same land cover class may differ from site to site due to the adjustment based on the field observations. The following information has been collected during the field campaign activities to enhance the model parameterizations and therefore

retrieval quality: vegetation geometric parameters such as height, DBH, branch lengths, branch densities, branch diameters, leaf properties, stem and branch dielectric constants. The specifics of data collection and sampling design during ground field campaigns are discussed in a separate document, "AirMOSS Field Campaign Description and Protocol."

f) Soil Texture
The USDA Soil Survey Geographic (SSURGO) Database is used for the sites within the United States and the Harmonized World Soil Database (HWSD) produced by the International Institute for Applied Systems Analysis (IIASA) is used for the BERMS site in Canada.

g) Soil Temperature
For each site, in-situ soil temperature measurements taken at the vicinity of the flux tower during the flight time are used as an input parameter to the soil dielectric model used in the forward model. The soil temperature is assumed to be the same throughout the site and this assumption does not impact the accuracy of the forward model since the mineral soil dielectric constant does not vary much with temperature.

Note that no meteorological data are used in the RZSM retrievals such as data from rain gauges. The RZSM retrievals are not constrained using meteorological data such as constraining that the RZSM must be higher for a scene after a rain than before it.

TABLE 4. Input data of AirMOSS L2/3-RZSM retrieval algorithm for monospecies vegetation

Data Type	**Data Source**
Radar Backscatter (HH/VV)	AirMOSS L1-S0 (3 arcsec)
Incidence Angle	AirMOSS L1-S0 (3 arcsec)
Terrain Slope	SRTM (Version 2)
Land Cover Class	NLCD (US) GlobCover (Canada)
Vegetation Parameters	USDA FIA Allometrics Field Measurements
Soil Texture	SSURGO (US) HWSD (Canada)
Soil Temperature	In-situ Measurements

4.2 Baseline Algorithm for Mixed-Species Vegetation

To retrieve root-zone soil moisture using backscattering coefficients data, we apply two-step approach. A flowchart of the entire process is illustrated in Fig. 10. Each individual pieces are explained in details in subsequent sections. The overall retrieval can be separated into two steps. The first step (Retv r1 and Retv r1p2) is to retrieve 'effective' dielectric constant of the ground, with the by-product of above ground biomass and ground roughness. The second step (Profile Retv) converts the effective dielectric constant of the ground into the soil moisture profile. The detailed description of these two step are given in 4.2.1 and 4.2.2.

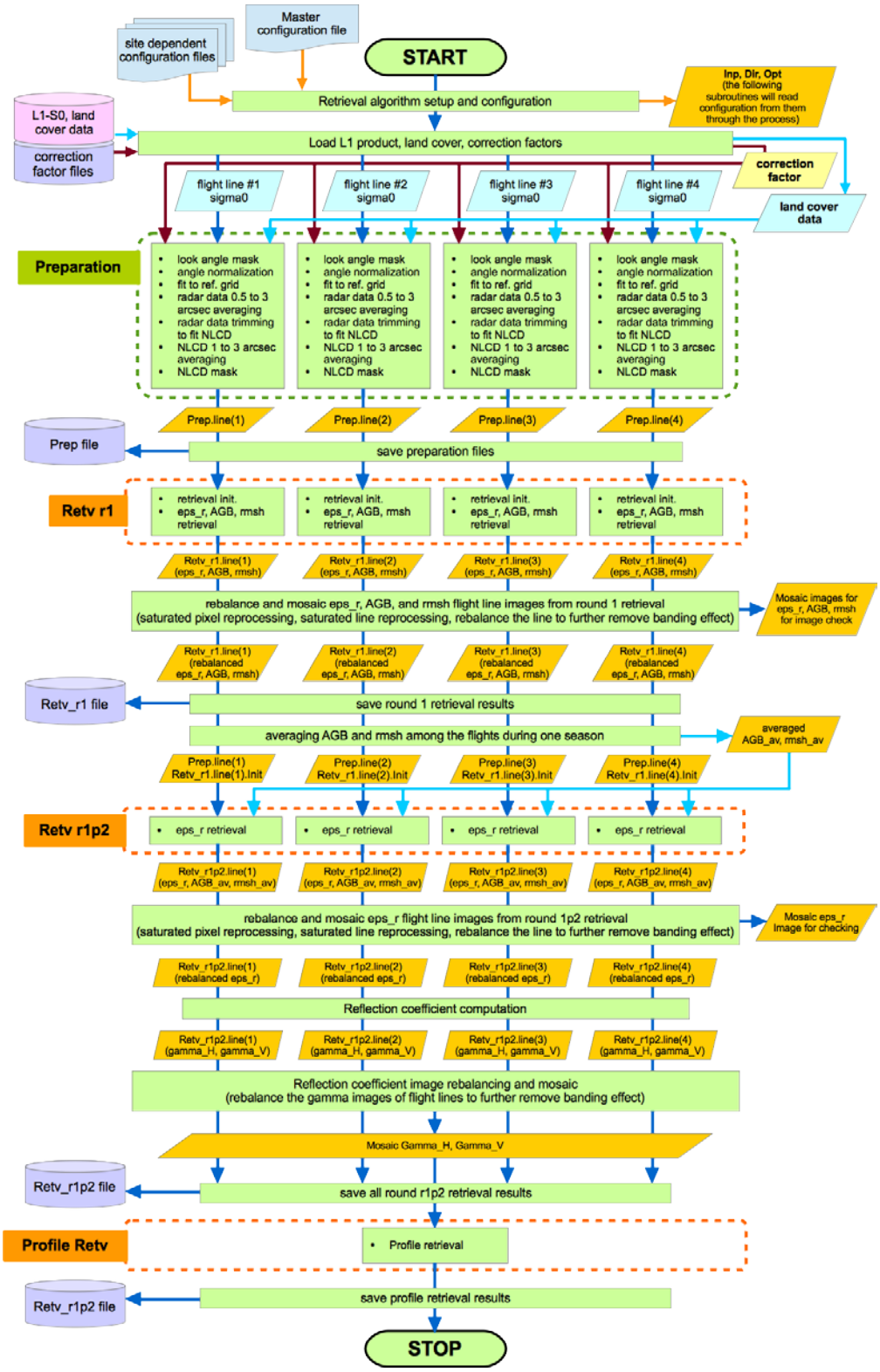

Fig. 10: The flowchart showing the whole process on soil moisture profile retrieval from Radar data

4.2.1 First step: Retrieval of dielectric constant of ground from radar data

In the mixed species vegetation scenarios, it is impractical to calculate contribution of scattering each individual tree species. Thus, we model the scattering mechanisms of the forest using physical based data-fitted model with three components, namely direct, direct-reflect, and ground components. The direct component consists of the scattering from trunks, branches, and leaves directly back to the sensor. The direct-reflect terms, also known as 'double bounce' term, is the scattering that includes interaction between trees and ground. This includes the trunk-ground, crown-ground scattering. The ground component is the scattering directly back from ground. The pictorial presentation of these components is illustrated in Fig. 11. The model can also be written in equation form as shown in Eq. 1. The detailed equation of the model is given by Saatchi et.al (Saatchi *et al.* 1997). This model which is represented by Eq. 1 is considered 'simplified forward' model

Notice that the backscattering coefficients depend not only on the dielectric constant of the ground (directly tied to soil moisture), which reside in the Fresnel reflection coefficients and S_{pq} (scattering from ground), but it also depends on the Above Ground Biomass (W) and the ground roughness (or RMS height denoted by s in the equation). The coefficients A, B, C and α, β, δ are fitted coefficients for particular type of forests.

