Algorithm Theoretical Basis Document
The AirMOSS Level 4 Root-Zone Soil Moisture product

Wade Crow[1], Sushil Milak[1,2], and Rolf Reichle[3]

[1] USDA ARS Hydrology and Remote Sensing Laboratory
[2] Science Systems Applications, Inc.,
[3] NASA GSFC Global Modeling and Assimilation Office

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 estimates' 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 actual 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;
(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 reduce uncertainty in estimates of NEE through the development of methodologies that 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 (Soil Moisture Active Passive) mission and assessing the impact of fine-scale heterogeneities in its coarse-resolution products.

AirMOSS surveys were conducted over ten regions of approximately 100 km × 25 km surrounding FLUXNET tower sites within nine different 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 provide radar backscatter measurements at 100 m spatial resolution and at select sub-weekly, seasonal, and annual time scales.

The AirMOSS radar and associated RZSM benchmark datasets, a first of their kind, represent 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 100 m resolution (Level 2/3 RZSM – hereinafter "L2/3 RZSM"), continuous estimates of RZSM obtained via the integration of the Level 2/3 product with a land surface model (Level 4 RZSM – hereinafter "L4 RZSM"), estimates of NEE at 1 km resolution through ecosystem modeling (Level 4A NEE – hereinafter "L4 NEE"), and integrated North American NEE estimates at 50 km resolution (Level 4B NEE).

Table 1. Summary of the nine North American biomes covered by the AirMOSS baseline investigation.

Biome #	Biome type IGBP* Vegetation class	Example site Name and location
1	Boreal forest/evergreen needle-leaf, mixed forest, cropland	BERMS*, Saskatchewan, Canada
2	Boreal transitional /mixed forest	Howland, ME and Harvard, MA forests
3	Temperate forest/mixed forest, cropland	Duke forest, North Carolina
4	Temperate forest/evergreen needle-leaf	Metolius, Oregon
5	Temperate grasslands/Crops	MOIST#, Oklahoma
6	Mediterranean forest/woody savanna	Tonzi Ranch, California
7	Desert and shrub/open shrubland & 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

*BERMS: Boreal Ecosystem Research and Monitoring Sites
#MOISST: Marena Oklahoma In Situ Sensor Testbed

1.2 Measurement Approach

The AirMOSS instrument is flown on a Gulfstream-III (G-III) or similar aircraft, equipped with a P-band Synthetic Aperture Radar (SAR), over nine major biomes of North America. The AirMOSS radar is a modification of the L-band radar in NASA's Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) program. For AirMOSS, the L-band front-end electronics and antenna were replaced with components that operate at P-band.

The AirMOSS project is producing L2/3 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

100 km x 25 km at each of the 9 biomes listed in Table 1. For addition details see the AirMOSS L2/3 ATBD.

In-ground soil parameters and ancillary data described here 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 campaigns in late 2015. The sensors measure surface soil temperature, soil matric potential, soil moisture content at several different depths. Precipitation amounts are also measured at each site. At select sites, AirMOSS also used a low altitude airborne platform (Purdue University's modified Beechcraft Duchess Airborne Laboratory for Atmospheric Research) to measure the spatial variations of atmospheric CO_2 fluxes with the Picarro wavelength-scanned cavity ring-down spectrometer.

The NEE estimate and accompanying uncertainty reduction estimates is 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 50 km x 50 km grid cells within the North American region. The appropriate statistical weighting and interpolation values for each grid cell will be 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

1.3 Motivation for L4 RZSM Data Product

The AirMOSS L2/3 RZSM algorithm theoretical basis document (ATBD) describes retrieval of the L2/3 RZSM estimates on a temporally-intermittent basis (i.e., roughly 3-9 acquisitions per year concentrated within 1 to 3 intensive periods of 10 to 14 days in duration). These retrievals will be presented in the form of a 2nd-order polynomial (valid up to 95 cm[1] in depth) describing the variation of volumetric soil moisture with depth with a lateral ground resolution of 100 m.

Many key applications for AirMOSS data products, including the calculation of net ecosystem exchange (NEE) by the L4 NEE product, require temporally continuous RZSM estimates. As a consequence, a physically-realistic interpolation method is required to convert intermittent L2/3 RZSM retrievals into temporally continuous estimates.

2. Overview and Approach

2.1 Product/Algorithm Objectives

The objective of the AirMOSS L4 RZSM algorithm is the realistic temporal interpolation of intermittent AirMOSS L2/3 RZSM retrievals into a temporally-continuous, multi-layer, hourly

---

[1] Reduced to 45 cm at the Tonzi and MOISST sites listed in Table 1.

soil moisture product. Per the project level requirements, the L4 RZSM products have the same spatial resolution (3-arcsecs or ~100 m) as the underlying L2/3 RZSM retrievals.

2.2 Historical Perspective

The realistic temporal interpolation of L2/3 RZSM estimates requires the use of a hydrologic process model capable of capturing the impact of temporal variations in L2/3 RZSM arising from the dynamic interaction of infiltration, evapotranspiration, baseflow and drainage processes on root-zone soil moisture storage. From this perspective, L2/3 RZSM interpolation can be recast as a data analysis problem where non-continuous L2/3 RZSM retrievals are used to fix L2/3 RZSM at observations times and the hydrologic model is used to interpolate between observations. This general model can be executed using either a state or parameter updating strategy.

