A Method for Retrieving High-Resolution Surface Soil Moisture From Hydros L-Band Radiometer and Radar Observations

Xiwu Zhan, Paul R. Houser, Jeffrey P. Walker, and Wade T. Crow, Member, IEEE

Abstract—NASA’s Earth System Science Pathfinder Hydropheric States (Hydros) mission will provide the first global scale space-borne observations of Earth’s soil moisture using both L-band microwave radiometer and radar technologies. In preparation for the Hydros mission, an observation system simulation experiment (OSSE) has been conducted. As a part of this OSSE, the potential for retrieving useful surface soil moisture at spatial resolutions of 9 and 3 km was explored. The approach involved optimally merging relatively accurate 36-km radiometer brightness temperature and relatively noisy 3-km radar backscatter cross section observations using a Bayesian method. Based on the Hydros OSSE data sets with low and high noises added to the simulated observations or model parameters, the Bayesian method performed better than direct inversion of either the brightness temperature or radar backscatter observations alone. The root-mean-square errors of 9-km soil moisture retrievals from the Bayesian merging method were reduced by 0.5 % vol/vol and 1.4 % vol/vol from the errors of direct radar inversions for the entire OSSE domain of all 34 consecutive days for the low and high noise data sets, respectively. Improvement in soil moisture estimates using the Bayesian merging method over the direct inversions of radar or radiometer data were even more significant for soil moisture retrieval at 3-km resolution. However, to address the representativeness of these results at the global and multyear scales, further performance comparison studies are needed, particularly with actual field data.

Index Terms—Microwave remote sensing, observing system simulation experiment, radar measurements, soil moisture, space-borne radiometry.

I. INTRODUCTION

Soil moisture is a critical hydropheric state variable that often limits the exchanges of water and energy between the atmosphere and land surface, controls the partitioning of rainfall among evaporation, infiltration and runoff, and impacts vegetation photosynthetic rate and soil microbiologic respiratory activities. Accurate measurement of this variable across the global land surface is thus required for global water, energy, and carbon cycle sciences and applications. Remote sensing technology has the potential to provide global information on spatial and temporal soil moisture variation; something that is not possible using sparsely distributed in situ point measurements. Thus, the development of a space-borne soil moisture remote sensing capability is a high priority [1], [2].

Microwave soil moisture remote sensing has been explored for several decades [3], [4]. The fundamental principle of microwave soil moisture remote sensing is the large contrast in dielectric properties of water and soil; water has a relative complex dielectric constant of about 80 for the real part as compared to about 3.5 for dry soil. Thus, the real part of relative dielectric constant for wet soil can be around 40. This large dielectric constant difference between wet and dry soil correspondingly impacts the soil emissivity (passive microwave) and soil reflectivity (active microwave) and, hence, the thermal microwave emission (brightness temperature) and backscattering cross section from the land surface, respectively.

It has been shown that a passive microwave radiometer with less than 1-K radio brightness noise sensitivity can measure near-surface soil moisture with a root-mean-square error of 1%–2% vol/vol for bare soil [5]. However, this 1-K radio brightness noise sensitivity can only be obtained for space-borne radiometer footprints larger than 40 km with current satellite antenna technology [6], [7], greatly limiting their applicability.

Experiments using active microwave (radar) have indicated that 3.5% vol/vol soil moisture retrieval accuracy for spatial resolutions down to 1 km may be achievable for soil surfaces with vegetation cover shorter than 15 cm [8]–[10]. However, because of the greater impact of vegetation, surface roughness, and topography on radar signals than on radiometer observations, typical root-mean-square error values of such fine spatial resolution soil moisture retrievals from radar are significantly higher [9].

To overcome the individual limitations of the two approaches (passive radiometric remote sensing has higher accuracy but lower spatial resolution and active radar remote sensing has higher spatial resolution but lower accuracy), the NASA Earth System Science Pathfinder Hydropheric States (Hydros) mission is combining these two technologies. The Hydros mission will use both an approximately 40-km L-band microwave radiometer and 3-km radar to provide global scale space-borne land surface soil moisture observations at three-day repeat intervals [11]. Hydros will provide coarse, fine, and medium resolution soil moisture products. The coarse resolution (40 km) product will be derived from the radiometer...
land surface brightness temperature observations using both single-channel methods with ancillary data to correct for the impacts of soil temperature, vegetation moisture content, and surface roughness, and multichannel methods with reduced reliance on ancillary data [12]. The fine resolution (3 km) product will be derived from the radar observations [11], while the medium resolution (10 km) product will be derived from combining the coarse resolution radiometer and fine resolution radar observations [11]. This paper describes a Bayesian approach to merging these radiometer and radar observations.

This Bayesian approach optimally merges both relatively accurate coarse resolution radiometer and relatively noisy high resolution radar observation data sources with a preliminary soil moisture background guess. The preliminary background field used in this demonstration is a single channel passive microwave radiometer inversion. For simplification, our observing system simulation experiment (OSSE) adopts an easily nested fine, medium and coarse resolution grid of 3, 9, and 36 km, respectively. Moreover, we assumed square, nonoverlapping radiometer and radar footprints. However, Hydros will actually have overlapping conically scanned elliptical footprints that could provide even more constraining data for higher resolution retrievals [13].