$$\sigma_{pq} = direct + direct - reflect + ground \quad (1)$$

$$direct = \mathrm{A_{pq}} \mathrm{W}^{\alpha_{pq}} cos\theta \left(1 - \exp\left[-\frac{\mathrm{B_{pq}} \mathrm{W}^{\beta_{pq}}}{\cos\theta}\right]\right)$$

$$direct - reflect = C_{pq} W^{\delta_{pq}} \Gamma_{pq} sin\theta \left(exp\left[-\frac{B_{pq} W^{\beta_{pq}}}{cos\theta}\right]\right)$$

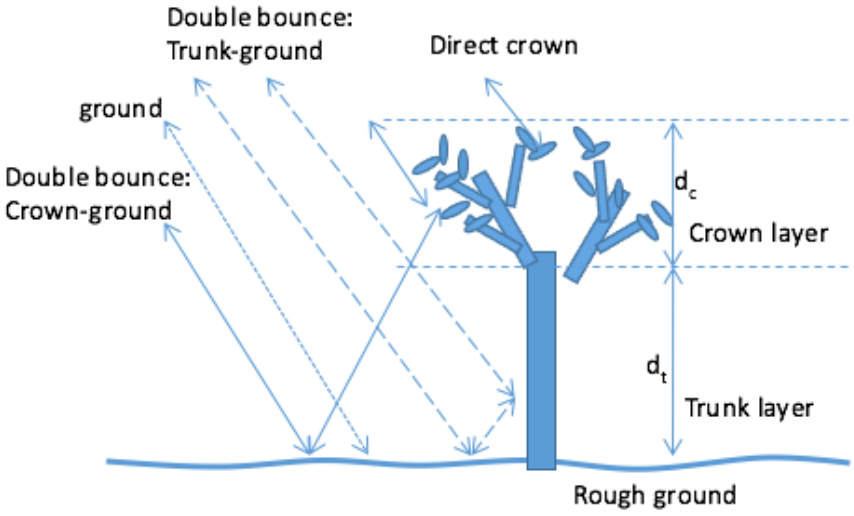

Fig. 11: Scattering mechanisms of forest scattering model

$$ground = S_{pq}\left(exp\left[-\frac{B_{pq}W^{\beta_{pq}}}{cos\theta}\right]\right)$$

where $\Gamma_{pq} = \left|R_{pq}\right|^2(exp[-4k^2s^2cos\theta^2])$, W is biomass (in Mg/ha), R_{pq} is the Fresnel reflection coefficient of the ground for polarization pq, pq denotes polarization which can be vv, hh, hv, or hv, s is the RMS height of the rough ground, k is the wave number, and S_{pq} is the scattering from the rough ground using Oh's model (Oh *et al*., 1992).

The coefficients α, β, δ are the 'shape' coefficients. They depend on the characteristics of the forests in terms of tree geometry such as trunk height and diameter, branch length and diameter and leaf size and density. On the other hand, coefficients A, B, C represent relative and absolute contribution of each of the three scattering mechanism. The Forward model coefficients are determined using previously obtained data that matches with ground observation/measurements.

Note that the S_{pq} is the contribution from rough surface scattering from the ground. We use the model provided by Oh (1992) which provide simple expression to evaluate as a function of soil dielectric constant and the RMS height of the rough ground. Other rough surface scattering models can also be used. Some examples are the Integral Equation Method (IEM) model, small perturbation method and Kirchhoff approximation. A few of these has been tested. However, we found that

(1) The contribution from rough surface scattering is comparatively smaller than the direct and direct-reflect scattering, particularly in the forested environments.
(2) Oh 92 model is preferred because of its simplicity and its expression that directly link the backscattering coefficients to the soil dielectric constant. His newer model (Oh 2004) expresses the backscattering coefficients in terms of soil moisture content, which is less useful in our case.
(3) The small perturbation method is also a possible alternative. However, we found that it usually underestimates the contribution from ground surface scattering. This might be because the surface roughness, which is about 2 cm is very small compared to the P-band wavelength.

Determining the coefficients for forward model

The intensive distorted Born approximation model (Saatchi *et al.*, 1997) is used to simulate the backscattering from the forest. The underline physical interpretation of scattering contributions from the forest is the same as the simplified forward model we use for retrieval. However, each of the components can be expand and written in more rigorous manner. For example, the direct terms consist of direct scattering from trunk, branch and leaves, instead of being lumped together into the first term of Eq. 1. Furthermore, there are a few more inputs to construct the model including the size distribution, density, and orientation of trunks, branch, and leaves. For the US site (Howland, Harvard, and Duke), there are tremendous amount of ground measurements for

the models. Therefore, we are able to construct and simulate backscattering coefficients with fewer assumption. The shape parameters are obtained from this simulation and the coefficients fit. Then, the backscattering coefficient of AirSAR and the ground measurements are used to determine the coefficient A, B and C as explained in more detail in Troung-Loi *et al.* (2015). See Table 5.

TABLE 5. Simplified forward model (Eq. 1) coefficients

Chamela forest

	A	B	C	α	β	δ
HH	0.17038117	0.0097499	0.015	0.1817	0.9727	1.28
VV	0.1	0.0092264265	0.032516427	0.1952	0.9921	1.361
HV	0.064323202	0.0094882129	0.001	0.2289	0.9827	1.49

Northeast US (Howland, Harvard, Duke forest)

	A	B	C	α	β	δ
HH	0.1	0.00767714	0.001403255	0.16351	0.95303	1.81032
VV	0.028704653	0.015	0.00239	0.21654	0.91264	1.9396
HV	0.0269	0.0023876037	0.0005	0.25673	0.932835	1.7513

La Selva forest

	A	B	C	α	β	δ
HH	0.0230638	0.00257578	0.00263325	0.3	1	1
VV	0.00971005	0.00429297	0.0034001	0.5	1	1
HV	0.00203221	0.00343438	$9.46966.10^{-5}$	0.5	1	1.5

Inversion approach: Levenberg-Marquardt non-linear least square method

In the inversion, we first initialize the three unknown (dielectric constant, above ground biomass, and surface roughness) To initialize the vegetation biomass value, we use an empirical model developed by Saatchi *et al.* (2011) and given by

$$\sqrt{W_0} = a_0 + a_1\sigma_{HH} + a_2\sigma_{HV} + a_3\sigma_{VV} \quad (2)$$

Where the coefficients a_0, a_1, a_2, a_3 are computed using limited ground estimates of biomass from forest inventory plots. This equation is then applied to the entire SAR image to get a biomass map. To initialize the two soil parameters, a segmentation of non-forested and forested areas is done. The soil moisture and roughness are estimated over bare surfaces. Their mean value is computed and allocated to forested areas. Initialization is now done for the three parameters and the inversion process can be run. The steps of this process are summarized in the diagram in Fig. 12.

TABLE 6. Coefficients for initializing biomass for retrieval

	a_0	a_1	a_2	a_3
Harvard, Howland, and Duke forest	2.33764	6.82745	110.726	-10.9808
La Selva forest	0.73	42.13	323.02	71.51

For Chamela forest, we use $W = 360.14\sigma_{HV}^{0.797}$ instead.