However, the sparse temporal availability of the AirMOSS L2/3 RZSM presents unique challenges for both strategies. State updating, typically referred to as data assimilation, is based on correctively modifying model state predictions to account for changes in the likelihood of that state associated with the acquisition of a related observation. For the L4 RZSM product, the key states will be soil moisture within a set of vertically-discrete soil layers. However, given the dissipative nature of hydrologic models, state updating is only effective when the time scales between successive observations are small relative to the persistence scale of updated states. Otherwise, hydrologic model states will relax back to nominal (i.e., non-updated) model levels between observations. Therefore, in the limit as observations become less frequent, the assimilated observations will be effectively ignored by the overall data assimilation analysis. Unfortunately, the typical time scale of persistence for RZSM fluctuations (typically on the order of weeks) is shorter than the expected interval between clusters of L2/3 RZSM retrievals at a particular AirMOSS study site (typically on the order of months). Consequently, state updating is not an effective interpolation strategy for the L4 RZSM algorithm.

Hydrologic models also require input parameters. In contrast to hydrologic model states, these parameters are externally proscribed and not internally generated by the model. The modification of model parameters in response to observations is typically referred to as model calibration. Consequently, changes in parameters can be assumed to persist beyond the response time scales of states. As a result, model calibration is typically a more effective method of temporally transmitting information from sparse observations throughout continuous model runs. However, in practice, such approaches face a severe dimensionality problem as spatially-distributed patterns of multiple land surface parameters must be individually-calibrated to produce the correct spatial pattern of soil moisture observed by the hydrologic process model. In addition, required spin-up times (to achieve equilibrium in all model states after changing a parameter) are extremely long (> 10 years) for the hydrologic model at certain sites. This represents a significant computational barrier for batch calibration schemes requiring a large number of end-to-end PIHM runs. As a result, we found that the robust calibration of the model was not possible given available computational resources and the length (and temporal frequency) of available L2/3 RZSM retrievals.

Due to these challenges, a relatively-simple interpolation approach was instead applied which provided robust state updating based on feasible computation loads (see Section 4.1.2b below).

2.3 Instrument/Product Characteristics

2.3.1 Instrument/Calibration Aspects (affecting product)

The AirMOSS instrumentation impacts on the L4 RZSM algorithm flow directly from its use of L2/3 RZSM data products. See the L2/3 RZSM ATBD for relevant details.

2.3.2 Data Product Characteristics

This section provides a summary of the AirMOSS L4 RZSM product specifications.

2.3.2a Geophysical parameters

The AirMOSS L4 RZSM product includes the following four components:

> (1) layer 1 soil moisture (0-10 cm vertical average) in volumetric soil moisture (VSM) units (m^3 m^{-3}),
>
> (2) layer 2 soil moisture (10-40 cm vertical average) in VSM units,
>
> (3) layer 3 soil moisture (40-100 cm vertical average) in VSM units,
>
> (4) and layer 4 soil moisture (100-200 cm vertical average) in VSM units.

to be generated independently over each of the AirMOSS study sites listed in Table 1.

2.3.2b Spatial resolution, posting, and coverage

For each study site in Table 1, posting and coverage of the L4 RZSM corresponds to that of the L2/3 RZSM product. As such, the L4 RZSM product is posted at 3-arcsec (~100 m) grid defined within each of the study sites. As discussed in Section 4, the fundament coverage units of the L4 RZSM products are individual hydrologic catchments. These catchments were selected to maximize overlap with L2/3 RZSM coverage at each site; however, spatial gaps in L4 RZSM coverage persist along the edges of the study sites and in any area lacking adequate L2/3 RZSM coverage (see Section 4.1.2c).

2.3.2c Temporal resolution and sampling

The L4 RZSM product has an hourly temporal resolution. Individual soil moisture values will correspond to instantaneous values obtained at 0:00 UTC, 1:00 UTC, … , and 23:00 UTC.

2.3.2d Latency

The project level latency requirement for the L4 RZSM product is 6 to 9 months after the relevant AirMOSS SAR data collection depending on the year. The SAR science data used to produce the L2/3 RZSM estimates was collected between September 2012 and September 2015. Each individual L4 RZSM delivery will be based on a reprocessing of the entire L4 RZSM product back to a September 1, 2012 baseline date (using all available AirMOSS L2/3 RZSM estimates acquired at a particular site during the entire mission life).

2.3.2e Error Estimates

Product error estimates are derived via the comparison of L4 RZSM algorithm output with independent, ground-based soil moisture observations acquired within individual AirMOSS study sites.

3. Physics of the Problem

3.1 System Model

The proposed L4 RZSM algorithm is based on a hydrologic model that makes continuous predictions about the evolution of soil moisture states in response meteorological drivers such as rainfall and incident radiation. The conceptual basis of these predictions are conservation principles known to operate in nature. In essence, the model is designed to conserve both water (converting precipitation inputs into evaporation, runoff, and storage change) and energy. Given realistic forcing and parameterizations of water and energy flux processes, these conservation principles ensure at least some first-order reliability in the model as an interpolator of temporally sparse soil moisture observations.

3.2 Radiative Transfer and Backscatter

Not applicable.

