Soil moisture algorithm validation using data from the Advanced Microwave Scanning Radiometer (AMSR-E) in Mongolia

Thomas J. Jackson, Ruud Hurkmans, Ann Hsu and Michael H. Cosh

USDA ARS Hydrology and Remote Sensing Lab, 104 Bldg. 007 BARC-West, Beltsville, MD 20705, USA
E-mail: tjackson@hydrolab.arsusda.gov

Abstract
New satellite-based low frequency microwave radiometers are being evaluated for estimating soil moisture. These include the Advance Microwave Scanning Radiometer (AMSR) instruments on the NASA Aqua satellite and the Advanced Earth Observing Satellite (ADEOS-II). Although there have been satellite instruments in space previously with these frequencies, the AMSR programs include the first ever commitments to providing a soil moisture product. As part of its algorithm selection process, the Japanese Aerospace Exploration Agency (JAXA) has been evaluating four different soil moisture algorithm approaches including a single channel method. This algorithm was evaluated here using validation data collected in Mongolia. Data quality issues, approaches to validation and algorithm performance are analyzed and discussed. Overall, the algorithm performed as expected based upon previous research.

Riassunto
E’ in corso la valutazione di nuovi radiometri a microonde da satellite per la stima dell’umidità del terreno. Questa include il sensore Advanced Microwave Scanning Radiometer (AMSR) montato sui satelliti Aqua della NASA e ADEOS-II (Advanced Earth Observing Satellite) della Japanese Aerospace Exploration Agency (JAXA). Anche se ci sono stati in passato strumenti operanti a queste frequenze, il programma AMSR comprende i primi impegni per garantire un prodotto di umidità del terreno. Come parte del suo processo di selezione degli algoritmi JAXA sta valutando quattro approcci per un algoritmo di estrazione dell’umidità del treno, compreso un metodo basato su un solo canale radiometrico. Questo algoritmo viene qui discusso, usando per la validazione degli archivi di dati raccolti in Mongolia. Vengono esaminati inoltre le questioni della qualità dei dati, gli approcci di validazione e le prestazioni dell’algoritmo, che sono state buone, come previsto dai risultati della ricerca svolta in precedenza.

Introduction
The year 2002 began what may be a new era in global soil moisture mapping. Three satellite low frequency (<7 GHz) microwave radiometers were launched into Earth orbit. Two of these were the Advance Microwave Scanning Radiometer instruments on the NASA Aqua (AMSR-E) and the Japanese Aerospace Exploration Agency (JAXA) Advanced Earth Observing Satellite (ADEOS-II AMSR). The third was the Coriolis Windsat instru-
ment. Each system includes dual polarization 6 and 10 GHz channels as well as several higher frequencies. With these sensors it is possible to map the Earth every two to three days. Although there have been satellite instruments in space previously with similar microwave frequencies, the current AMSR satellite programs include the first ever commitments to providing a soil moisture product. Only data from AMSR-E will be utilized here.

As part of its algorithm selection process, JAXA has been evaluating four different algorithm approaches [Koike et al., 2000]. One of these is the single channel algorithm described in Jackson [1993]. In order to make a selection from the alternatives, a number of validation analyses are being conducted. The first of these utilizes a data set collected in Mongolia in 2002. In the investigation conducted in this paper, the single channel algorithm was evaluated using the Mongolia data. Validation of satellite soil moisture estimates using ground-based observations can be challenging and requires a number of tradeoffs. Some of these are discussed in the context of the Mongolia Match-Up data set.

**AMSR-E data**
The Aqua AMSR-E instrument includes dual polarization data at frequencies 6.925, 10.65, 18.7, 23.8, 36.5, and 89.0 GHz. Ground footprint sizes are 75x43, 51x29, 27x16, 32x18, 14x8, and 6x4 km respectively. The viewing angle of AMSR is a constant 55° and the total swath is 1445 km. Overpass times are 1:30 am (descending-D) and 1:30 pm (ascending-A) in local time. The nominal footprint spacing is 10 km. Additional details on AMSR-E can be found at [http://wwwghcc.msfc.nasa.gov/AMSR/](http://wwwghcc.msfc.nasa.gov/AMSR/) and at [http://sharaku.eorc.nasda.go.jp/AMSR/index_e.htm](http://sharaku.eorc.nasda.go.jp/AMSR/index_e.htm) Preliminary studies indicate that there is widespread radio frequency interference (RFI) in the C band channels (6 GHz), especially in populated regions and within the U.S. [Njoku et al., 2003]. Therefore, it is likely that the most useful channels for soil moisture retrieval will be those operating at the slightly higher X band (10 GHz).