The inversion procedure uses the Levenberg-Marquardt non-linear least-squares method to estimate the structural parameters: vegetation biomass, dielectric constant and ground surface roughness. The Levenberg-Marquardt is a non-linear least squares curve fitting defined here by:

$$S(W, \varepsilon, s) = \sum_{i=1}^{n}\left[\sigma_{pq_i} - f(W, \varepsilon, s)\right]^2 \quad (3)$$

This procedure is based on the minimization of the distance between data and model computations given a priori estimate of the parameters and boundary conditions (Markwardt, 2009). This is performed in an iterative way, updating the solution at each iteration until a small error is reached. In this case we run the optimization process over the entire SAR image pixel by pixel. σ_{pq} is the backscattering value recorded by the SAR system at this particular pixel, f is the model equation, W, ε and s are the parameters to be optimized for this pixel. Since no textural soil data is available the empirical relationship of Topp (1980) is used to compute the soil moisture from the dielectric constant.

To avoid meaningless values, some limits are set for the three parameters:

$$\varepsilon_{min} < \varepsilon' < \varepsilon_{max}, \quad W_{min} < W' < W_{max}, \quad s_{min} < s' < s_{max}$$

And ε' and ε_0 are the estimated and initial dielectric constants respectively. *W'* and W_0 are the estimated and initial biomass values respectively. s' and s_0 are the estimated and initial ground surface roughness values respectively. These are parts of input parameters for retrieval algorithm and their values are listed in Table 6.

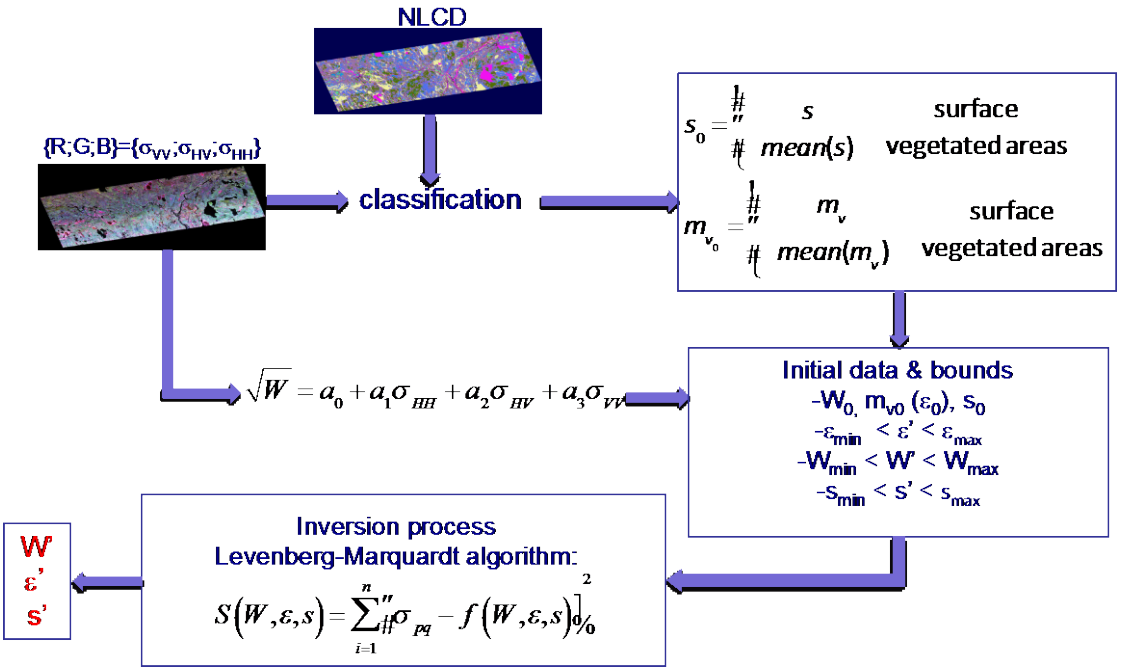

Fig. 12: Diagram summarizing the steps of the inversion process using a local optimization algorithm (Retv r1).

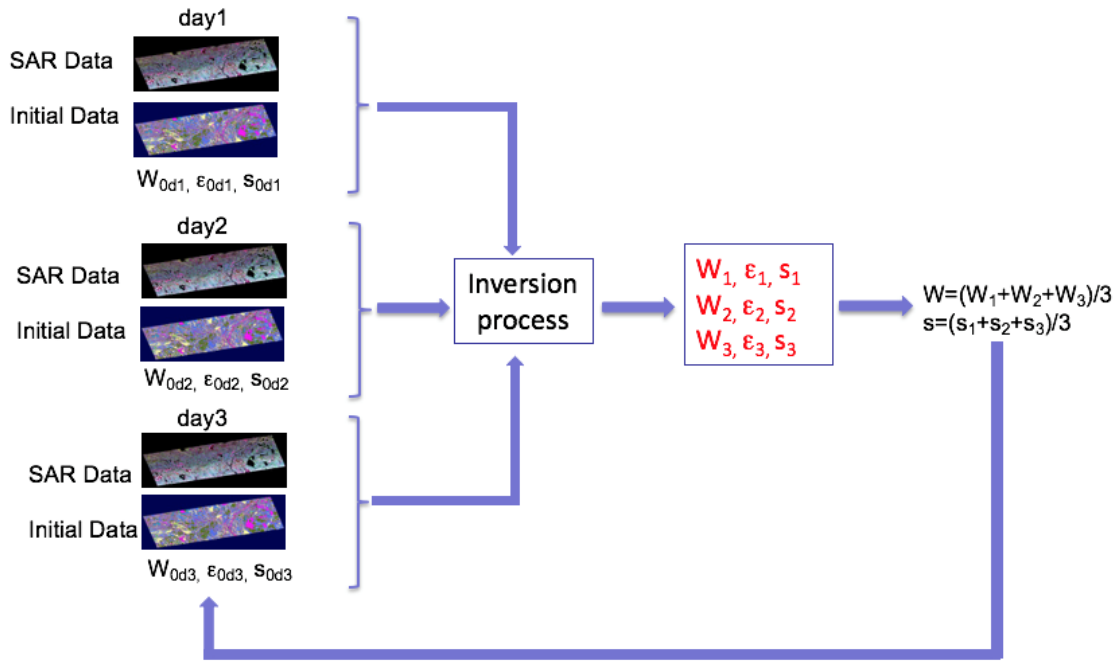

Fig. 13: Diagram for soil moisture retrieval Step 1.2 (Retv r1p2)

Improved dielectric constant retrieval: Step 1.2

Under the assumption that within about 10 days of radar data acquisition for each season, the above ground biomass and the surface roughness should be unchanged, we can further condition the retrieval by using average of above ground biomass and average of surface roughness from previous retrieval as shown in Fig. 13.

4.2.2 Second step: Soil profile retrieval using data cube method

The concept of the soil moisture profile retrieval is using the data cube approach where all the possible soil moisture profile is simulated and the observable is recorded. The selection of the result is based on picking the profile with minimize cost function. First, we explain the construction of the data cube and, later, we explain the retrieval process.

The data cube approach starts with the forward calculation of the observables (reflection coefficients and average dielectric constant) from the soil profiles. The reflection coefficients (v and h pols) are derived from coherent scattering using multi-layer transmission line model, illustrated in Fig. 14.