3.3 Parameter and Model Uncertainties

The hydrologic model at the core of the L4 RZSM algorithm requires a diverse set of soil, geologic and vegetation parameters to accurately capture the flux and storage of water within (and out of) a hydrologic catchment. Failure to accurately specify these parameters can degrade the accuracy of model predictions. While all efforts will be made to minimize calibration error by considering all available stream flow and soil hydraulic property information, model calibration errors will likely persist and impact the accuracy of L4 RZSM estimates.

In addition, the quality of the ancillary forcing data will also impact the model uncertainties. In particular, three of the ten AirMOSS sites are outside the United States: BERMS, Chamela, and La Selva. hydrologic model output (and thus subsequent L4 RZSM estimates) at these sites are likely be of relative lower quality.

4. L4 RZSM Algorithm

4.1 Theoretical Description

4.1.1. Baseline and Option Algorithm Overview

As illustrated in Figure 1, the L4 RZSM algorithm consists of three components:

(1) a hydrologic model,

(2) a bilinear integration procedure,

(3) and pre- and post-processing stages required to transform RZSM products between specified AirMOSS project formats and the format requirements of the hydrologic model.

Each element is described in greater detail within the following sub-sections.

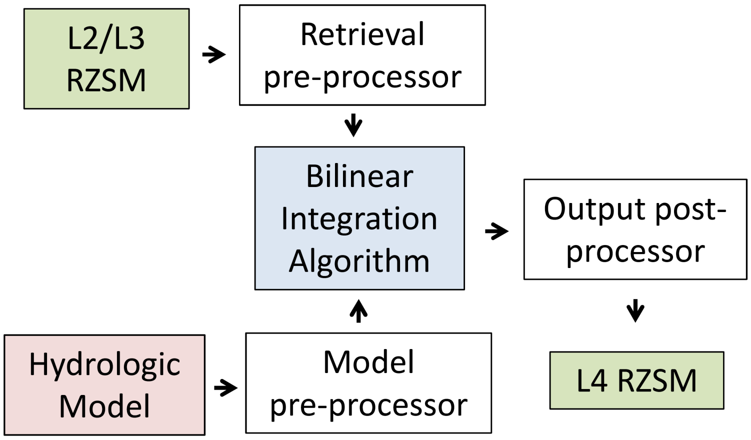


Figure 1. Schematic of the AirMOSS L4 RZSM processing system.

4.1.2 Mathematical Description of the Algorithm

4.1.2a Hydrologic Model

The Penn State Integrated Hydrologic Model (PIHM) is a multi-process, multi-scale hydrologic model where the major hydrological processes (illustrated in Figure 1) are fully coupled using the semi-discrete finite volume method [Qu and Duffy, 2007; Yu et al., 2014] Table 2 shows all these processes along with the original and reduced governing equations. For channel routing and overland flow which is governed by the St. Venant equations, both the kinematic wave and the diffusion wave approximations are included. For saturated groundwater flow, the 2-D Dupuit approximation is applied. For unsaturated flow, either shallow groundwater assumption in which unsaturated soil moisture is dependent on groundwater level or 1-D vertically integrated form of Richards's equation can be applied. From physical arguments, it is necessary to fully couple channel routing, overland flow and subsurface flow in the ordinary differential equations (ODE) solver. For those processes whose governing equations are partial differential equations (PDE), PIHM first discretizes in space via the finite volume method. This results in a system of ODEs representing those processes within the control volume. Within the same control volume, a local ODE system is formed for the complete dynamics of the finite volume by combining other processes whose governing equations are ODE's, (e.g. the snow accumulation and melt process),. After assembling the local ODE system throughout the entire domain, the global ODE system is formed and solved by a state-of-art ODE solver. Snowmelt, vegetation and evapotranspiration are assumed to be weakly coupled. That is, these processes are calculated at the end of each time step, which is automatically selected within a user specified range in the ODE solver.

