Estimating vegetation water content during the Soil Moisture Active Passive Validation Experiment 2016

Michael H. Cosh
William A. White
Andreas Colliander
Thomas J. Jackson
John H. Prueger
Brian K. Hornbuckle
E. Raymond Hunt
Heather McNairn
Jarrett Powers
Victoria A. Walker
Paul Bullock
Abstract. Vegetation water content (VWC) is an important land surface parameter that is used in retrieving surface soil moisture from microwave satellite platforms. Operational approaches utilize relationships between VWC and satellite vegetation indices for broad categories of vegetation, i.e., “agricultural crops,” based on climatological databases. Determining crop type–specific equations for water content could lead to improvements in the soil moisture retrievals. Data to address this issue are lacking, and as a part of the calibration and validation program for NASA’s Soil Moisture Active Passive (SMAP) Mission, field experiments are conducted in northern central Iowa and southern Manitoba to investigate the performance of the SMAP soil moisture products for these intensive agricultural regions. Both sites are monitored for soil moisture, and the calibration and validation assessments had indicated performance issues in both domains. One possible source could be the characterization of the vegetation. In this investigation, Landsat 8 data are used to compute a normalized difference water index for the entire summer of 2016 that is then integrated with extensive VWC sampling to determine how to best characterize daily estimates of VWC for improved algorithm implementation. In Iowa, regression equations for corn and soybean are developed that provided VWC with root mean square error (RMSE) values of 1.37 and 1.10 kg/m², respectively. In Manitoba, corn and soybean equations are developed with RMSE values of 0.55 and 0.25 kg/m². Additional crop-specific equations are developed for winter wheat (RMSE of 0.07 kg/m²), canola (RMSE of 0.90 kg/m²), oats (RMSE of 0.74 kg/m²), and black beans (RMSE of 0.31 kg/m²). Overall, the conditions are judged to be typical with the exception of soybeans, which had an exceptionally high biomass as a result of significant rainfall as compared to previous studies in this region. Future implementation of these equations into algorithm development for satellite and airborne radiative transfer modeling will improve the overall performance in agricultural domains. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JRS.13.014516]

Keywords: vegetation water content; soil moisture content; Landsat.
The operational algorithms use a long-term average (climatological) vegetation parameterization based on MODIS data in order to facilitate a low latency in the soil moisture estimate. This approach introduces some uncertainty if the actual vegetation conditions deviate from the climatology. Previous experiments have addressed the vegetation parameterization by conducting physical sampling to provide accurate estimates of vegetation conditions at the time of the experiment. For instance, the Soil Moisture Experiments in 2002, SMEX02, established equations for corn and soybeans for central Iowa during the field campaign in early June of 2002. Follow-up experiments furthered this analysis for short time periods of intensive observation periods, but most focused only on a limited two- to three-week period of time and did not include peak biomass conditions of agricultural crops. These methodologies usually take advantage of high-resolution optical datasets such as Landsat, to produce 30 m vegetation index maps. From these maps, relationships can be developed to estimate VWC. In sparsely vegetated regions, the normalized difference vegetation index (NDVI) can be used effectively. For more densely vegetated regions, such as in the U.S. Corn Belt, a more resilient index is necessary. Reference demonstrated that the normalized difference water index (NDWI) in central Iowa is a suitable measure for dense vegetation, especially corn. However, that study did not have the peak biomass conditions that can be observed during the reproductive stages of corn and soybeans.

Estimating the water content of a crop at peak biomass and peak moisture content presented challenges for the design and implementation of Soil Moisture Active Passive Validation Experiment 2016 (SMAPVEX16). SMAPVEX16 provides a reference dataset for not only peak corn and soybean water content, but canola, spring wheat, black beans, and oats. Vegetation sampling was incorporated into the experimental design with the ultimate intention of producing a high-resolution (30 m), daily VWC map for the study region to serve as a ground truth map for the algorithm improvement efforts. This study will review the collection, results, and analysis that led to this high-resolution dataset, which is critical to the microwave satellite community for soil moisture retrieval.