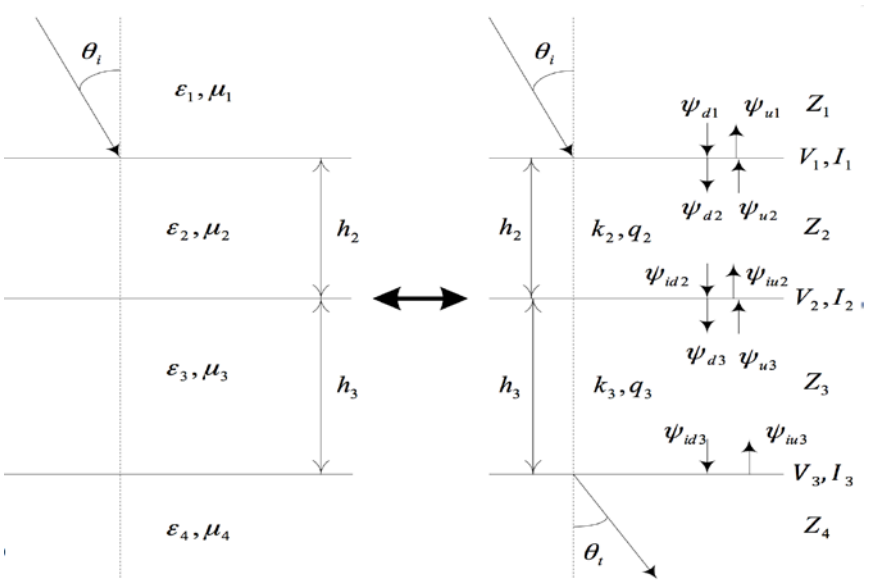

Fig. 14: Multi-layer transmission line model.

The soil layer is assumed to have planar interfaces with different clay fraction and soil moisture content in each layers. We calculate dielectric constant in each layer from the clay fraction and the soil moisture content using Mironov model (Mironov, 2009). Then, the Fresnel reflection coefficients R_s can be calculated by

$$R_S = \frac{A + \frac{B}{Z_N} - Z_1\left(C + \frac{D}{Z_N}\right)}{A + \frac{B}{Z_N} + Z_1\left(C + \frac{D}{Z_N}\right)}$$

where

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} A_3 & B_3 \\ C_3 & D_3 \end{bmatrix} \cdots \begin{bmatrix} A_{N-1} & B_{N-1} \\ C_{N-1} & D_{N-1} \end{bmatrix}$$

$$A_m = D_m = \cos q_m h_m \ , B_m = jZ_m \sin q_m h_m \ , \ C_m = \frac{j \sin q_m h_m}{Z_m}$$

$$Z_m = \frac{\omega \mu_m}{q_m} \ , q_m = \beta_m \cos\left(\theta_m\right) - j\alpha_m$$

for H-polarization

for V-polarization $Z_m = \dfrac{q_m}{\omega \mu_m} \ , q_m = \beta_m \cos\left(\theta_m\right) - j\alpha_m$

and the propagation constant in each layer is given by

$$\alpha_m + j\beta_m = j\frac{2\pi f}{c}\sqrt{\varepsilon_{ri}}$$

In the equation above, we write the expression based on four-layer medium (air, soil1, soil2, and soil3 where soil 3 extends to infinity). The expression can be generalized to *N*-layer in which the total ABCD matrix is the multiplication of ABCD matrix 1, ..., *N*-1. The incidence angle at the top layer is θ_i and the dielectric constant at each layer is ϵ_{ri}.

The reflection coefficients calculated for these formulation is the same as R_{pq} shown in Eq. 1. Next, the average soil moisture is calculated with the following formula

$$Mv_{avg} = \frac{\int M_v exp(-j\alpha z)dz}{\int dz}$$

This is soil moisture as seen by the radar with average weighted by attenuation α which is a function of depth. In each of the site, we gather all combination of soil's clay profile provided The USDA Soil Survey Geographic (SSURGO) Database. For Duke, Harvard, and Howland site, there are approximately 7000, 2300, and 2100 possible soil profiles within the area inside the radar swath. For each of these possible clay profile, we construct data cube by varying soil moisture profile. We assume the soil moisture profile follow second-order polynomial as a function of depth from the surface, i.e.

$$Mv(z) = az^2 + bz + c \qquad 0 < z < 1 \qquad z \text{ is depth from the surface in meters}$$

Although, the parameter a, b, c can be mathematically any numbers, in our calculation, they are allowed to vary only under the following rules to fit the physics of soil moisture:

0 <= c <= 50 (soil moisture at the surface is smaller than 50%)

0 <= $Mv(z)$ <=50 (soil moisture is always positive and less than 50% at any depth)

We pre-compute the reflections coefficients (R_{vv}, R_{hh}) and the average soil moisture with varying *a*, *b*, and *c* for each of the clay profile and save it for the soil moisture retrieval after the soil moisture retrieval are performed using the fit from Eq. 1.

Soil moisture profile parameters retrieval with data cube using cost function

From previous section, we retrieve the soil moisture profile parameters (*a*, *b*, and *c*) using data cube by comparing the datacube observables ($R_{vv}^{'}$, $R_{hh}^{'}$ and $\acute{M}_v$) against the observables from radar data (R_{vv}, R_{hh}, M_v). The observables from radar data are the results of plugging in the retrieved information (soil moisture m_v, biomass W, and soil surface roughness s) into equation (1) along with the backscatter (σ_{pq}). However, this time, we treat Γ_{pq} as unknown.

The cost function to find the parameters *a*, *b*, and *c* are

$$\Phi = A\left|\acute{M}_v - M_v\right| + B\left|\acute{R_{vv}} - R_{vv}\right| + C\left|\acute{R_{hh}} - R_{hh}\right| + D|a| + E|b|$$

As the expression suggested, the cost is weighted error of observables and also a function of polynomial parameters *a* and *b*. We found that there are multiple *a*, *b*, *c* parameters that can provide the same cost. The added cost function $D|a|$ and $E|b|$ is to narrow down the choice under the assumption that the more probable profile shape is the one where moisture content varies the least as a function of depth (small *a* and small *b*).

From the observation of the ground sensor soil moisture profile behavior as a function of time, there usually is only small change among the flight days within a season (usually within 10 days), particularly, the soil profile shape. Thus, we enforce extra cost function to constraint soil moisture profile solution in such a way that the solution favors only small change is soil moisture profile among the flight dates in the same season. This is done by first finding likely candidates from imposing the first cost function within prescribe threshold. Then, among the candidates, evaluate the following cost function

$$\Upsilon = F\left|a_i - a_j\right| + G\left|b_i - b_j\right| + H\left|c_i - c_j\right|$$

where a_i is the candidate solutions for date i. This extra cost function optimizes the solutions in the way that the solution both matches the retrieved average moisture and reflection coefficients and is consistent from day to day.

Constraints on profile polynomial retrieval

To obtain meaningful and physically possible retrieval, we constraint the potential candidates for profile retrieval based on knowledge of the measured soil moisture profile and its temporal – spatial variability. When the soil moisture profile measurements on the ground exists, which happen in most case for Duke, Harvard, and Chamela, we calculate the polynomial fit of the profile and the statistics of the coefficients of the fit. This information is used to further constrain the possible solutions.

Algorithm options

There are a number of retrieval options that impact the retrieval results. They are listed in Table 7.