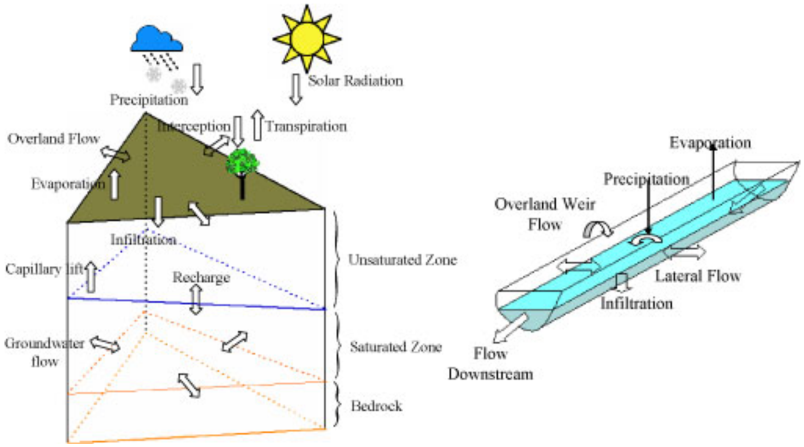
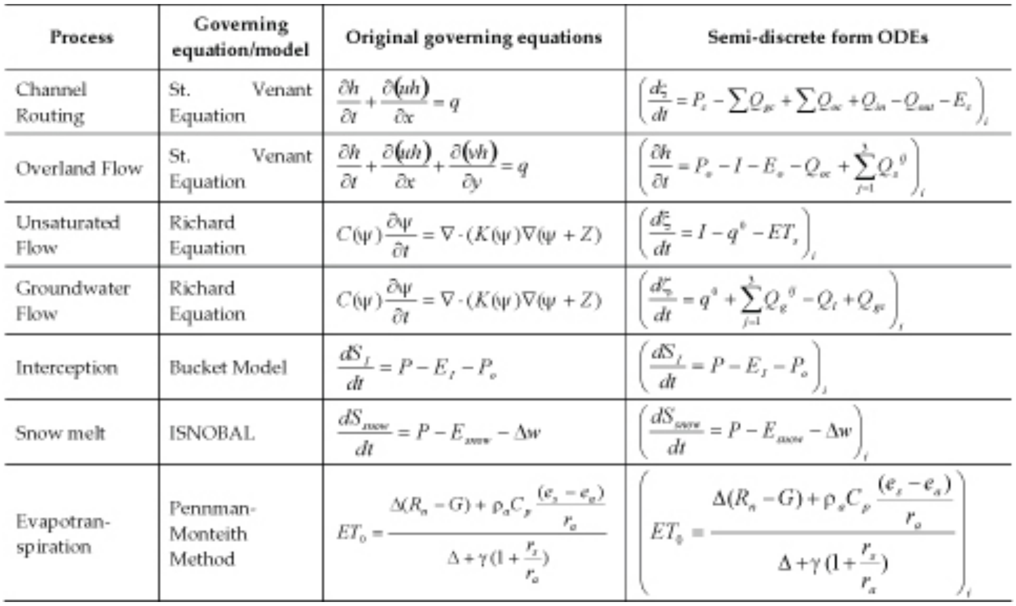


Figure 2. Schematic of physical processes represented in PIHM.

PIHM has recently been modified to calculate root-zone soil moisture within four discrete layers of the unsaturated zone using a semi-discrete numerical version of Richard's equation and a vertically-varying root density profile [Yu et al., 2014]. Note that flow within the root-zone remains 1-D (vertical-only) and 2-D (i.e. lateral flow) aspects of PIHM are confined to the saturated zone.



Process	Governing equation/model	Original governing equations	Semi-discrete form ODEs
Channel Routing	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} = q$	$\left(\frac{d\zeta}{dt} = P_c - \sum Q_{gc} + \sum Q_{oc} + Q_{in} - Q_{out} - E_c\right)_i$
Overland Flow	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} + \frac{\partial (vh)}{\partial y} = q$	$\left(\frac{\partial h}{\partial t} = P_o - I - E_o - Q_{oc} + \sum_{j=1}^{3} Q_s^{ij}\right)_i$
Unsaturated Flow	Richard Equation	$C(\psi)\frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi)\nabla(\psi + Z))$	$\left(\frac{d\xi}{dt} = I - q^0 - ET_s\right)_i$
Groundwater Flow	Richard Equation	$C(\psi)\frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi)\nabla(\psi + Z))$	$\left(\frac{d\zeta}{dt} = q^0 + \sum_{j=1}^{3} Q_g^{ij} - Q_c + Q_{gc}\right)_i$
Interception	Bucket Model	$\frac{dS_I}{dt} = P - E_I - P_o$	$\left(\frac{dS_I}{dt} = P - E_I - P_o\right)_i$
Snow melt	ISNOBAL	$\frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w$	$\left(\frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w\right)_i$
Evapotranspiration	Pennman-Monteith Method	$ET_0 = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})}$	$\left(ET_0 = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})}\right)_i$

Table 2. Mathematical summary of hydrologic processes represented within the PIHM model.

The spatial domain of individual PIHM simulations is always defined by a hydrologic catchment which is resolved via a triangular irregular network (TIN) mesh describing fundamental model elements. There, each 25 km x 100 km SAR imaged area (over which AirMOSS L2/3 RZSM retrievals are obtained) was decomposed into 30-50 hydrologic catchments, which, in turn, will be topographically-characterized using a TIN mesh. Every attempt was made to maximize the completeness of coverage over the SAR imaged area. However, due to computational constraints and the irregular shape of catchments, complete coverage could not be guaranteed. The area of individual TIN elements is non-uniform. However, the goal of the mesh generation process will be to obtain a TIN which an average element area on the order of 100^2 m^2 to 200^2 m^2.

4.1.2b Bilinear integration technique

As motivated above, the integration of AirMOSS L2/3 RZSM retrievals in the PIHM model is based on a relatively simple bilinear-interpolation integration approach. Figure 3 provides a step-by-step schematic for the approach for a single PIHM soil layer.

The background for the approach are PIHM-generated soil moisture predictions acquired for multiple soil moisture layers (Figure 3a). These time-series values are referred to as $Y^j(t_i)$ where t_i refers to discrete acquisition times for the *i*th L2/3 RZSM retrieval and *j* indexes a particular PIHM model layer of interest (see Section 2.3.2a for a description of these layers). Likewise, at discrete times t_i, L2/3 RZSM soil moisture profiles estimates are vertically-integrated to match the vertical spatial support for PIHM soil moisture layer *j* ($X^j(t_i)$; Figure 3b).