2 Study Sites

During the summer of 2016, an experiment focused on the monitoring of surface soil moisture, and vegetation was conducted at two locations that serve as core validation sites for the SMAP mission (Fig. 1). The experiment was the second such postlaunch validation experiment, entitled the SMAP Validation Experiment in 2016 (SMAPVEX16), with the first taking place around Tombstone, Arizona. These domains were established to coincide with the SMAP Core Validation Sites operated by the USDA Agricultural Research Service near Ames, Iowa (42.03 N, 93.62 W; South Fork), and Agriculture and Agri-food Canada (AAFC) near Winnipeg, Manitoba (49.90 N, 97.14 W).11

The South Fork experimental watershed covers an area of ~1300 km² and has a subhumid climate with an average annual precipitation of 800 mm. The crop land portion is dominated by two row crops, corn, and soybeans. In 2016, the crop proportions were 55% corn and 27% soybeans, with an approximate annual crop rotation of corn-corn-soybean. The average field size is ~800 m × 800 m (quarter section in the Public Land Survey System). The soil texture is dominated by loam and silty clay loam (Des Moines Lobe), and the soil is not well drained with a landscape-featuring kettles. Tile drainage is growing in the region as a means of reducing excess soil moisture. The study areas were defined as 36-km boxes, based on the NASA EASE-Grid 2.0.

The Red River Watershed of southern Manitoba experiences extremes in soil moisture conditions. In the period from 1966 through 2016, 32% of crop loss insurance payments were for drought/heat and 39% were for excess moisture according to Manitoba Agricultural Services Corporation. The watershed is dominantly utilized for annual cropping with some areas also used for forage and pasture. A typical rotation is a cereal crop alternating with an oilseed or pulse crop. Field sizes are variable and can range from as little as 20 ha (45 acres) up to as much as 260 ha (640 acres) with 65 ha (160 acres) being common.
The SMAPVEX16-MB intensive sample site is located in the vicinity of Carman and Elm Creek, Manitoba, within the Red River Watershed. More than 85% of the site is dominated by annual crops consisting of canola, soybean, corn, spring wheat, winter wheat, oats, and edible beans. Only a small fraction (<5%) is under grassland and pasture. The soil texture varies significantly across the site. The east side is composed of fine-textured soils developed on glaciolacustrine sediments of the Red River Plain that formed from Glacial Lake Agassiz deposits. These soils experience frequent shrinking and swelling during the dry–wet cycles due to the presence of smectite clays. The surface soil texture changes abruptly on the

Fig. 1 (a) SMAPVEX16 study sites and landcover, in 36-km EASE-2 grid cells in (b) Manitoba, Canada, and (c) Iowa.
west side of the site to loamy fine sands developed on lacustrine beach deposits of the Lower Assiniboine Delta. The topography across the site is mainly flat to gently undulating with slopes from 0% to 2%. The soils are generally imperfectly to poorly drained which, together with the flat topography, accentuates flooding issues during periods with excess moisture.

3 Methods of Analysis

The computation of VWC maps for croplands is a multistep process that must be done in sequence. First, a land cover classification must be conducted. Next, a field campaign must be designed and executed to collect the physical ground truth of water content. Optimizing collection during Landsat 8 overpasses is desirable, but cloud cover is often a problem in the central plains of the U.S. One way to increase probability of cloud-free pixels is to locate the study area in a region that is on the border of two Landsat swaths, which is the case for the South Fork study region, resulting in twice as many scenes to select from. Next, the atmospherically corrected reflectances can be computed to a vegetation index for each overpass day. Linear interpolations of the vegetation index are then produced to provide a daily estimation of vegetation index for each day of physical sampling. The locations of field sampling can then be identified in the imagery and the vegetation index extracted for comparison. Regression equations are then computed for each dominant crop type so that the vegetation index can be converted to a physical estimate of VWC. These equations are then applied to the daily vegetation index maps to produce daily VWC maps, which can be used for analysis of aerial and satellite product algorithms. Below is a more detailed accounting of this process for the SMAPVEX16 study regions.