TABLE 7. Mixed-species vegetation algorithm option

Option	Values used in Retrieval	Description
Options for applying correction factor		
corr_fac	'on'	'on/off': whether to apply the correction factor

ref_urban, ref_water, ref_barren, ref_forest, ref_deciduous_forest, ref_evergreen_forest, ref_mixed_forest	Corresponding values from NLCD or Globcover database	NLCD or Globcover values of the corresponding land cover types
Options for post binding removal and mosaic		
cutoff_ratio	15%	Percentage of the samples to ignore at the high- and low-end of the statistical distribution when computing the sample statistics
line_balancing	'on'	'on/off': whether to apply the re-balancing to further remove the banding in the retrieval among flight lines
satpixel_throwback	'off'	'on/off': whether to re-retrieve the pixels that might be mistaken as vegetated pixels
satline_criteria	Site dependent values	A threshold, above which if the percentage of the saturated pixels in a flight line is, the line will be considered as a saturated flight line
satline_throwback	'on'	'on/off': whether to re-retrieve the saturated flight line with further bias correction
mosaicmtd	'latest line'	'average/higher look angle/latest line': the method to use when mosaic the flight lines
Options for line preparation		
lookangmask (Inp.minlookang, Inp.maxlookang)	'on'	'on/off': whether to apply the look angle mask to only retrieve pixels having look angles between Inp.minlookang (20 degrees) and Inp.maxlookang (55 degrees)
NLCDmask (Inp.NLCDmasklist)	'on'	'on/off': whether to apply the land cover mask (based on the Inp.NLCDmasklist)
angcomp	'on'	'on/off': whether to apply angle compensation (i.e. normalization)
angcompopt	'incang'	'incang/lookang': whether to use incidence angle or look angle for angle compensation

plyfit (Inp.plyfit_order)	'on'	'on/off': whether to use polynomial fitting for angle compensation, and the order of the polynomial to use
anglerange (Inp.angleinterval, Inp.anglemin, Inp.anglemax)	'partial'	'partial/full': whether to use the partial or full angle range for the angle compensation. If partial, the range is defined by Inp.anglemin and Inp.anglemax. The angle interval is given by Inp.angleinterval.
Options for retrieval initialization		
vegdet.alg **vegdet.hvthreshold**	'hv threshold' -15 (dB)	'hv threshold': to set the method used for determine the vegetated/non-vegetated pixels during retrieval initialization, and the hv threshold value
vegdet.nonvegpixel_throwback	'on'	'on/off': for those non-vegetated pixels having saturated retrieval values after using the bare surface algorithm, whether to treat them as vegetated pixels
AGBretv.alg (Inp.AGBretv.coef)	'method2' for Chamela 'method1' for rest sites	algorithm to use for biomass initial retrieval and the coefficients
Baresoilretv.alg	'Oh92'	algorithm to use for dielectric constant and roughness initial retrieval on bare surface
Options for retrieval		
baresoil_opt	'Oh92'	algorithm to use in the retrieval for bare soil
retv.rmsh_lb, retv.rmsh_ub **retv.AGB_lb, retv.AGB_ub** **retv.er_lb, retv.er_ub**	0, 0.2 0, 250 2, 55	lower and upper bounds for roughness, biomass, and dielectric constant retrievals
retv.incang_mtd **retv.incangval**	'fix' 40 (degrees)	'fix/variable': whether to use a fixed value or a variable for the incidence angle used in the retrieval; and the incidence angle value if set to be fixed
retv.channelweight	'on'	'on/off': whether to apply different weight for HH, VV, and HV channels

Options for reflectivity computation		
GammaCalmtd	‘use_sigma0’	‘use_sigma0/use_er’: method to calculate the reflection coefficients
Options for profile retrieval		
datacubetype	‘1-layer_poly’	‘’1-layer_poly/2-layer_poly: datacube type options for profile retrieval
multday	‘on’	‘on/off’: whether to use all three-day information or single day information to retrieve profile

4.3 Retrieval Error Improvement and Assessment

The existing uncertainty in the retrieval results are associated with radar calibration errors, vegetation parameterization errors, surface roughness assumptions, and inaccuracies in the scattering and inversion models. We have continued to improve the accuracy of retrievals as the delivery of the RZSM products has progressed since 2012. The upper and lower bounds placed on the unknowns have been improved based on in-situ data. We have also progressively changed the implementation of the inversion code and its subroutines as we found more accurate and efficient ways for the mathematical operations. Moreover, given the project latency requirements, we are going to reprocess some of the flights using the recalibrated radar data over the sites with larger retrieval inaccuracy.

4.3.1 Validity Depth

The retrieval error generally increases when deeper points from installed probes are used in validation. This increase in error is expected as penetration depth of electromagnetic waves decreases with depth. We can therefore define a `threshold depth' for each site as the depth up to which the retrieval is mathematically valid. The threshold depth depends on the number of modeled layers and their thickness in the forward model. The validity depth for each site is different because of its different vegetation, soil texture, and soil layers. However, it can be shown that the choice of 50 cm is reasonably accurate for all sites. All 2012 and 2013 RZSM products from BERMS, Metolius, MOISST, Tonzi Ranch, and Walnut Gulch (except for Tonzi Ranch 2013) are valid up to 95 cm. All 2014 and 2015 RZSM products from BERMS, Metolius, MOISST, Tonzi Ranch, and Walnut Gulch are valid up to 45 cm. Tonzi Ranch 2013 products are valid up to 55 cm.

4.3.2 Unknown Bounds

The retrieval algorithm of AirMOSS only uses the HH and VV channels due to calibration inaccuracy of the HV channel. Therefore, the corresponding inverse problem is ill-posed as 3 unknown parameters (i.e., a, b, and c) are retrieved with only 2 data points. Some regularization

is thus necessary to overcome the effect of ill-posedness. The method we have applied for inversions at used for BERMS, Metolius, MOISST, Tonzi Ranch, and Walnut Gulch is constraining each unknown with upper and lower bounds based on in-situ soil moisture data at each of these sites. Considering the assumption that RZSM is represented with quadratic

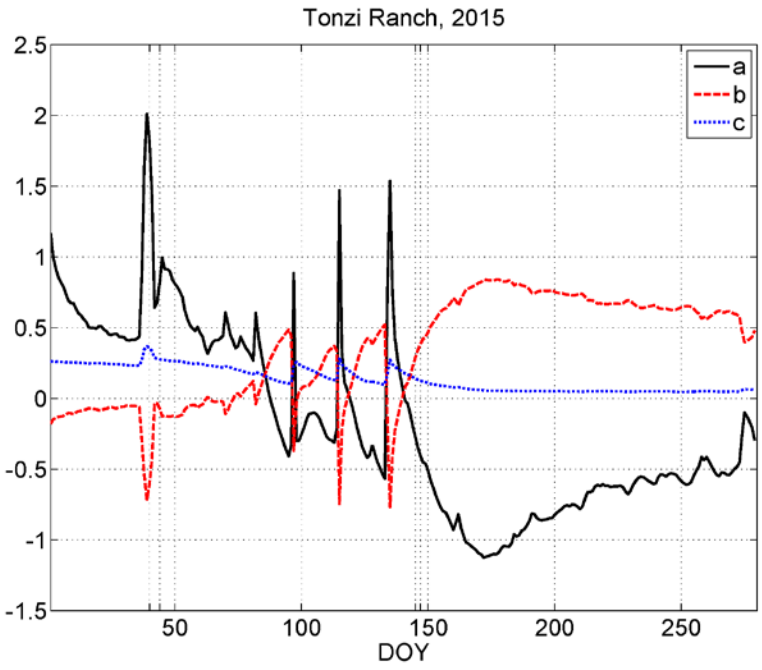

Fig. 15: The variation of the coefficients of the quadratic function fitted to the average in-situ soil moisture profile in Tonzi Ranch in 2015

function, in-situ soil moisture profiles at each site are fitted with a quadratic function and the corresponding polynomial coefficients are observed throughout the year. An upper and lower bound is empirically selected for each flight date based on the behavior of the coefficients within a time period that includes the flight date. Fig. 15 shows the variation of the coefficients of the quadratic function fitted to the average in-situ soil moisture profile in Tonzi Ranch in 2015.