**Soil Moisture Algorithm**
For a vegetated site, the brightness temperature ($T_B$) value measured is a combination of the radiation emitted by the vegetation and the radiation emitted by the underlying soil as modified by the vegetation. This is described by the radiative transfer equation [Jackson, 1993]:

$$T_B = \Gamma(e, T_S) + (1 - \omega)T_C(1 - \Gamma) + (1 - e_c)(1 - \omega)T_C(1 - \Gamma)\Gamma$$  \[1\]

Where $T_B$ is the brightness temperature, $T_S$ and $T_C$ are temperatures of soil and canopy respectively, $\omega$ is the single scattering albedo, $e$ is the surface emissivity and $\Gamma$ is the transmissivity of the canopy. At microwave wavelengths, the single scattering albedo can be very small. When this term is set to zero and if it is assumed that the physical temperatures of soil and canopy are nearly the same, Equation [1] reduces to:

$$e = \left( \frac{T_B}{T_S} \right) = 1 - (1 - e_c)\Gamma^2$$ \[2\]

Where $e$ is the emissivity of the surface, which may or may not be covered with vegetation. Inverting Equation [2] gives $e_c$ expressed as a function of two observations, $T_B$ and $T_S$:

$$e_c = 1 - \frac{1 - T_B/T_S}{\Gamma^2}$$ \[3\]

In the current version of the algorithm used here, the surface temperature ($T_S$) is estimat-
ed using brightness temperature from the vertically polarized 37 GHz band according to the following equation [De Jeu, 2003]

\[ T_s = 0.861 T_{B,37V} + 52,550 \]  

[4]

The transmissivity of the canopy \( \Gamma \) can be described by:

\[ \Gamma_p = \exp\left( -\frac{\tau_p}{\cos(\theta)} \right) \]  

[5]

Where \( p \) is polarization, \( \tau \) is the optical depth of the canopy and \( \theta \) is the incidence angle of the radiation. \( \tau \) is determined by:

\[ \tau = b \times VWC \]  

[6]

Where \( b \) is a vegetation parameter that depends on land use and frequency, and \( VWC \) is the vegetation water content in [kg m\(^{-2}\)] [Jackson and Schmugge, 1991]. The vegetation water content can be obtained from Normalized Difference Vegetation Index (NDVI) data [Jackson et al., 1999] using linear relationships:

\[
\begin{align*}
VWC &= 2.0 \times NDVI & \text{when } 0.36 < NDVI \leq 0.50 \\
VWC &= 2.5 \times NDVI & \text{when } 0.20 < NDVI \leq 0.36 \\
VWC &= 3.0 \times NDVI & \text{when } 0.00 < NDVI \leq 0.20
\end{align*}
\]  

[7]

The surface emissivity, \( e_r \), is corrected for the influence of surface roughness according to Choudhury et al. [1979] yielding the smooth surface emissivity \( e_s \):

\[ e_s = 1 + (e_r - 1) \exp(h \cos^2 \theta) \]  

[8]

Where \( h \) is a roughness parameter \((=4s^2k^2)\), which is proportional to the root mean square (RMS) height variations of the soil surface \( s \), and wave number \( k = \frac{2\pi}{\lambda} \) where \( \lambda \) is wavelength in cm). With the inverted Fresnel equation for horizontally polarized radiation the dielectric constant \( \varepsilon \) is computed from the smooth surface reflectivity \( R_s (=1 - e_s) \):

\[ \varepsilon = \sin^2 \theta + \left[ \cos \theta \left( \frac{1 - \sqrt{R_s}}{\sqrt{R_s} - 1} \right) \right]^2 \]  

[9]

From \( \varepsilon \) the volumetric soil moisture is computed using the dielectric mixing model of Wang and Schmugge [1980].

Before brightness temperature values are used in the algorithm, they are screened for rainfall using the algorithm presented in Ferraro et al. [1997]. In addition the data are also screened for certain land uses. Retrieval is only possible in areas with low vegetation content, so some categories such as forests are left out.