II. HYDROS OSSE DATA SET

The data set used in this paper was generated in the Hydros OSSE. The OSSE was designed to mimic as closely as possible the specific Hydros sensor and orbital characteristics to prepare Hydros algorithms for its planned 2010 launch. The OSSE includes four elements: 1) a land surface model to generate 1-km resolution geophysical data fields; 2) a microwave emission and backscatter model (MEBM) to generate radiometric brightness temperature \( T_B \) and radar backscatter \( \sigma \) simulations at 1 km; 3) aggregation of the generated 1 km \( T_B \) and \( \sigma \) data to Hydros radiometer (36 km) and radar (3 km) footprints; 4) addition of ancillary data errors and soil moisture retrieval from the simulated Hydros observations using candidate methods. Complete OSSE details for radiometer-only soil moisture retrievals were described in [12]. Details relevant to this combined radar and radiometer soil moisture retrieval approach are summarized below.

A. Generation of 1-km Geophysical Data Fields

The TOPLATS hydrological model [14] was used to simulate high-resolution (1 km) geophysical fields such as top 5 cm soil moisture and temperature within the United States Southern Great Plains; the 575 000 km\(^2\) Red-Arkansas River Basin. The simulation was run for 34 days from May 26 (Day of Year 146) to June 29 (Day of Year 179), 1994. Results and methodology for the TOPLATS application to this domain are described in greater detail by [15]. Additional details about the OSSE domain are presented in [12]. The 1-km surface soil moisture data fields from the TOPLATS model nature runs are used to simulate Hydros radar and radiometer observations, and are also aggregated to 3, 9, and 36 km to be used as soil moisture “truth” data in evaluating the OSSE soil moisture retrievals.

B. MEBM

The TOPLATS model derived 1-km soil moisture and temperature fields described above were used to compute radar backscattering and radiometer brightness temperatures at the Hydros frequencies (1.42 GHz for radiometer and 1.26 GHz for radar), polarizations (horizontal and vertical labeled as subscripts \( h \) and \( v \) for radiometer and \( hV \) and \( vV \) for radar), and incident angle \( (40^\circ) \) using the forward MEBM described herein at 1-km spatial resolution.

\[ T_B = \frac{1}{1 - \left( \frac{T_v}{T_s} \right)} \left( \frac{T_s}{T_0} \right) \]

\[ T_V = \frac{1}{1 - \left( \frac{T_v}{T_s} \right)} \left( \frac{T_s}{T_0} \right) \]

where \( T_B \) is radiometer brightness temperature at polarization \( p \) (\( h \) or \( v \)), \( T_s \) is soil effective temperature, and \( T_V \) is vegetation canopy temperature (K). The soil microwave effective temperature \( T_s \) is estimated as the average of the soil skin temperature and the 5-cm soil temperature. For vegetated surfaces the canopy temperature \( T_V \) is assumed to be equal to the soil skin temperature. Symbols \( e_p \), \( \gamma \), and \( \omega \) are the rough soil emissivity, vegetation transmissivity, and single scattering albedo, respectively.

The influence of surface roughness on emissivity is approximated as \( e_p = 1 - \left( 1 - e_{sp} \right) \exp \left( -h \cos^2 \theta \right) \], where the roughness parameter \( h \) (cm) is related to (one tenth of) the root-mean-square surface height and \( e_{sp} \) is the emissivity of the equivalent smooth soil surface at polarization \( p \), determined by the soil dielectric constant and sensor look angle using the Fresnel equations.

The vegetation transmissivity represents the transparency of the vegetation layer to the microwave signal and is related to the vegetation optical depth (opacity) parameter \( \gamma \) by \( \gamma = \exp \left( -\tau / \cos \theta \right) \), where \( \theta \) is the sensor look angle (degrees). An unpolardized nadir vegetation opacity is assumed and is related to the average columnar vegetation water content \( W \) (kg-m\(^{-2}\)) over the pixel by \( \tau = hW \), where the coefficient \( h \) depends on vegetation type [16]. A typical value for the single scattering albedo was obtained from literature for each land cover type.

For water surfaces, the observed brightness temperature was taken as the skin temperature multiplied by the Fresnel smooth surface emissivity for fresh water at the skin temperature [17]; wind-induced surface roughness was neglected.

The land cover classes and the corresponding biophysical parameters used in the Hydros OSSE are listed in Table I. The parameter values for bare soil, crop, and natural vegetation were taken from published literature [18]–[21]. The canopy vegetation water content was obtained from the Advanced Very High Resolution Radiometer (AVHRR) NDVI dataset of the same time period in 1995 and the published relationship in Jackson et al. [18].

Backscatter Model: The radar backscatter model for a vegetation-covered soil is the sum of three components [22]

\[ \sigma^V_{\text{PS}} = \sigma^{\text{scat}}_{\text{PS}} \exp \left( -2\tau \right) + \sigma^V_{\text{PS}} + \sigma^{SV}_{\text{PS}} \]
where $\sigma_{ipq}^s$ represents the total radar scattering cross section, $\sigma_{ipq}^d$ represents the direct scattering cross section of the soil surface modified by the two-way attenuation through vegetation, $\sigma_{ipq}^{s \nu}$ represents the direct scattering cross section of the soil surface, and $\sigma_{ipq}^{s \nu}$ represents the scattering interaction between the soil and vegetation. Subscripts $pq$ represent polarization ($h_1v_1$, or $h_1v$). The scattering components are modeled as soil moisture and vegetation characteristic functions [22]–[24]. The following models give backscattering in power ratio.