4.3.3 Probe Behavior

The behavior of each in-situ soil moisture probe is studied before the corresponding data are used in validation and calculation of RMSE. Soil moisture probes in Walnut Gulch sometimes report values for only 2 depths. We only use Walnut Gulch probes that reports soil moisture values for at least 3 depths. Needless to say, if a probe malfunctions, hence reporting non-physical values, the corresponding values are excluded in the calculation of daily average of soil moisture. Moreover, some of the probes are installed close to developed areas such as roads in MOISST. These probes are excluded in the calculation of the retrieval error.

4.3.4 Daily Average of Measured Soil Moisture

Each AirMOSS site has several installed probes each of which measures soil moisture at several depths every 30 minutes or hour. These probes are located either within one radar pixel or in adjacent radar pixels. The measured value by a probe at each depth can be used in two different fashions. The value measured around the flight time can be directly used for validation against retrieved soil moisture. However, to remove uncertainty in measurements, we have been using the daily average of the measured moisture except for flights a few hours before or after precipitation. Even in such cases, the use of daily average as opposed to hourly value would not make a significant change in the accuracy.

4.3.5 Bias Removal

The error in the retrieved soil moisture values correspond to several factors including mathematical inaccuracies of forward and inverse models, physical inaccuracies of forward and inverse models imposed by penetration depth, radar calibration, installed probes, and inaccuracies in land cover classification as well as vegetation and soil parameters. These factors can cause systematic biases present in our RZSM products. Each site has characteristics that make the bias unique to that site. Taking into account variables such as yearly precipitation and improvement of radar calibration throughout the mission, we consider the bias a site-dependent parameter that needs to be calculated for each year separately. We have used the following simple method for removing bias from the products by first calculating the average error in the

$$\mathrm{RMSE} = \sqrt{\frac{\sum \left(\hat{m}_v - E\left(\hat{m}_v - m_v\right) - m_v\right)^2}{N}}$$

retrieved values of soil moisture and then evaluating the RMSE based on bias-free sets of points:

where $\widehat{m_v}$ and m_v denote the estimated and actual value for soil moisture at each measurement point, $E(\widehat{m_v} - m_v)$ is the average retrieval error over all measurement points in the site throughout the year, and N is the total number of measurement points in the site throughout the year.

4.3.6 Reprocessing Recalibrated Flights

Given that some of the flights, especially the ones from 2013—when AirMOSS radar experienced some hardware complexities—have been recalibrated, we have reprocessed the flights and redelivered the corresponding RZSM products. The reprocessed data show noticeable improvement in error. Fig. 16 shows a comparison between the retrieval errors of Tonzi Ranch 2013 flights using old calibration (red) and new calibration (black).

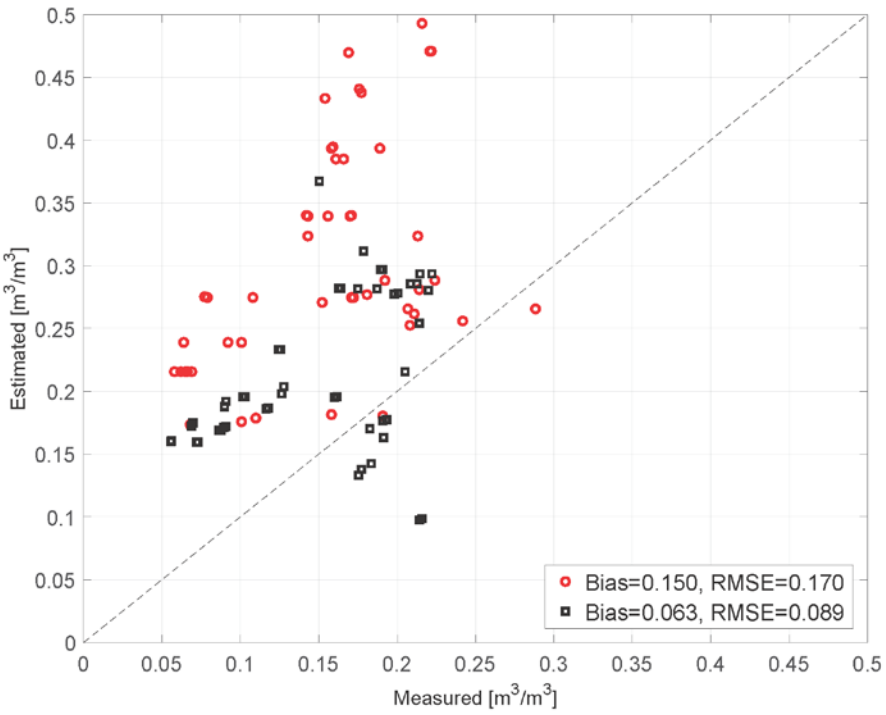

Fig. 16: Comparison between retrieval errors of Tonzi Ranch 2013 flights using old calibration (red) and new calibration (black)

4.4 Practical Considerations

Unless otherwise noted, the following considerations apply to both algorithm types described above.

4.4.1 Numerical Computation Considerations

Using the SPM-based and KA-based models for the ground scattering simulations, the computational cost of predicting the total backscattering coefficient only depends on the cost of the forest scattering model. Calculation of the monospecies vegetation model output takes approximately 0.2s for one set of polarimetric P-band backscattering coefficients on an AMD 3.4 GHz CPU. Computation time of the forward scattering model increases in proportion to frequency (determined roughly by the number of cylindrical harmonics required for a convergent solution of scattering from cylinders).

The computational cost of the inversion algorithm depends on many factors including the cost of the forward model evaluation, measurement parameters (i.e., number and value of frequency points and number of observation angles), and inversion parameters. The value of frequency affects the forest scattering model convergence, hence its computational cost. The number of frequency points and observation angles directly affect the cost of evaluation of the cost function

as evident from its definition above. The measurement parameters, detailed in Tabatabaeenejad and Moghaddam (2006), also directly affect the convergence speed. These parameters are usually problem-specific and are chosen empirically. There is always a trade-off between accuracy of inversion and its convergence speed.

4.4.2 Calibration and Validation

The baseline validation plan is to evaluate the L2/3-RZSM retrievals against profile soil moisture observations acquired with ground instrumentation installed at each AirMOSS flux tower site. As explained in Section 2, more intensive field sampling is planned during the flight campaigns to collect spatially diverse validation data.

4.4.3 Interface Assumptions

Input Interface

The primary input to the AirMOSS L2/3-RZSM product is the AirMOSS L1-S0 product, which were provided as binary format floating point data in individual polarimetric channels, not the compressed Stokes matrix L1-S0 product. In addition, ancillary data, described earlier, are in HDF-5 format.

Output Interface

The outputs of the L2/3-RZSM processing are RZSM retrievals with properties explained earlier in this document. The output format is in HDF-5, and the primary user of the product is the L4-RZSM product.

4.4.4 Test Procedures

The AirMOSS L3/2-RZSM algorithm have undergone comprehensive testing and quality control with simulated and measured radar data prior to and during AirMOSS flight campaigns. The effectiveness of the simulated annealing method for radar retrievals, which has been the basis of the monospecies retrieval algorithm, has been extensively validated previously and results have been published in peer review literature (Tabatabaeenejad and Moghaddam, 2009 and 2011; Tabatabaeenejad *et al.*, 2012a).