The soil moisture algorithm requires land cover information not only to screen the data but also to provide the vegetation \( b \) parameter, which is a static database. The soil texture database required for dielectric constant is also static. Vegetation water content is estimated using NDVI. As originally implemented this too was a static database derived from historic data. This database and alternative approaches using dynamic data sets will be described in a later section.
Mongolia data Match-Up data set
A data set was provided by JAXA consisting of AMSR-E brightness temperature and ground based soil moisture and temperature data for a region in Mongolia. Some specifics of the data set were:
- Time period DOY 182 to 254 of 2002 (July 1 - September 21);
- Products covered latitudes 45.0N – 47.5N and longitudes 105.5E – 108.0E.

The in situ or ground observation data used here is based on data from a network of soil moisture and meteorological stations distributed in this region of Mongolia [Kaihotsu et al., 2003]. There are two types of stations, automated weather stations (AWS) and automated soil and temperature stations (ASSH). Locations are shown in Figure 1. All stations include one point of soil moisture and temperature at depths of 3 and 10 cm collected every hour. Data are recorded onto loggers and are downloaded once every few months during the year. Only the ASSH data were used for soil moisture in this study. Figure 1 shows the spatial domain of the study as well as the locations of the ground observations.

Two types of products were provided, point and grid. The point data consisted of an ASCII file georeferenced for each of twelve soil moisture ground sampling sites. For each date and ascending (A) and descending (D) overpass time there is a record consisting of all AMSR-E channels for the closest satellite footprint to the site followed by the soil moisture and temperature at depths of 3 and 10 cm for the previous and following twelve hour period.

Grid data were generated on a 0.1-degree basis for the 2.5 by 2.5 degree spatial domain. These are shown in Figure 1. Cell values were generated for both satellite and ground data by interpolation and extrapolation of the observations and resampling to the fixed grid.

Figure 1 - Mongolia Match-Up data set spatial domain and in situ site locations.
One can expect a relatively large number (~600 to 800) of AMSR-E footprints to fall within the study area box, therefore, the resulting map products should be representative. However, only the twelve ASSH soil moisture sites shown in Figure 1 were used to generate the soil moisture and temperature grid data over the entire domain. This data will be discussed in more detail in a later section.

Landscape conditions in the Mongolia study region are compatible with the expected capabilities of C and X band soil moisture retrieval algorithms. Vegetation consists mostly of sparse grasses and smooth surfaces. Figure 2 is a photograph from one of the sites. The soils are sandy loams and there appears to be a significant amount of surface rocks in most photographs.

Validity of comparing data products

There are several different ways that the data might be compared in soil moisture retrievals:

**Point Data:** Comparing one ground based soil moisture point to an AMSR footprint (~ 50 km) is not statistically reasonable or physically defensible. Figure 1 includes an example of a nominal C/X band footprint that shows this disparity in scales. To illustrate the flaws of this approach we conducted a temporal stability analysis of the 3 cm in situ soil moisture data collected over the three-month period.

This technique, temporal stability analysis [Vachaud et al., 1985; Mohanty and Skaggs, 2001; Cosh et al., 2004], exploits the hypothesis that a soil moisture field maintains its spatial pattern over time. The principle variable used in this analysis is the relative difference, calculated by

\[
\delta_{i,j} = \frac{S_{i,j} - \bar{S}_j}{\bar{S}_j}
\]

where \(S_{i,j}\) is the soil moisture at time \(j\) for site \(i\). A mean and standard deviation for each site are then calculated using all samples in time, \(j\). These are plotted in rank order (based upon the mean) to produce a mean relative difference plot. Stability of a site is measured by the standard deviation of the relative difference. Low values indicate stability, while high values indicate instability.

If the sensor location is demonstrated to be stable at long time scales, it is possible to use this to an advantage. These locations can be used to reliably estimate the areal mean. If the mean relative difference plot shows a bias, indicated by a large deviation from zero, an appropriate correction will have to be made. With this knowledge the number of sites within a network could be reduced, yet the network would remain reliable.

Here we will use this analysis a bit differently. We will use temporal stability analysis to illustrate the weakness of assuming that a single point can be compared to a satellite foot-
print, especially without the apriori knowledge of a temporal stability analysis.

Figure 3 is a mean relative difference plot for the surface soil moisture data from the
Mongolia network. Also plotted are the errors bars equivalent to the standard deviation of
relative differences for each site. Comparing these results to information in Figure 1, there
appears to be no relationship between geographic location within the study region and
mean relative difference, indicating that sampling sites can be randomly selected. Also,
there was a single site that appeared to be significantly wetter than the surrounding region,
Site ‘F0’. This site is also less stable in time, as indicated by the large standard deviation
of relative differences. More detailed surface studies are necessary to confirm whether this
site is representative of the surrounding region or if it is anomalous.