For each pixel in every L2/R RZSM image, an off-set (*O*) is calculated between PIHM-modeled and L2/3 RZSM retrieved soil moisture for layer *j* (Figure 3c):

$$O^j(t_i) = X^j(t_i) - Y^j(t_i). \qquad (1)$$

Next, these (temporally-discrete) offsets $O^j(t_\mathrm{i})$ are temporally-interpolated (on a layer-by-layer basis) to produce a continuously-interpolated product $\hat{O}^j(t)$ using a simple, one-dimensional bilinear-interpolation operator **B** applied in time (Figure 3d):

$$\hat{O}^j(t) = \mathbf{B}[O^j(t_i)]. \qquad (2)$$

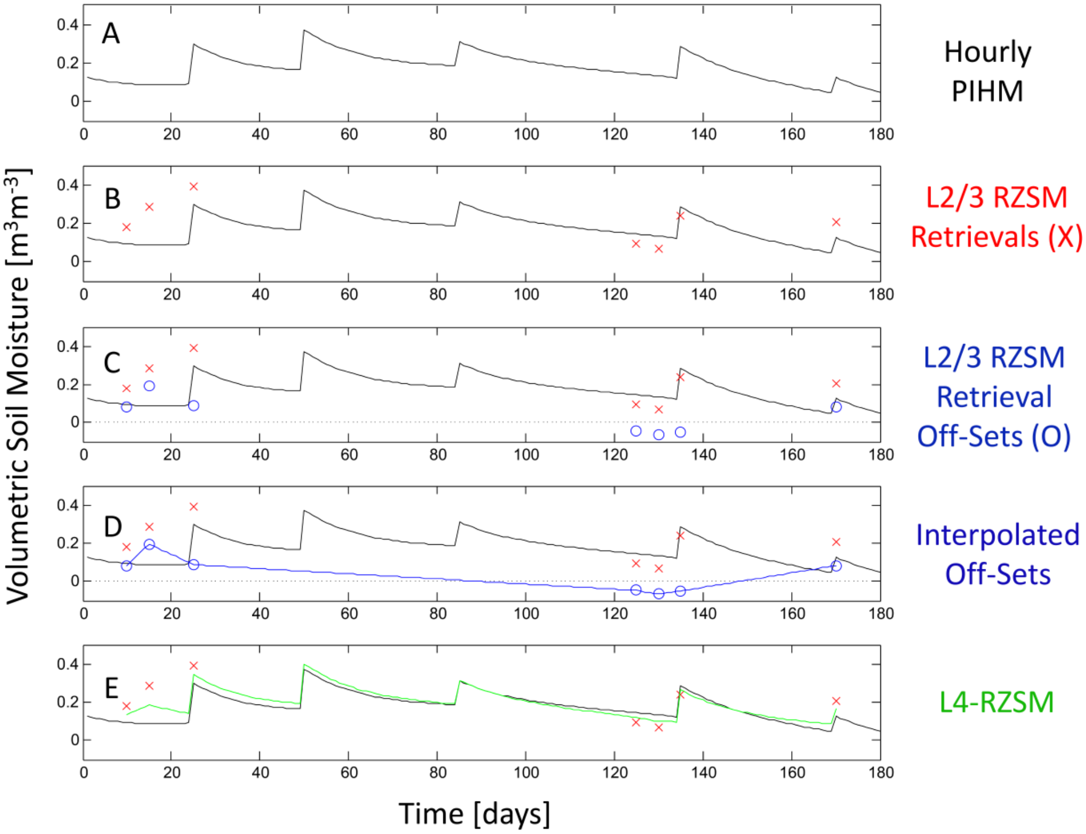


Figure 3. Schematic of the bilinear integration scheme for the case $w = 0.5$.

Finally, weighted averaging is applied to merge $\hat{O}^j(t)$ with comparable continuous soil moisture predictions acquired from the PIHM model $Y^j(t)$ to obtain a continuous L4 RZSM prediction $Z^j(t)$ corresponding to soil layer j:

$$Z^j(t) \; = w^j \, \hat{O}^j(t) \; + (1 - w^j) \; Y^j(t) \qquad (3)$$

where the weighting factors w^j are defined as the ratio between the error variance for $Y^j(t_\mathrm{i})$ and the sum of the error variance of $Y^j(t_\mathrm{i})$ and the error variance of $X^j(t_\mathrm{i})$ (Figure 3e). These error variances are layer-dependent and based on the comparison of both $Y^j(t_\mathrm{i})$ and the error variance of $X^j(t_\mathrm{i})$ against independently-acquired ground based observations at each PIHM study site. However, due to difficulties in measuring such errors, derived values of w^j are applied as spatial and temporal constants across all AirMOSS study sites. Table 3 provides values of w^j applied for each of the four PIHM soil layers.

PIHM Layer	**Depth (cm)**	**Weight (w)**
1	0-10	0.055
2	10-40	0.085
3	40-100	0.033
4	100-200	0.000

Table 3. Values of derived weights (w^j) applied in (3).