5 FUTURE ALGORITHM ENHANCEMENTS

Enhancements to both the forward scattering and the inversion algorithms are ongoing and have been implemented through the course of the AirMOSS mission and upon satisfactory validation. In particular, the following enhancements are planned:

Enhanced scattering model of soil surface and subsurface: the current choice of SPM to calculate backscattering from the soil surface and profile along with KA for the coherent specular scattering can be replaced with the SEBCM (Duan and Moghaddam, 2012; Duan and Moghaddam, 2013), which is a numerical solution. SEBCM has a significantly larger region of validity (with respect to surface roughness) than SPM and KA, and produces the bistatic scattered fields so that the backscattering and forward scattering coefficients are computed all at

once. It also includes all orders of scattering, coherent and incoherent, and therefore is expected to be more accurate all-around. The disadvantage of SEBCM is its higher computational complexity compared with the SPM and KA analytical techniques. SEBCM is, however, significantly more efficient than MoM, and holds the promise of being a practical choice for the forward model as the computational resources are expanded.

Model of sloped ground: the current surface and vegetation scattering models assume that the ground surface is flat (horizontal). However, in reality this is not always the correct assumption as several of the AirMOSS flight lines include locations with considerable slopes. Whereas the baseline algorithm excludes pixels with slopes of larger than 5 degrees, it is desirable to relax this limitation such that retrievals can be achieved for sloped terrain. For scattering from bare surfaces (or the direct ground scattering contribution from vegetated surfaces), the modification is straightforward and is achieved by the proper modification of the local incidence angle using the local slope. The scattering from vegetation volume is unaffected by the terrain slope. The principal scattering mechanism whose implementation becomes more complicated is the double bounce mechanism, where the bistatic scattering in the forward direction from both the trees (trunks and branches) and the ground must be calculated. This will be achieved by using the SEBCM (Duan and Moghaddam, 2012; Duan and Moghaddam, 2013), which produces the bistatic full vector fields. Once again, the current limitation is the higher computational cost of SEBCM.

6 ACKNOWLEDGEMENTS

Critical review and comments from JPL's Dr. Elaine Chapin are gratefully acknowledged. The general structure of this document was derived from that of the SMAP mission Level 2 high-resolution radar product ATBD, whose lead authors we acknowledge for providing us with that draft document.

7 REFERENCES

Burgin, M., D. Clewley, R. Lucas, and M. Moghaddam, "A generalized radar backscattering model based on wave theory for multilayer multispecies vegetation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4832–4845, Dec. 2011.

Chapin, E., A. Chau, J. Chen, B. Heavey, S. Hensley, Y. Lou, R. Machuzak, and M. Moghaddam, "AirMOSS: an airborne P-band SAR to measure root-zone soil moisture," in *Proc. IEEE Radar Conference*, Atlanta, GA, May 2012, pp. 693–698.

Chauhan, N. and R. Lang, "Radar modeling of a boreal forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 4, pp. 627–638, Jul. 1991.

Corana, A., M. Marchesi, C. Martini, and S. Ridella, “Minimizing multimodal functions of continuous variables with the “simulated annealing” algorithm,” *ACM Trans. Math. Softw.*, vol. 13, no. 3, pp. 262–280, Sep. 1987.

Demir, M. A., J. T. Johnson, and T. J. Zaidel, “A study of the fourth-order small perturbation method for scattering from two-layer rough surfaces,” *IEEE Trans. Geosci. Remote Sens.*, in press, 2012.

Dobson, M. C., F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes, “Microwave dielectric behavior of wet soil—Part II: Dielectric mixing models,” *IEEE Trans. Geosci. Remote Sens.*, vol. GE-23, no. 1, pp. 35–46, Jan. 1985.

Dobson, M. C., F. T. Ulaby, T. Le Toan, A. Beaudoin, E. S. Kasischke, and N. Christensen, “Dependence of radar backscatter on coniferous forest biomass,” *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 412–415, Mar. 1992.

Dobson, M. C., F. T. Ulaby, L. E. Pierce, T. L. Sharik, K. M. Bergen, J. Kellndorfer, J. R. Kendra, E. Li, Y. C. Lin, A. Nashashibi, K. Sarabandi, and P. Siqueira, “Estimation of forest biophysical characteristics in Northern Michigan with SIR-C/X-SAR,” *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 877–895, Jul. 1995.

Duan, X. and M. Moghaddam, “3-D vector electromagnetic scattering from arbitrary random rough surfaces using Stabilized Extended Boundary Condition Method for remote sensing of soil moisture,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 87–103, Jan. 2012.

Duan, X. and M. Moghaddam, “Bistatic vector 3D scattering from layered rough surfaces using stabilized extended boundary condition method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2722–2733, May 2013.

Dubois, P. C., J. Van Zyl, and E. T. Engman, “Measuring soil moisture with imaging radar,” *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 915–926, Jul. 1995.

Durden, S. L., J. J. Van Zyl, and H. A. Zebker, “Modeling and observation of the radar polarization signature of forested areas,” *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 3, pp. 290–301, May 1989.

El-Rayes, M. A. and F. T. Ulaby, “Microwave dielectric spectrum of vegetation—Part I: Experimental observations,” *IEEE Trans. Geosci. Remote Sens.*, vol. GE-25, no. 5, pp 541–549, Sep. 1987.

El-Shenawee, M., C. Rappaport, E. Miller, and M. Silevitch, “Three-dimensional subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces:

Use of the steepest descent fast multipole method (SDFMM),” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 1174–1182, Jun. 2001.

Fung, A. K., Z. Li, and K. S. Chen, “Backscattering from a randomly rough dielectric surface,” *IEEE Trans. Geosci. Remote Sens.*, vol. 30**,** no. 2, pp. 356–369, Mar. 1992.

Fuks, I. M. and A. G Voronovich, “Wave diffraction by rough interfaces in an arbitrary plane-layered medium”, *Waves in Random Media*, Vol. 10, 253–272, 2000.

Haddad, Z., P. Dubois, and J. Van Zyl, “Bayesian estimation of soil parameters from radar backscatter data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 1, pp. 76–82, Jan. 1996.

Imhoff, M., “Radar backscatter and biomass saturation: ramifications for global biomass inventory,” *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 511–518, Mar. 1995.

Ishimaru, A., *Wave Propagation and Scattering in Random Media*. New York: Wiley-IEEE Press, 1997.

Israelsson, H., L. M. H. Ulander, J. L. H. Askne, J. E. S. Fransson, P.-O. Frolind, A. Gustavsson, and H. Hellsten, “Retrieval of forest stem volume using VHF SAR,” *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 36–40, Jan. 1997.

Israelsson, H., L. M. H. Ulander, T. Martin, J. I. H. Askne, “A coherent scattering model to determine forest backscattering in the VHF-band,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 238–248, Jan. 2000.

Kapp, D. and G. Brown, “A new numerical method for rough-surface scattering calculations,” *IEEE Trans. Antennas Propagat.*, vol. 44, no. 5, pp. 711–721, May 1996.

Kong, J.A., *Electromagnetic Wave Theory*. Cambridge, MA: EMW Publishing, 2000.

Kuo, C.-H., and M. Moghaddam, “Electromagnetic scattering from multilayer rough surfaces separated by media of arbitrary dielectric profiles for remote sensing of soil moisture,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 349–367, Feb. 2007.

Le Toan, T., A. Beaudoin, J. Riom, and D. Guyon, “Relating forest biomass to SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 403–411, Mar. 1992.