The regression model of [24] is used to compute the soil surface scattering as

$$\sigma_{lh}^{h} = 10^{-2.75}, \left( \frac{\cos^{1.5} \theta}{\sin^{0.1} \theta} \right) \cdot 10^{0.025 \varepsilon' \tan \theta} \cdot (ks \cdot \sin \theta)^{1.4} \cdot \lambda^{0.7}$$

$$\sigma_{vh}^{h} = 10^{-2.35}, \left( \frac{\cos^{3} \theta}{\sin^{0.1} \theta} \right) \cdot 10^{0.065 \varepsilon' \tan \theta} \cdot (ks \cdot \sin \theta)^{1.1} \cdot \lambda^{0.7}$$

where $\lambda$ is the wavelength (cm), $k = 2\pi/\lambda$ is the wave number (cm$^{-1}$), $s$ is the surface root-mean-square height (cm), and $\varepsilon'$ is the real part of the dielectric constant. The Oh et al. model [26] is used to relate the co- and cross-polarized terms $\sigma_{vh}^{h}$ and $\sigma_{vh}^{v}$ by

$$\sigma_{vh}^{h} = 0.23 \sqrt{r_{s0}[1 - \exp(-ks)]} \times \sigma_{vh}^{v}$$

where $r_{s0}$ is the Fresnel (smooth surface) reflectivity at nadir.

A simplified single-scattering canopy model [25] parameterized (as in the emission model) in terms of $\omega$ and $\tau$ is used, which represents the vegetation as a collection of randomly oriented disks. This model is expected to provide a reasonable approximation at L-band. The volume scattering expressions are

$$\sigma_{pp}^{v} = 0.744\omega \cos \theta[1 + 0.53\omega\tau_0 - 0.24(\omega\tau_0)^2] \times [1 - \exp\{-2.12\omega\tau_0 \sec \theta\}]$$

$$\sigma_{lv}^{v} = \omega \cos \theta[0.04kw_0 - 0.018(\omega\tau_0)^2] + 0.006(\omega\tau_0)^3] \times [1 - \exp\{-11.7\omega\tau_0 \sec \theta\}]$$

The equivalent surface-volume scattering expressions are

$$\sigma_{pp}^{av} = 1.9\omega \cos \theta[1 + 0.9k\tau_0 + 0.4(\omega\tau_0)^2] \times [1 - \exp\{-1.93\omega\tau_0 \sec \theta\}] \times \exp\{-0.84(k\tau_0)^2 \cos \theta\} r_{sp}$$

$$\sigma_{lv}^{av} = 0.013\omega \cos \theta[1 + 7.85\omega\tau_0 + 7.9(\omega\tau_0)^2] \times [1 - \exp\{-6.9\omega\tau_0 \sec \theta\}] \times \exp\{-1.02^\tau_0 \cos \theta\}$$

$$\times \exp\{-2.9(k\tau_0)^2 \cos \theta\} \left( \frac{T_{sv} + T_{sh}}{2} \right)$$

where $r_{sp}$ and $r_{sh}$ are the smooth surface reflectivity at $v$ and $h$ polarization, respectively. For water surfaces, the radar backscatter is not computed theoretically but is assigned experimentally observed values. The following values observed by the PALS airborne radar in the Southern Great Plains 1999 experiments [27] are used: $\sigma_{sv} = -23$ dB, $\sigma_{sh} = -27$ dB, and $\sigma_{lv} = -38$ dB.

### Dielectric Model:

The real part of the soil dielectric constant $\varepsilon'$ required by the Fresnel smooth surface reflectivities $r_{sv}$ and $r_{sh}$, is computed as a function of surface soil moisture and sand clay fractions. The computation is done using the empirical relations of Dobson et al. [23] and the correction factor of Peplinski et al. [28]. Parameters required in these models include the soil bulk density $\rho_b$ and the specific density $\rho_s$ of solid soil particles. These were assigned uniform values of $\rho_b = 1.40$ g cm$^{-3}$ and $\rho_s = 2.66$ g cm$^{-3}$, respectively.

### C. Aggregation

A 1-, 3-, 9-, and 36-km nested grid was defined for this simulation study as shown in Fig. 1. The 1-km grid represents the resolution of the underlying geophysical data fields (TOPLATS model 1-km nature run) while the 3-km grid represents the resultant radar footprint measurements to be used by Hydros to minimize the signal noise in 1-km observations; the 36-km grid represents the 40-km Hydros radiometer footprint measurements; and the 9-km grid represents the 10-km combined radiometer-radar Hydros soil moisture product resolution. The 9- and 36-km resolutions were used because they represent integer multiples of the radar data and combined product, respectively, making the OSSE simpler in this initial demonstration.