During SMAPVEX16, aircraft overflights were conducted during two separate intensive observation periods at each study site. In Iowa, flights and field sampling occurred from May 25 to June 5, and August 3 to 16. In Manitoba, intensive sampling occurred from June 13 to June 20 and July 10 to July 22. Field teams were deployed during aircraft overflight days to collect physically based soil moisture samples for calibration of the aircraft-based L-band radiometer and radar (passive active L-band system). On nonoverflight days, vegetation sampling, surface roughness sampling, and land cover classification surveys were conducted. More detailed information on vegetation sampling will be presented below.

3.1 Land Cover Classification

Land cover was classified separately within each 36-km study area. Land cover for Manitoba was generated by AAFC, at a 30-m resolution, using a decision tree-based methodology on Landsat 8 and RADARSAT-2 imagery, supported by ground truth data, which became available at the end of the calendar year.

Land cover for Iowa was classified using a supervised classification scheme, prior to the release of the 2016 USDA-NASS cropland data layer (CDL). The maximum likelihood classification was trained by using ground truth observations of 1290 farm field polygons. Two Landsat 8 scenes, acquired on July 22 and August 23, 2016, were masked for clouds and nonrow crops and classified using a training and testing dataset. Nearly the entire domain of the Iowa study site was surveyed during SMAPVEX16 to determine crop type and row direction per major parcel. Quality assessment bands included with the Landsat 8 surface reflectance data product were used to create cloud masks. Cloud-contaminated pixels in the classified August image were filled with results from the July image. Nonrow crop areas were extracted from the 2015 CDL and overlaid on the classification results. A confusion matrix at 10,000 points showed good agreement with ground truth with an overall accuracy of 0.942 (Table 1).

Landsat 8 images acquired over the domains were often cloudy. The clearest images for the summer of 2016 were downloaded and processed (Table 2). A 30-m grid was created in ArcGIS, representing Landsat pixel boundaries. With the grid overlaying recent 1-m resolution aerial RGB imagery, grid cells were manually selected around each of the 970 vegetation sampling points, avoiding mixed pixels, and pixels near field edges. Grid cell centroids were used to extract Landsat pixel values from each scene. Surface reflectances for the near-infrared band ($R_{\text{NIR}}$) and a shortwave-infrared band ($R_{\text{SWIR}}$, approximate center wavelength 1.61 to 1.65 $\mu$m)
were extracted. Using the data quality flags to determine whether a pixel was cloud free, reflectance values (up to 9) around each sampling point were averaged. NDWI was calculated:

\[
\text{NDWI} = \frac{R_{\text{NIR}} - R_{\text{SWIR}}}{R_{\text{NIR}} + R_{\text{SWIR}}},
\]

and linearly interpolated between image dates, to give daily estimates at each ground sampling point. We learned from Ref. 6 and other studies that NDWI is a more appropriate vegetation index for VWC, compared with the NDVI, because the shortwave-infrared band is directly sensitive to foliar water content, whereas leaf chlorophyll strongly absorbs in the red band used in NDVI. The respective sensitivities of NDWI and NDVI to changes in VWC were validated with ground-based reflectances obtained with the Cropscan Multispectral Radiometer MSR-16.

### 3.2 Cloud Filling

Cloud-contaminated pixels within the study areas were filled to obtain complete coverage on each date. Three NDWI images for Manitoba, and five for Iowa, showed the expected behavior of VWC, increasing and then decreasing over the growing season. Noncrop pixels did not require cloud-filling since they were assigned constant VWC values, per land cover type, for the whole season.