The approach summarized in (1-3) is independently applied to each of the top four PIHM soil moisture layers (i.e., 0-10 cm, 10-40 cm, 40-100 cm, and 100-200 cm) within each L2/3 RZSM spatial grid cell. However, note that, due to the sharp reduction of L2/3 RZSM retrievals at depths below 100 cm, no updating is actually applied to the fourth PIHM soil moisture layer (i.e., $w^4 = 0$; Table 3). In addition, due to site-specific quality issues, the third layer was also not updated at the Tonzi Ranch site.

The interpolation step in (2) is preformed only between successive L2/3 RZSM pairs for a given L2/3 pixel. Therefore, no L4 RZSM value is provided for pixels where less than two retrievals are available (within the entire Sept. 2012 to Sept. 2015 L2/3 RZSM data acquisition period). In addition, no value is provided for temporal periods outside of the first and last available L2/3 RZSM retrieval for a given pixel (see Figure 3e).

This approach to missing L2/3 values is fairly conservative in that it presumes the existence of (essentially) spatially-continuous AirMOSS L2/3 RZSM fields. This assumption is reasonably valid for L2/3 RZSM acquisitions at most sites. The exception being retrievals at the Tonzi Ranch site which contain a large number of missing values due to site-specific data quality and radio frequency interference issues. Therefore, at the Tonzi Ranch alone, a modified approach is followed where missing L2/3 RZSM retrievals are replaced with the spatial average of (non-missing) L2/3 RZSM retrievals whose pixel center is located within the same PIHM TIN. For cases in which the extent of missing L2/3 RZSM pixels spans two (or more) TINs, the spatial average of (non-missing) L2/3 RZSM values in both TINs is used. This results in a significantly more continuous L4 RZSM data product at the Tonzi Ranch site.

4.1.2c Pre- and post-processing

The L4 RZSM algorithm also requires specific pre- and post-processing algorithms to resolve differences in data format between PIHM input requirements and the specified format of the AirMOSS L2/3 and L4 data products. In particular, the "retrieval pre-processing" step in Figure 1 consists of transforming the HDF5 format of the L2/3 RZSM product into the tabular ascii format required by PIHM. Likewise, the "model pre-processing" step involves re-sampling of the (irregular TIN-based) PIHM soil moisture predictions onto a 3-arsec grid that is consistent with the L2/3 RZSM retrievals.

Similarly, the “post-processing” step in Figure 1 consists of re-sampling the (irregular) TIN mesh back onto a regular 3-arcsec square grid and converting re-sampled RZSM fields back into HDF5. The post-processing step is also used to define and structure required meta-data into the final AirMOSS L4 RZSM HDF5 file.

Each HDF5 file contains hourly soil moisture values for the entire day. The data projection is mercator using the WGS84 datum. Each HDF file has a data group of browse layer (browse), latitude grid (lats), longitude grid (lons), 0-10 cm layer volumetric soil moisture dataset (sm1), 10-40cm layer volumetric soil moisture dataset (sm2), a 40-100 cm layer volumetric soil moisture dataset (sm3), and a 100-200 cm soil moisture dataset (sm4). Furthermore, each of the layer volumetric soil moisture datasets consist of up of 24 pages where each page represent respective hourly soil moisture grid values at 0-23 UTC hours for that day, for each soil layer.

However, due to vertical restrictions in the accuracy of the L2/3 RZSM retrievals, the L4 RZSM soil moisture for the 100-200 cm layer (sm4) is not included in the product at all sites. In addition, the 40-100 cm layer soil moisture (sm3) was not produced for the Tonzi Ranch AirMOSS site. The L4 RZSM values for these layers are set to missing data values.

The following naming template is utilized with L4RZSM product to identify the biome and calendar time information: L4RZSM_[Sitename]_[Date in YYYYMMDD]_[File version ID].h5 as in:“L4RZSM_Moisst_20130303_v5.h5”. Note that AirMOSS “Sitename” labels and all grid specifications are equivalent to those utilized in the L2/3 RZSM product.

4.1.3 Ancillary Data Requirements

Meteorological input forcing includes time series of: precipitation, temperature, relative humidity, wind velocity, solar radiation, vapor pressure, leaf area index, roughness length, interception storage factor and snow melt factor. As a baseline these products are obtained from the hourly North American Land Data Assimilation Version 2 (NLDAS-2) forcing dataset. Local tower-based meteorology observations and topographic corrections to solar radiation, precipitation and air temperature will also be considered. For sites outside of the NLDAS-2 domain (i.e., the La Selva, BERMS and Chamela sites), baseline forcing data are acquired from the Modern-Era Retrospective analysis for Research and Applications (MERRA). In particular, an off-line, global land replay of the system (MERRA-Land) based on a MERRA (0.50^o x 0.66^o, hourly) atmospheric reanalysis plus precipitation corrections based on daily, 0.50^o Climate Prediction Center (CPC)-Global-Unified rain gauge analysis.

PIHM simulates hydrologic state variables in space and time using pedologic, topographic, geologic and vegetation information derived from: a digital elevation model (DEM), a bedrock elevation map, a soil texture map, and a land cover classification map. Watershed delineation, stream definition and the horizontal variations in the depth of constraining layers are obtained from the DEM and bedrock map and used to obtain a TIN mesh. Parameter lookup tables are defined to assign physical parameters to individual TIN elements based on characteristics identified in these maps. Whenever possible, PIHM parameters were defined to be consistent with ancillary data defined for the AirMOSS L2/3 RZSM and the L4 NEE product algorithms.