The 36-km radiometer observations were computed by linearly averaging the simulated 1-km brightness temperatures in each 36-km grid cell, and the 3-km radar observations were generated similarly by averaging the underlying 1-km radar backscatter fields (in power ratio not decibels). For soil moisture

### Table 1

**LAND COVER CLASSIFICATION KEY AND ASSOCIATED ROUGHNESS AND VEGETATION PARAMETERS**

<table>
<thead>
<tr>
<th>Class</th>
<th>Category Name</th>
<th>$h$ [cm]</th>
<th>$\omega$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Crop/mixed farming</td>
<td>0.15</td>
<td>0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>Short grass</td>
<td>0.10</td>
<td>0.05</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>Evergreen needleleaf tree</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>4</td>
<td>Deciduous needleleaf tree</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>5</td>
<td>Deciduous broadleaf tree</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>Evergreen broadleaf tree</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>7</td>
<td>Tall grass</td>
<td>0.10</td>
<td>0.05</td>
<td>0.20</td>
</tr>
<tr>
<td>8</td>
<td>Desert</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>Tundra</td>
<td>0.10</td>
<td>0.05</td>
<td>0.20</td>
</tr>
<tr>
<td>10</td>
<td>Irrigated crop</td>
<td>0.15</td>
<td>0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>11</td>
<td>Semidesert</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>12</td>
<td>Bog or marsh</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>13</td>
<td>Inland water</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>14</td>
<td>Evergreen shrub</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>15</td>
<td>Deciduous shrub</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>16</td>
<td>Mixed woodland</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>17</td>
<td>Short grass/crop</td>
<td>0.12</td>
<td>0.05</td>
<td>0.17</td>
</tr>
<tr>
<td>18</td>
<td>Tall grass/crop</td>
<td>0.12</td>
<td>0.05</td>
<td>0.17</td>
</tr>
<tr>
<td>19</td>
<td>Crop/mixed woodland</td>
<td>0.12</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>20</td>
<td>Crop/evergreen needleleaf tree</td>
<td>0.12</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>21</td>
<td>Crop/deciduous broadleaf tree</td>
<td>0.12</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>22</td>
<td>Irrigated crop/deciduous broadleaf tree</td>
<td>0.12</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>23</td>
<td>Short grass/mixed woodland</td>
<td>0.10</td>
<td>0.08</td>
<td>0.20</td>
</tr>
<tr>
<td>24</td>
<td>Evergreen needleleaf/short grass</td>
<td>0.10</td>
<td>0.08</td>
<td>0.20</td>
</tr>
<tr>
<td>25</td>
<td>Evergreen needleleaf/evergreen broadleaf</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
</tr>
</tbody>
</table>
retrieval and evaluation purposes, the hydrological model soil temperature and moisture outputs, and surface characteristic parameters \((W, S, C, h, \omega; \delta)\) were also linearly averaged from 1 km to fields of grid cell sizes 3, 9, and 36 km. Special care was taken with averaging water pixels: parameters \(W, h, \omega, \) and \(\delta\) were assumed to have values of zero for the water pixels while soil texture parameters and soil moisture were not included in the averaging for water pixels. There was no special temperature treatment. These aggregation rules were considered the most appropriate for simulating actual measurements.

The 36-km radiometric brightness temperature and 3-km radar observations could also be derived from the aggregated 36-km/3-km surface parameters by applying the MEBM equations directly to the 36-km radiometer or 3-km radar footprints. However, the MEBM equations were developed based on point scale field observations, thus, the 1-km emission and backscatter simulations based on these models were believed to be more realistic than the direct 36-km/3-km simulations, so the above approach for simulating the Hydros observations was chosen for this OSSE.

**D. Observations**

Measurement error was included in the OSSE by adding noise to the observation data described above. The radiometer measurement error includes instrument noise (that limits the measurement precision) and calibration error. To simulate the measurement error, spatially independent Gaussian noise with a standard deviation of 1.5 K was added to the 36-km brightness temperature simulations. Additionally, the noise was assumed to be uncorrelated between radiometer polarizations and subsequent days of the simulation. A spatially independent error assumption was appropriate in this case as radiometer footprints were treated separately in the analysis. However, as several radar footprints are used simultaneously for analysis of a single radiometer footprint, more realistic error introduction was required. Radar measurement error depends on the signal-to-noise ratio and the number of independent samples or “looks” averaged in each measurement, as well as calibration error [29]. Because the number of radar “looks” is higher toward the edges from the center of satellite track, the radar measurement errors may be spatially correlated and decrease in magnitude from the center to the edges. To simulate the radar measurement error, a Gaussian noise with zero mean and standard deviation increasing from the edges to the center of track was added to each 3-km radar backscatter (in decibels). Due to limited knowledge about actual Hydros radar noise levels, two scenarios were simulated: low and high radar noise. In the low noise case, the standard deviation was increased from 0.1 dB at the edges to 1.0 dB at the center of the satellite track while for the high radar noise case the standard deviation was increased from 1.0 dB at the edges to 2.0 dB at the center of the satellite track. Table II summarizes the noise characteristics used to simulate radar and radiometer observation errors, including the noises added to the ancillary data of the simulations.

**E. Ancillary Data**

Soil moisture retrieval from the perturbed radiometer and radar observations requires ancillary information on surface roughness, vegetation water content, and surface temperature. Since there are multipolarized observations for each 3-km pixel, it is possible to also retrieve some of these variables simultaneously with soil moisture. However, in this demonstration only soil moisture is retrieved with the surface roughness, vegetation water content and surface temperature prescribed with a certain noise level.

By adding noise to the radar backscatter and radiometer simulations, and the corresponding ancillary data used in subsequent soil moisture retrieval, the resulting data are intended to simulate reality. The noise added to ancillary data is summarized in Table II and described in detail below. While soil and skin temperatures were assumed to be known from ECMWF forecasts with 1.5 K Gaussian error, soil roughness and vegetation water content were considered to be much more uncertain, requiring special consideration.