Figure 3 indicates that very few of the network sites could be used to reliably estimate a
large-scale average in this region. Only “A6” and “E4” have no bias and low standard devi-
ations. Some other sites such as “GUS” have low standard deviations but are biased.
Without correcting for bias a comparison to footprint estimates could be erroneous.

Grid Cell: Retrieval of soil moisture from the microwave data on a cell-by-cell basis is
probably valid due to the high density (~10 km) of the AMSR-E sampling but results need
to be carefully interpreted at the grid cell scale. However, comparing these results to the
gridded soil moisture generated using the 12-in situ points is not valid over the entire study
domain. As shown in Figure 1, the point samples cover only a portion of the region. Much
of the region must be gridded by extrapolation of pattern based upon this small number of
points located in the center of the region. In addition, inferring spatial patterns from 12

Figure 3 - Temporal stability analysis plot with mean relative difference values plotted versus rank.
Error bars are equivalent to the one standard deviation of the relative differences and site labels are
included for reference.
points distributed over a region this size is also not justified based upon known sources of variation in soil moisture.

Subset of Region: After considering the issues described above, we concluded that the only valid comparison that can be performed with the Mongolia Match-Up data provided is to compare the retrieved soil moisture for the portion of the gridded domain that encompasses the twelve-in situ points to the average of those twelve points. This subset of the region is shown in Figure 1.

For each day/pass, soil moisture was retrieved for each of the original grid cells in this subset. These were averaged to obtain a single value. The in situ soil moisture data for the closest hourly observation were then averaged. These two averages were then compared.

Estimating VWC from Vegetation Indices
The soil moisture algorithm described above involves two vegetation characteristics: a vegetation type parameter (b), which is fixed by land cover, and the vegetation water content (VWC), which will exhibit daily, seasonal, and annual variability related to environmental and meteorological factors. In this approach it is hypothesized that the VWC can be adequately estimated using readily available satellite based vegetation index products. The most widely available of these is the Normalized Difference Vegetation Index (NDVI), which is computed from red and near infrared measurements.

NDVI and other products have been generated from several satellite sensors for over twenty years. To compensate for cloud cover, products are usually available as composites over multiple days. They are also high resolution (as compared to microwave measurements). Ingesting and processing these data dynamically in an algorithm is certainly feasible, however, during initial algorithm development this wasn’t possible. Therefore, in order to facilitate the soil moisture algorithm during initial formulation we chose to establish a global set of 10-day interval NDVI values based upon the NOAA AVHRR NDVI Pathfinder data set (http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/LAND_BIO/GLBDST_main.html). This data set provides 10-day composite values between 1982 and 1999.

For each 10-day interval we averaged all years to establish a climate average set of 10-day NDVI values. Figure 4a shows a plot of the average value for the Mongolia subset region for the studied time period. Our approach facilitated analysis but it ignores the annual variation in the NDVI resulting from meteorological and cultural influences. If a region experiences drier or wetter than normal conditions it might result in quite different NDVI values. This is an inherent limitation of this database, however, it was our intention to use more timely input on NDVI later in algorithm application. Figure 4a includes the Mongolia subset regional average for each year of record to illustrate this inter-annual variability.

In order to account for NDVI variability in this particular match up test, we extracted concurrent vegetation index coverage from the NASA Moderate Resolution Instrument (MODIS) land products for the region in 2002. Two indices are plotted in Figure 4b along with the climate averages that the retrieval algorithm default would have provided. For the NDVI, the data match up well in the early portion but diverge in the later portion. This likely reflects low rainfall during the season.

The second index plotted in Figure 4b is the Enhanced Vegetation Index (EVI). This is an alternative index being produced from MODIS data (http://lpdaac.usgs.gov/modis/mod13a2.html). It is similar to NDVI but attempts to minimize the effects of two very important sources of temporal and spatial variability that are not related to vegetation, aerosols and the soil background. Our review of ground photography indicates that the soil background (especially with higher rock fractions and dry vegetation) will be brighter than
conditions encountered in the development of our VWC estimation methods. Brighter soil backgrounds should result in generally lower NDVI values for a corresponding VWC [Huete et al., 1994]. In order to account for this factor we used the EVI values to estimate VWC. Figure 5 shows the spatial and temporal distribution of the EVI for the study area.
and time period. Generally higher values in July and early August would be expected based upon the data in Figure 4.