Liang, P., M. Moghaddam, L. Pierce, and R. Lucas, “Radar backscattering model for multilayer mixed-species forests,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2612–2626, Nov. 2005.

Lin, Y.-C. and K. Sarabandi, “A Monte Carlo coherent scattering model for forest canopies using fractal-generated trees,” *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 440–451, Jan. 1999.

Markwardt, C. B., "Non-Linear Least Squares Fitting in IDL with MPFIT," in Astronomical Data Analysis Software and Systems XVIII, vol. 411, 2009, pp. 251-254.

Mironov, V. L., L. G. Kosolapova, and S. V. Fomin, “Physicallly and mineralogically based spectroscopic dielectric model for moist soils,” *IEEE Trans. Geosci Remot Sens*., vol 47, no. 7, pp. 2059 – 2070, July 2009.

Moss, C.D., T. M. Gezegorcyk, H. C. Han, and J. A. Kong, “Forward–backward method with spectral acceleration for scattering from layered rough surfaces,” *IEEE Trans. Antennas Propagat.*, vol. 54, no. 3, pp. 1006–1016, Mar. 2006.

Nghiem, S., R. Kwok, S. Yueh, J. Kong, C. Hsu, M. Tassoudji, and R. Shin, “Polarimetric scattering from layered media with multiple species of scatterers,” *Rad. Sci.*, vol. 30, no. 4, pp. 835–852, 1995

Notarnicola, C. and F. Posa, “Inferring vegetation water content from C- and L-band SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3165–3171, Oct. 2007.

Oh, Y., K. Sarabandi, and F. T. Ulaby, “An empirical model and an inversion technique for radar scattering from bare soil surfaces,” *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 370–381, Mar. 1992.

Oh, Y. “Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces,*” IEEE Trans. Geosci. Remote Sens*, vol. 42, no. 3, pp 596 – 601, Mar 2004.

Picard, G., T. Le Toan, and F. Mattia, “Understanding C-band radar backscatter from wheat canopy using a multiple-scattering coherent model,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 7, pp. 1583–1591, Jul. 2003.

Peplinski, N. R., F. T. Ulaby, and M. C. Dobson, “Dielectric properties of soils in the 0.3–1.3-GHz range,” *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 803–807, May 1995. (1995a)

Peplinski, N. R., F. T. Ulaby, and M. C. Dobson, “Corrections to “Dielectric properties of soils in the 0.3–1.3-GHz range”,” *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 6, p. 1340, Nov. 1995. (1995b)

Ranson, K. J. and G. Sun, “Mapping biomass of a northern forest using multi-frequency SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 388–396, Mar. 1994.

Saatchi, S. S. and K. C. McDonald, “Coherent effects in microwave backscattering models for forest canopies,” *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 1032–1044, Jul. 1997.

Saatchi, S. S. and M. Moghaddam, “Estimation of crown and stem water content and biomass of boreal forest using polarimetric SAR imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 697–709, Mar. 2000.

Saatchi, S., M. Marlier, R. L. Chazdon, D. B. Clark, and A. E. Russel, “Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass,” *Remote Sensing of Environment*, Vol. 115, no. 11, pp. 2836-2849, Nov 2011.

Shi, J. C., J. Wang, A. Y. Hsu, P. E. O'Neill, and E. T. Engman, “Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp.1254–1266, Sep. 1997.

Shimada, J., "AirMOSS Science Product Grid Specifications," Jet Propulsion Laboratory Interoffice Memorandum 334-JGS-20120113-002, Feb. 2012.

Soubret, A., G. Berginc, and C. Bourrely, “Backscattering enhancement of an electromagnetic wave scattered by two-dimensional rough layers,” *J. Opt. Soc. Am. A., Opt. Image Sci.*, vol. 18, no. 11, pp. 2778–2788, Nov. 2001.

Sun, G. and K. J. Ranson, “A three-dimensional radar backscatter model of forest canopies,” *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 372–382, Mar. 1995.

Tabatabaeenejad, A. and M. Moghaddam, “Bistatic scattering from layered rough surfaces,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2102–2115, Aug. 2006.

Tabatabaeenejad, A. and M. Moghaddam, “Inversion of dielectric properties of layered rough surface using the simulated annealing method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2035–2046, Jul. 2009.

Tabatabaeenejad, A. and M. Moghaddam, “Retrieval of surface and deep soil moisture and effect of moisture profile on inversion accuracy,” *IEEE Geosci. Remote Sensing Lett.*, vol. 8, no. 3, pp. 477–481, May 2011.

Tabatabaeenejad, A., M. Burgin, and M. Moghaddam, “Potential of L-band radar for retrieval of canopy and subcanopy parameters of boreal forests,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2150–2160, Jun. 2012. (2012a)

Tabatabaeenejad, A., M. Burgin, and M. Moghaddam, “Impact of subsurface layers on subcanopy soil moisture retrieval from radar,” presented at *IEEE IGARSS*, Munich, Germany, Jul. 2012. (2012b)

Tabatabaeenejad, A, X. Duan, and M. Moghaddam, “Coherent scattering of electromagnetic waves from two-layer rough surfaces within the Kirchhoff regime,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3943–3953, Jul. 2013.

Tabatabaeenejad, A., M. Burgin, X. Duan, and M. Moghaddam, “P-band radar retrieval of subcanopy and subsurface soil moisture profile as a second-order polynomial: first AirMOSS results,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 645–658, Feb. 2015.

Thirion, L., E. Colin, and C. Dahon, “Capabilities of a forest coherent scattering model applied to radiometry, interferometry, and polarimetry at P- and L-band,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 4, pp. 849–862, Apr. 2006.

Topp, G. C. "Electromagnetic Determination of Soil Water Content: Measurements in Coaxial Transmission Lines," *Water Resources research*, vol. 16, no. 3, pp. 574–582, June 1980.

Truong-Loi, M.-L., S. Saatchi, and S. Jaruwatanadilok, “Soil moisture estimation under tropical forests using UHF radar polarimetry,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1718–1727, Apr. 2015.

Tsang, L., R. T. Shin, and J. A. Kong, *Theory of Microwave Remote Sensing.* New York: Wiley Interscience, 1985.

Tsang L., C. H. Chan, K. Pak, H. Sanganni, A. Ishimaru, and P. Phu, “Monte Carlo simulations of large-scale composite random rough-surface scattering based on the banded-matrix iterative approach,” *J. Opt. Soc. Am. A., Opt. Image Sci.*, vol. 11, no. 2, pp. 691–696, Feb. 1994.

Ulaby, F., K. Sarabandi, K. McDonald, M. Witt, and M. C. Dobson, “Michigan microwave canopy scattering model,” *Int. J. Remote Sens.*, vol. 11, no. 7, pp. 1223–1253, 1990.

Van Zyl, J. and Y. Kim, *Synthetic Aperture Radar Polarimetry.* Hoboken, NJ: John Wiley & Sons, 2011. [Online]. Available: http://descanso.jpl.nasa.gov/SciTechBook/SciTechBook.cfm.

Wang, Y., J. Day, and G. Sun, “Santa Barbara microwave backscattering model for woodlands,” *Int. J. Remote Sens.*, vol. 14, no. 8, pp. 1477–1493, May 1993.

Xu, F. and Y.-Q. Jin, “Multiparameter inversion of a layer of vegetation canopy over rough surface from the system response function based on the Mueller matrix solution of pulse echoes,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 2003–2105, Jul. 2006.