The roughness parameter for each pixel was assigned according to its land cover type from a classification map as listed in Table I. Considering the large surface roughness variation that exists within each land cover type and the potential land cover classification error, two error perturbation levels were considered. In the low noise scenario, the zero mean Gaussian noise was assumed to have a standard deviation magnitude of 5%
of the roughness parameter while in the high noise case the standard deviation was assumed to be 15% of the roughness parameter.

Vegetation water content ($W$) was computed from the normalized difference vegetation index (NDVI) observed by the AVHRR on a NOAA satellite for the same time period in 1995 as the Hydros OSSE simulations. Because of the potential for significant errors in the NDVI values and the NDVI-$W$ relationship used in the calculation [21], two noise scenarios were also considered for this parameter. In this case, the resultant $W$ map was assumed to have a Gaussian noise with a relative magnitude of 20% of the computed $W$ for the low noise case and 50% of the computed $W$ for the high noise case.

Fig. 2 gives the spatial distribution of the vegetation water content computed from the NDVI map, the roughness parameter, soil texture (sand and clay fractions) and soil moisture for a typical day (day 155 of 1994). The central part of the Red-Arkansas river basins is seen to be dry and less vegetated as compared to the eastern part.

III. BAYESIAN MERGING METHOD

According to the Bayesian Theorem, state vector $[X^a]$ of a system can given by

$$[X^a] = [X^b] + [K][[Z] - h([X^b])]$$  \hspace{1cm} (10)

if their background estimates $[X^b]$ conditioned on observations $[Z]$ are known [30]. In this Bayesian system state update equation

$$[K] = \frac{[P^a[H^T]}{[H][P^a[H^T] + [R]]}$$  \hspace{1cm} (11)

and

$$[H] = \frac{\partial h([X^b])}{\partial [X]}$$  \hspace{1cm} (12)

where the $n \times 1$ system state vectors $[X^a]$ and $[X^b]$ are the system analysis (i.e., the optimal estimate) and background (i.e., the initial estimate) fields, respectively, with $n$ being the number of system state variables (i.e., soil moisture). $[Z]$ is an $m \times 1$ vector that contains the $m$ number of observations. $[K]$ is an
$n \times m$ matrix of analysis weights based on the respective uncertainties of the background states and observations. The observation function $h([X])$ maps the system state space to the observation space where the first derivative (Jacobian $[H]$ of $h([X])$ is an $m \times n$ matrix called the observation operator. $[P]$ is the $n \times n$ error covariance matrix of the background field and $[H]$ is the $m \times m$ error covariance matrix of the observations.

When (10) is applied to a dynamic system and both system states and error covariances are updated with observations and model predictions sequentially, it is called the Kalman filter [31] and the coefficient $[K]$ is called the Kalman gain. The Kalman filter has been used for combining observations with a model forecast in hydrological sciences for a decade [32]–[35]. Before its application to dynamic systems as the Kalman filter system state update equation, (10) was used to merge two Gaussian random variables based on the Bayesian theory. Thus, it can also be used to combine different types of observations for a single time step, in order to find optimal retrievals of the corresponding state variables assuming these observations have Gaussian distributions. The Hydros mission will provide coarse resolution radiometric brightness temperature and finer resolution radar backscatter observations at a range of polarizations. If a coarse resolution soil moisture estimate is treated as the system background $[X^b]$, and the radar backscatter and/or radiometer brightness temperature are treated as system observations $[Z]$, then $[X^a]$ becomes the optimal near-surface soil moisture retrieval. In this case, the observation function $h([X^b])$ is the MEBM described in the previous section.

In the most pure application of (10), soil moisture retrievals can be obtained at the 3-km radar resolution. However, it can also be applied for retrievals at 9- and 36-km directly. In all cases the background soil moisture field $[X^b]$ is taken as a best-guess estimate; an approximation of the 36-km pixel average soil moisture $x_c$ in this application. The observation vector contains the 236-km brightness temperatures from the radiometer ($u$ and $h$ polarizations) and the 3 radar backscatter observations ($hh$, $vv$, and $hv$ polarizations) for each of the 144 3-km pixels, totaling 434 observations and 144 background and analysis states (see Fig. 1). For this implementation see (13)–(16) at the bottom of the page. In these equations, $x_{f,i} (i = 1, 144)$ is the near-surface soil moisture and $\sigma_{pq,i}(1 \leq p < q \leq 144)$ is the $pq$ polarized radar backscatter for the $i$th fine resolution pixel.

The observation function $h([X])$ is the brightness temperature and backscatter computed with the nonlinear MEBM at the background soil moisture value $x_c$. The observation operator $[H]$ multiplied by the state vector $[X]$ yields the brightness temperature and backscatter computed with the linearized MEBM at the background soil moisture value $x_c$.

Fig. 3 is a data flowchart for the above Bayesian method and the traditional direct inversion method to retrieve soil moisture from the OSSE data set. In the absence of observation and model parameter perturbations, it should be possible to retrieve the truth soil moisture field perfectly if the same radiometer and radar MEBM is used in the retrieval process as in the forward modeling at the original 1-km resolution. However, since anticipated errors have been added to the aggregated observations and model parameters, the retrieved soil moisture will not be exactly the same as the truth soil moisture fields underlying the simulation. The purpose of this paper is to demonstrate the Bayesian method and compare results with the direct inversion method when using the same OSSE data set.