In particular, soil texture, organic matter and bulk density information is required to estimate soil and aquifer hydraulic parameters. PIHM also requires vertical soil hydraulic conductivity, porosity, residual porosity, horizontal area fraction of macro-pore, vertical area fraction of macro-pore, macro-pore depth, macro-pore horizontal hydraulic conductivity, macro-pore vertical hydraulic conductivity, Van Genuchten alpha and beta parameters. Baseline values for these soil parameter were based on the Soil Survey Geographic Database (SSURGO) soil classification; however, whenever possible, this baseline will be updated using actual *in situ* soil texture and hydraulic parameter measurements. In particular, limited water retention point data available at select sites were incorporated into the Rosetta model [Schaap et al., 2001] to provide improved prior estimates of soil hydraulic parameters.

Land cover and vegetation parameters inputs include: land cover classes and their maximum leaf area index, minimum stomatal resistance, reference stomatal resistance, albedo, vegetation fraction, Manning's roughness coefficient and root zone depth. As a baseline, these parameters were defined based on land cover classifications contained within the National Land Cover Database (NLCD).

Parameter	**Hydrological Process**
Horizontal Matrix Conductivity (Geological)	Subsurface flow
Vertical Matrix Conductivity (Geological)	Subsurface flow
Horizontal Macropore Conductivity	Subsurface flow
Vertical Macropore Conductivity	Subsurface flow
Macropore Depth	Subsurface flow
Infiltration Rate	Infiltration
Porosity	Subsurface flow and recharge
Van Genuchten (alpha & beta)	Subsurface flow and recharge
River Manning Roughness	Channel routing
River Bed Conductivity (Horizontal)	Channel routing
River Bed Conductivity (Vertical)	Channel routing

Table 4. PIHM model parameters calibrated via stream flow comparisons.

In addition, a subset of PIHM parameters impacting key hydrologic processes (see Table 4) were further modified via calibration against available stream flow observations. In particular, for each US AirMOSS site, USGS stream flow stations were mapped onto the biome modeling domain to select a single candidate watershed for stream flow calibration. Using forcing data and USGS stream flow data, a temporal calibration window (between 2 and 4 weeks in length) was chosen which demonstrated a clear stream flow recession period in response to a significant precipitation event. Each (spatially-distributed) parameter in Table 4 was assigned a spatially constant multiplying factor which scaled its prior value. The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) was then used to select a unique set of these multiplying factors which minimized the mean-squared-difference between PIHM and observations stream flow within this period. Once defined for the (single) calibrated basin, the same constant multiplying factors were applied to spatially-distributed parameters within the entire study site. Finally, limited *ad hoc* calibration was applied to match ground-based soil moisture observations acquired at various ground sites.

The project level requirements call for all calibration and ancillary data used to generate the standard data products to be delivered to the Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics (ORNL-DAAC) by investigation close out. However, data merely taken from standard data bases have not been re-archived.

4.2 Practical Considerations

4.2.1 Numerical Computation Considerations

The L4 RZSM algorithm described above is computationally intensive. PIHM, and other models like it, require fine spatial resolutions and short computation steps (on the order of minutes) in order to maintain computational stability. In order to meet this computational challenge, the L4 RZSM algorithm was run on the NASA Pleiades Supercomputer to minimize wall clock run times.

4.2.2 Programming/Procedural Considerations

PIHM is a multi-process, multi-scale hydrologic model coded in the C computing language. It is a platform independent model. PIHMGIS is a geographic information system (GIS) interface to the PIHM model and is a useful tool to create PIHM input datasets and execute PIHM runs. Both PIHMGIS and ArcGIS are utilized to create required input datasets.

The entire Level 4 RZSM algorithm is based on Perl and Matlab scripts from which individual PIHM model simulations are called in C. In particular, a Perl script executes a single PIHM simulation for a given set of parameters and calculates PIHM RZSM output. Likewise, a Matlab script collects this output, compares it to corresponding L2/3 RZSM retrievals and generates an updated set of parameters by calling an automated calibration algorithm (e.g., SCE-UA). These parameters are then passed back to the PERL script for the next PIHM model iteration.

4.2.3 Ancillary Data Availability/Continuity

All ancillary data indentified in Section 4.1.3 are publically available. Ongoing generation of the NLDAS-2 meteorological forcing data set is expected to be maintained throughout the AirMOSS project.

4.2.4 Calibration and Validation

Validation of the L4 RZSM product is primarily be against profile soil moisture observations acquired with ground instrumentation installed at each AirMOSS study site. The validation strategy for the L4 RZSM algorithm parallels the activities described by the L2/3 RZSM ATBD. The single key difference is that L4 RZSM products are evaluated continuously in time while validation of the L2/3 RZSM product is limited only to intermittent retrieval acquisition times. It should be noted that there is no program-level requirement on the accuracy of the L4 RZSM product.