**Results and Discussion**

**General Temporal Behavior of \( T_B \) and Soil Moisture**

The average \( T_B \) and soil moisture for the subset region were plotted versus time. Data were separated into A and D plots in Figure 6. A review of these plots shows that soil moisture varied over a relatively small range during the extended study period: \( \sim3\text{-}10\% \). This suggests sandy soils with low precipitation amounts. Most of the variation occurred during two periods, one around DOY 200 and the other around DOY 250. \( T_B \) data were not available for a significant period in the middle of this data set. Unfortunately, based upon the values on DOY 221 it appears that some interesting conditions may have occurred during this gap. We were not provided the soil moisture data for those days without AMSR coverage.

The \( T_B \) data show a range of 240 to 275 K over the study period for the ascending daytime orbits (225 to 255 K for the descending nighttime orbits). This is a fairly large range for C band,
which should indicate good potential for soil moisture retrieval. Several large drops in $T_B$ occur that correspond to increases in soil moisture. However, in every case the drop in $T_B$ lasts a very short period of time (one to two days). Following the decrease, the $T_B$ increases to the same level as the preceding date, yet soil moisture shows a gradual decrease over 5 or 6 days (as expected). There is obviously a physical reason why a disconnect occurs between soil moisture and $T_B$. One explanation we can offer is that AMSR is responding to a shallower depth of soil than the in situ sensors. Deeper layers take longer to dry than the near surface. This is a likely condition following rain events.

Soil Moisture and Temperature Retrieval Results.
We assembled EVI and soil ancillary data sets and computed retrieval algorithm parame-
Soil temperature was estimated using the method described in De Jeu (2003) with the AMSR-E 37 GHz V data. Results for the A and D data are shown in Figure 7.

The A data set predictions are excellent with a 1.60°C standard error of estimate (SEE) and low bias (Tab. 1). The D data sets have a larger SEE (2.98°C), which is mostly the result of a larger bias. The 37 GHz channel is expected to respond more to the near surface. At night the surface could be cooler than deeper layers.

These temperature data were used in our retrieval algorithm to estimate soil moisture. Results are summarized in Figures 8 and 9 and in Table 1. The SEE values are within the expected level of performance of the retrieval algorithm. As in the case of soil temperature, the ascending (daytime) retrievals are better than the descending results.
Summary
For the first time an effort is underway to use satellite measurements to estimate and map global soil moisture. This is a result of the availability of new satellite-based low frequency microwave radiometers and in particular the Advance Microwave Scanning Radiometer instrument on the NASA Aqua. Both the NASA and JAXA AMSR programs include commitments to providing a soil moisture product from the satellite data. As part of its algorithm selection process, JAXA has been evaluating alternative soil moisture algorithm approaches including a single channel method. Here, this algorithm was evaluated using validation data.
collected in Mongolia. Various methods for comparing ground and satellite data were considered. Temporal stability analysis revealed that care must be taken in what data are used in validation tests. Overall the retrieval algorithm produced acceptable results for soil moisture and temperature, ascending results were better than descending. It was necessary to use a modified vegetation index in the estimation of the vegetation water content due to the unusual soil and surface properties in the test site. It is well known that vegetation has a significant effect on the ability to retrieve soil moisture. The conditions in Mongolia were relatively benign for the expected capabilities of AMSR, very low levels of vegetation. More challenging sites will be considered in the future along with more robust data sets.

Table 1 - Standard error of estimate and Bias for soil moisture and temperature in Mongolia.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Standard Error of Estimate</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil Moisture: Ascending</td>
<td>3.24%</td>
<td>-0.91%</td>
</tr>
<tr>
<td>Soil Moisture: Descending</td>
<td>3.38%</td>
<td>0.36%</td>
</tr>
<tr>
<td>Temperature: Ascending</td>
<td>1.60 °C</td>
<td>0.06 °C</td>
</tr>
<tr>
<td>Temperature: Descending</td>
<td>3.21 °C</td>
<td>-3.11 °C</td>
</tr>
</tbody>
</table>
Acknowledgment
This work was supported in part by the Japanese Aerospace Exploration Agency under the Advanced Earth Observing Satellite II Algorithm Research and Development Program.

References