This Bayesian method for soil moisture retrieval requires a background soil moisture field and its error covariance to merge with radar and radiometer observations. This background soil moisture can be simply a guess of the 36-km soil moisture and its error variance. Alternatively, the background may be estimated from direct inversion of the radiometer and radar data individually and the variance estimated from the ensemble of background fields. In this study, the inversion of a single polarization radiometer brightness temperature observation ($T_{Eh1}$) is used for the background field. This inverted soil moisture background field was then compared with the 3-km soil moisture inverted from the $hh$ polarized radar backscatter ($\sigma_{hh}$) to evaluate its error covariance. The 144 diagonal elements of matrix $[P]$ were thus assigned the error covariance and the off-diagonal elements set to be zero, which assumes that the 3-km pixel errors are uncorrelated. This assumption is appropriate as the noises added to the radar backscatter simulations and ancillary data can be treated as uncorrelated at the 3-km pixel resolution.

The diagonal elements of the observation error variance matrix $[R]$ were also assigned based on comparison of the truth radiometer and radar observations with radiometer and radar estimates when using the perturbed ancillary data, plus the radiometer and radar perturbations. For the low noise level, the perturbed ancillary data resulted in approximately 1.0 dB av-

\[
[X^b] = [x_{f,1} \quad x_c \quad x_{f,2} \quad x_c \quad \ldots \quad x_{f,144} \quad x_c]^T_{144 \times 1}
\]

\[
[Z] = [T_{Eh1} \quad T_{Eh1} \quad \sigma_{hh1} \quad \sigma_{hh1} \quad \ldots \quad \sigma_{hh1,144}]^T_{144 \times 1}
\]

\[
h([X]) = [T_{Eh}(x_c) \quad T_{Eh}(x_c) \quad \sigma_{hh1}(x_c) \quad \sigma_{hh1}(x_c) \quad \ldots \quad \sigma_{hh1,144}(x_c)]^T_{144 \times 1}
\]

\[
[H] = \\
\begin{bmatrix}
\frac{\partial T_{Eh}/\partial x_{f,1}}{\partial T_{Eh}/\partial x_{f,2}} & \frac{\partial T_{Eh}/\partial x_{f,2}}{\partial T_{Eh}/\partial x_{f,1}} & \ldots & \frac{\partial T_{Eh}/\partial x_{f,144}}{\partial T_{Eh}/\partial x_{f,144}} \\
\frac{\partial T_{Eh}/\partial x_{f,1}}{\partial T_{Eh}/\partial x_{f,2}} & \frac{\partial T_{Eh}/\partial x_{f,2}}{\partial T_{Eh}/\partial x_{f,1}} & \ldots & \frac{\partial T_{Eh}/\partial x_{f,144}}{\partial T_{Eh}/\partial x_{f,144}} \\
\frac{\partial T_{Eh}/\partial x_{f,1}}{\partial T_{Eh}/\partial x_{f,2}} & \frac{\partial T_{Eh}/\partial x_{f,2}}{\partial T_{Eh}/\partial x_{f,1}} & \ldots & \frac{\partial T_{Eh}/\partial x_{f,144}}{\partial T_{Eh}/\partial x_{f,144}} \\
\frac{\partial T_{Eh}/\partial x_{f,1}}{\partial T_{Eh}/\partial x_{f,2}} & \frac{\partial T_{Eh}/\partial x_{f,2}}{\partial T_{Eh}/\partial x_{f,1}} & \ldots & \frac{\partial T_{Eh}/\partial x_{f,144}}{\partial T_{Eh}/\partial x_{f,144}} \\
\end{bmatrix}
\]
average error in radar backscatter and 2.2 dB for the high noise level. The ancillary data impact on brightness temperature observations was approximately 2.5 K. The off diagonal elements of \([H]\) were again assigned zeros assuming that observation errors were uncorrelated with each other. While in reality the radar observations are uncorrelated with radiometer observations, it is likely that the different polarization brightness temperature and radar observation errors will be correlated. Moreover, the error between radar pixel observations within a radiometer footprint will also be correlated. However, such correlations were not included in the observation error covariance matrix of this demonstration because of the lack of knowledge on the degree of these correlations.

Once the fine resolution soil moisture retrievals are obtained from (10), they can be averaged to medium resolution pixels to generate the medium resolution soil moisture data product. Alternatively, the same approach can be applied using medium resolution (9 km) radar data and medium resolution retrievals attained directly, reducing the retrieval’s computational burden. Both the fine resolution and medium resolution retrievals are compared with the “truth” data to evaluate the performance of the method.

### IV. Retrieval Results

Using the Hydros OSSE data set described previously, the performance of the Bayesian merging method was examined by comparing the retrieved soil moisture values with their corresponding “truth” data and results from a “traditional” soil moisture retrieval method.

This traditional method is a simple mathematical iteration inversion method which treats the soil moisture retrieval problem as a mathematical equation solution problem. In the MEBMs summarized previously, a brightness temperature or radar backscatter is observed and all other model parameters are assumed to be “known.” Hence, the Newton’s search solution method [36] can be used to obtain the soil moisture value corresponding to the brightness temperature or backscatter observations. Given there are two brightness temperature observations and three backscatter observations for each 36- and 3-km pixel, respectively, the iterative inversions of the respective observations are averaged. This method is similar to the method used by Njoku and Li [6] for retrieving surface soil moisture from observations of the Advanced Microwave Scanning Radiometer of NASA’s Earth Observing System (AMSR-E). The performance of the Bayesian merging method is compared with this traditional iteration inversion base-line method.