For AirMOSS we have four sources of soil moisture data observations for L2/3 and L4 RZSM validation. Firstly, we will have the AirMOSS L2 In Ground Sensor Measurements (IGSM). These data are produced by the in situ soil moisture measurement equipment installed at most of the AirMOSS sites by the AirMOSS project. At some sites, such as La Selva, no in situ equipment was deployed as part of AirMOSS because similar instruments were already in place from other projects. Secondly, at the Tonzi Ranch and the MOISST sites we have data from the Soil moisture Sensing Controller And oPtimal Estimator (SoilSCAPE) arrays. Thirdly, at the majority of the AirMOSS sites soil moisture probes have been deployed as part of the Cosmic-ray Soil Moisture Observing System (COSMOS). Fourthly, during the campaigns when SAR over flights of the sites are planned, various members of the AirMOSS science team will be in the field collecting ground truth measurements. Soil moisture measurements using time domain reflectometers are planned as part of these campaigns. The spatial distribution of these measurements over the 25 km x 100 km area imaged by the SAR may be poor due to access issues.

4.2.5 Quality Control and Diagnostics

The L2/3 RZSM ATB specifies that any pixel characterized as: (i) numerically non-covergent, (ii) possessing excessive topographic slope, (iii) possessing inappropriate land cover for retrieval, (iv) possessing inappropriate soil texture for retrieval or (v) lacking adequate radar coverage for retrieval will be masked from L2/3 processing. All such pixels are discarded during the calculation of the offsets in Eq. (1) in the L4 RZSM algorithm. A second set of QC steps are additional “online” rules that exclude the inclusion of L2/3 RZSM retrievals into Eqs. (1-3) if (i) rain is falling, (ii) the soil is frozen, or (iii) the ground is fully or partly covered with snow.

4.2.6 Exception Handling

See sections 4.1.2b and 4.2.5.

4.2.7 Interface Assumptions

Inputs Interface

In addition to ancillary data requirements described in Section 4.1.3, the L4 RZSM algorithm requires the AirMOSS L2/3 RZSM product as input. The pre- processing required to interface the L2/3 RZSM product with the L4 RZSM algorithm are described in Section 4.1.2c.

Output Interface

L4 RZSM output described in Section 2.3.2a will be ingested by the AirMOSS L4 NEE algorithm. The post-processing required to convert calibrated PIHM predictions into a standardized HDF data format is described in Section 4.1.2c.

4.2.8 Test Procedures

Prior to the initial delivery of the AirMOSS L2/3 product in late 2012, the L4 RZSM algorithm was tested using proxy L2/3 RZSM products acquired using P-band AIRSAR observations.

4.2.9 Algorithm Baseline Selection

No option algorithms were carried; therefore, no baseline selection process was required.

5. Acknowledgements

The AirMOSS L4 team would like to thank Prof. Richard Cuenca (Oregon State University) for help with the soil hydrologic parameterization at the AirMOSS sites and Prof. Christopher Duffy (Pennsylvania State University) for consultation on implementation of the PIHM model.

6. Prototype Data Product Specifications

Example Meta data for L4 RZSM data file:

```
/ (96)
   Group size = 7
   Number of attributes = 3
      Datum = WGS84
      NOTE = Datasets (sm1, sm2, sm3) contains Level2/3 updates.
      Projection = LAT/LONG D000

browse (1655287992, 2)
   64-bit floating-point,    1650 x 3300
   Number of attributes = 0

lats (1611725104, 2)
```

```
    64-bit floating-point,    3300 x 1650
    Number of attributes = 0
lons (1568162216, 2)
    64-bit floating-point,    3300 x 1650
    Number of attributes = 0

sm1 (800, 2)
    32-bit floating-point,    1650 x 3300 x 24
    Number of attributes = 4
        Dataset description = 24 layers containing hourly soil moisture at depth 0-10cm. Contains Level2/3 updates.
        No Data = NaN
        Soil moisture unit = vol./vol.
        Time Zone = UTC

sm2 (1400, 2)
    32-bit floating-point,    1650 x 3300 x 24
    Number of attributes = 4
        No Data = NaN
        Soil moisture unit = vol./vol.
        Time Zone = UTC
        dataset description = 24 layers containing hourly soil moisture at depth 10-40cm. Contains Level2/3 updates.

sm3 (1672, 2)
    32-bit floating-point,    1650 x 3300 x 24
    Number of attributes = 4
        No Data = NaN
        Soil moisture unit = vol./vol.
        Time Zone = UTC
        dataset description = 24 layers containing hourly soil moisture at depth 40-100cm. Contains Level2/3 updates.

sm4 (1944, 2)
    32-bit floating-point,    1650 x 3300 x 24
    Number of attributes = 0
```

7. References

Schaap, M.G., Leij, F.J., and van Genuchten, M.T. (2001) Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology, 251:163-176.

Qu, Y., and Duffy, C.J. (2007) A semidiscrete finite volume formulation for multiprocess watershed simulation. Water Resources Research, 43(8), W08419.

Yu, X., Duffy, C., Kaye, J., Crow, W., Bhatt, G. and Shi, Y. (2014) Watershed Reanalysis of Water and Carbon Cycle Models at a Critical Zone Observatory, in Remote Sensing of the Terrestrial Water Cycle (eds. V. Lakshmi, D. Alsdorf, M. Anderson, S. Biancamaria, M. Cosh, J. Entin, G. Huffman, W. Kustas, P. van Oevelen, T. Painter, J. Parajka, M. Rodell and C. Rüdiger), John Wiley & Sons, Inc, Hoboken, NJ. doi: 10.1002/9781118872086.ch31.