#### A. Temporal Variations of Retrieval Errors

The root mean square error (RMSE) of the daily surface soil moisture Bayesian merging, radiometer-only, and radar-only retrievals at 3-, 9-, and 36-km resolutions are shown in Fig. 4. The Bayesian method RMSEs are smaller than the radiometer and radar inversions at all resolutions. The large 3- and 9-km radiometer inversion errors result from the large soil moisture variations across a 36-km pixel. However, the 36-km radiometer inversion errors are smaller than the radar inversion errors. While the Bayesian method advantage over the radar inversion method was not very large for the low noise data, it became very significant for the high noise data at all resolutions. Moreover, there was significant improvement from the Bayesian merging as compared to the 36-km radiometer inversions evaluated against the 3-and 9-km truth. The Bayesian merging retrievals were even slightly improved over the radiometer inversions at 36-km resolution. This is because the
Bayesian method optimally merged the different characteristic strengths of the two observation types. Moreover, all RMSE values calculated at the 9-km resolution were lower than those at 3-km because the errors at 3-km are smoothed during the 9-km averaging. This noise cancellation is the basis of creating the 10-km Hydros soil moisture data product (represented by the 9-km soil moisture in this OSSE).

Also shown in Fig. 4 is the average soil moisture content across the study area. The average basin soil moisture was generally drying down throughout the 34-day time period with four major rainfall events early in the period, and another less significant rainfall later in the period. There are two soil moisture related retrieval features. First, there is higher radiometer retrieval error corresponding with the rainfall events on days 154 and 162, but there is no obvious relationship between radiometer retrieval error and soil moisture status. This increase in retrieval error is a result of greater soil moisture heterogeneity at 36-km during and immediately following rainfall events than at other times. Consequently, the Bayesian retrievals also have greater error on those days. Second, the radar retrieval error is inversely proportional to the soil moisture content. This is because backscattering is nonlinearly related to soil moisture content, especially at the dry end. The same backscatter noise added to dry soil pixels will cause a much greater soil moisture retrieval difference than for wet soil pixels. Consequently the radar inversion error in Fig. 4 is smaller early in the time period, and larger later in the period. When the Bayesian method merges the two data sources into an optimal retrieval, the resulting RMSEs are more temporally consistent than either the radar or radiometer inversions. Table III lists the time and space averaged RMSEs of the Bayesian, radar inversion and radiometer inversion retrievals as calculated from the 3-, 9-, and 36-km truth data. Additionally it shows the error statistics for the medium resolution (9 km) Bayesian method, which are still smaller than those of the direct radar backscatter inversions at all three resolutions for both low and high noise data sets. This confirms the advantage of the Bayesian merging method for retrieving the medium resolution soil moisture. However, the RMSE values of the 9-km Bayesian merging retrievals and direct inversions are much larger than those of the 3-km Bayesian merging retrievals and direct inversions, respectively, because the 9-km retrievals or inversions lack the spatial details of 3-km soil moisture, and the simple averages of the 3-km MEBM parameters to 9-km smooth out features important to the nonlinear responsiveness of the MEBM. This highlights that representative parameters are critical to accurate soil moisture retrievals.

<table>
<thead>
<tr>
<th>Method</th>
<th>Low Noise Data Compared with truth at 36km 9km 3km</th>
<th>High Noise Data Compared with truth at 36km 9km 3km</th>
</tr>
</thead>
<tbody>
<tr>
<td>3km</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bayesian Merging</td>
<td>1.5 2.1 2.8</td>
<td>1.6 2.7 3.8</td>
</tr>
<tr>
<td>Radar Inversion</td>
<td>1.9 2.6 3.6</td>
<td>3.2 4.1 6.0</td>
</tr>
<tr>
<td>9km</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bayesian Merging</td>
<td>1.4 2.7 5.0</td>
<td>1.5 4.4 6.1</td>
</tr>
<tr>
<td>Radar Inversion</td>
<td>2.7 3.7 5.7</td>
<td>2.6 5.0 6.6</td>
</tr>
<tr>
<td>36km</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Radiometer Inversion</td>
<td>1.5 4.4 6.2</td>
<td>1.6 4.5 6.3</td>
</tr>
</tbody>
</table>
To measure the Bayesian method’s improvement over the radar inversion method, we have calculated the absolute and relative soil moisture retrieval error improvement for each pixel of the domain. The absolute error improvement is calculated as the radar inversion errors minus the Bayesian merging retrieval errors. For day 155, the Bayesian method’s domain average absolute error improvement are 1.4% v/v, 1.0% v/v, and 1.5% v/v at the 3-, 9-, and 36-km resolutions, respectively. The relative error improvement is the absolute error improvement divided by the soil moisture range of each pixel during the OSSE time period. For day 155, the domain average relative error improvements are 29%, 19%, and 21% at the 3-, 9-, and 36-km resolutions, respectively. These statistics demonstrate the advantage of the Bayesian merging method for retrieving soil moisture from the Hydros radar and radiometer observations.

**B. Spatial Characteristics of the Retrieval Errors**

To demonstrate the spatial characteristics of the soil moisture retrievals from the Bayesian merging method and the radar and radiometer inversion methods, the soil moisture retrieval errors from these methods are mapped at 3-, 9-, and 36-km resolutions in Fig. 5. These are for the high noise data set on day 155, which is representative of the entire simulation period. In these maps, most overestimation errors (in blue) occurred in the eastern areas of the basin where the dense forests associated with higher vegetation water content decrease the soil moisture sensitivity of radar and radiometer observations. When soil moisture retrievals are evaluated against the 36-km truth, the Bayesian merging method performance is similar to the radiometer ($T_B$) inversion, while the radar backscatter ($\sigma$) inversions have higher errors. However, at 9- and 3-km resolution the radiometer inversions have the highest error resulting from scale differences, while the radar inversions have an error magnitude greater than the Bayesian method. Moreover, the radar inversions had a typical overestimation of the true surface soil moisture in the eastern region where the vegetation parameters (e.g., roughness) have higher heterogeneity and the linear aggregation method for these parameters might have caused the radar inversions bias. The basin RMSEs listed in map captions show that the overall performance of the Bayesian method is better than the radar inversion method.

One may anticipate that the Bayesian merging soil moisture retrieval error should be high when both radiometer and radar inversions have high errors. This may not be true as shown in Fig. 4(a) and (b). Around day 154, both the radiometer and radar inversion RMSEs demonstrated error increases while the Bayesian merging RMSE might not. This is caused by the spatial distribution of the retrieval errors. As shown in Fig. 5, many pixels with high radiometer inversion errors do not have high radar inversion errors (e.g., those pixels in the southeast panhandle). Both radar and radiometer observations for each of those pixels restrain each other in the Bayesian merging method, so that
the overall performance of the Bayesian merging method may not necessarily correspond to increased errors in the inversions. However, when both the radar and radiometer inversions fail simultaneously, the Bayesian method may show relatively high error even though the domain average RMSEs of both the radar and radiometer inversions are not relatively high (see the small increased error in the Bayesian Merging RMSE curve at day 150).

V. CONCLUSION

A Bayesian merging method for retrieving surface soil moisture from reliable coarse resolution radiometer and noisy fine resolution radar observations has been proposed and demonstrated with the Hydros OSSE data sets. The method applies the system state update equation of the Kalman filter for combining the two different Hydros observations into a poor background soil moisture field. In this demonstration, a simple inversion of the horizontal polarization radiometer brightness temperatures ($T_{\text{BHz}}$) was used as the background soil moisture field and all three polarizations of 3-km radar backscatter and two polarizations of the 36-km radiometer brightness temperature used as the observation. The Bayesian merging method then produced a 3-km soil moisture data field. These 3-km soil moisture retrievals were then aggregated to 9-km to generate a medium resolution soil moisture data product. While the output of a sophisticated land surface prediction model could be used as the background for a Kalman filter style update (with the ability to also retrieve additional land surface states), the Bayesian merging method presented requires only those data fields used by the “traditional” remote sensing soil moisture retrieval algorithms and resulted in a model independent surface soil moisture field.

Based on the Hydros OSSE data sets, the Bayesian merging method produced the best soil moisture retrievals compared with traditional numerical inversions of the radar or radiometer observations. The average RMSE of 3-km soil moisture retrievals from the Bayesian method was 2.8%v/v compared with 3.6%v/v from the direct radar backscatter inversions for a low noise data set. For the more realistic high noise data set, the RMSE from the Bayesian method was 3.8%v/v compared with 6.0% from the direct radar backscatter inversions. In this Hydros OSSE study, the errors were added to the simulated Hydros radar and radiometer observations and ancillary data were hypothetic. The values and ranges of the simulation model parameters (such as vegetation water content, roughness parameter) may not be representative at the global scale or large model parameters (such as vegetation water content, roughness parameter) may not be representative at the global scale or large.

Acknowledgment

The authors appreciate the encouragement and constructive comments from Prof. D. Entekhabi and other members of the Hydros mission science team. The authors are also very grateful to the three anonymous reviewers for their critical comments and constructive suggestions.

REFERENCES


Xiwu Zhan received the Ph.D. degree in soil, crops, and atmospheric sciences from Cornell University, Ithaca, NY, in 1995.

He is currently with the Hydrology and Remote Sensing Laboratory of USDA Agricultural Research Service, Beltsville, MD. His research interests include land surface soil moisture and vegetation remote sensing, modeling, and data assimilation for regional and global water, energy, and carbon cycle studies.

Paul R. Houser received the Ph.D. degree in hydrology from the University of Arizona, Tucson, in 1996.

He is currently the Director of the Center for Research on Environment and Water, Calverton, MD, and a Professor of Global Hydrology at George Mason University, Fairfax, VA. His research interests include hydrologic remote sensing, development and application of hydrologic data assimilation methods, and global water and energy cycle studies.

Jeffrey P. Walker completed the B.Surv. and B.E. (civil) degrees at the University of Newcastle, Australia, in 1995 and the Ph.D. degree in environmental engineering at the same university in 1999.

Dr. Walker then joined the Hydrological Sciences Branch at the NASA Goddard Space Flight Center as a visiting scientist for two years, before joining the Department of Civil and Environmental Engineering at the University of Melbourne, Australia. His research has been focused on environmental sensing, earth system modeling and data assimilation.

Wade T. Crow (M’03) received the Ph.D. degree from Princeton University, Princeton, NJ, in 2001.

He is currently a Research Physical Scientist with the Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD. His research involves the application of land surface modeling, data assimilation and remote sensing technology to hydrology and agriculture.