For Manitoba, the three images were quite clear, with <1% of crop pixels cloud-contaminated. These areas were filled with the average NDWI from the other pixels in that scene with the same land cover for each date.

For Iowa, clouds were common. Ground sampling confirmed crop emergence in the June 4 image, though cloud cover was significant. March 16 was the closest available clear image showing pre-emergent crop fields. Crop fields were predominantly bare soil in both images, so the March image was used in place of the June 4 image.

The June 20 image was exceptionally clear. The few pixels identified as cloudy by the quality assessment bands occurred near water and were filled in with clear pixels from a prior year.
image. This was done for a few cloudy pixels in the March 16 image, as well. The July and August images both had scattered clouds. Both images were acquired at points in the growing season when corn and soybean VWC was high. NDWI for both images were similar. Therefore, the clouded pixels of the July image were assigned the pixel values from the August image.

Less than 1% of pixels in the July/August mosaic and the September 8 image remained cloud-contaminated. For these, the crop-specific relationships developed using clear pixels were used in conjunction with the most recent image to estimate their value. The magnitude of errors associated with cloud contamination was assumed very small for this study because of the restrictions placed on the Landsat scenes that were used.

3.3 Field Sampling Protocols

Vegetation characterization was an extensive part of the field campaign, requiring many teams to collect data across the spatial domain to account for potential cloud coverage during the Landsat overpass dates. Vegetation biomass and water content were collected in several fields in each domain. In Iowa, sampling occurred between May 25 and June 5, and August 3 and 16. In Manitoba, sampling occurred between June 13 and July 22. Field sampling teams sampled ~10 fields each in intensive observation period, at the beginning and end of each. During the May intensive observation period, the crops were just starting to emerge from the soil, so there was little to no vegetation to collect, which was noted as 0.0 kg of water per square meter (kg/m²). Sampling teams entered a field and selected a site 100 m into the field, to get beyond the edge rows. The second and third sampling sites were selected by walking 100 m diagonally across the field from the first site. At each site, a random row was selected, and the row spacing and direction were recorded. Row density was determined by measuring a meter down a row and counting the plants within that meter. These measurements were conducted three times at each site, for a total of nine measurements per field. Then, three plants were selected at each site and their characteristics were recorded. As an example, for corn, these characteristics included stem diameter above the bottom node, stem diameter below the top node, height, and leaf count. The plant was then cut into pieces and separated into different labeled bags for transport to the weighing facility. Vegetation was weighed dried and weighed again by each plant component sample for three days at 70°C. Initial samples were dried for longer to confirm that three days of drying were sufficient. Corn ears were dried for up to seven days, as significant weight loss was observed daily until day 7.

4 Results

First, NDWI imagery was generated for the Landsat overpass days, which were cloud-free or were cloud-filled as described above. Then, daily imagery of NDWI was computed between the scenes for each day between the first and last available scenes. For the locations where sampling
was conducted, the NDWI value was retrieved for the day samples from this daily dataset. When physically sampled vegetation was compared to the retrieved NDWI values for the sampling dates, regression equations from the data, shown in Figs. 2 and 3, were developed for each crop for each domain. Table 3 contains the regression equations for the Iowa study domain as an example. The regression equation was selected to be linear for interpolation purposes, as this was the most appropriate regression option given the data. The two intensive observation periods are separated in time, observing the early and peak periods of the growing season, and it is desirable to include other data for the vegetation stages of the crops. Luckily, two previous large-scale field experiments were conducted in Iowa in 2002 and 2005 that monitored VWC. Results from the SMEX02 and SMEX05 experiments are also plotted in Fig. 2 to demonstrate

![Interpolated NDWI with measured VWC for six crop types in Manitoba](image)

**Fig. 3** Interpolated NDWI with measured VWC for six crop types in Manitoba: (a) soybean, (b) wheat, (c) canola, (d) oats, (e) corn, and (f) black beans.

Table 3 The regression equations for VWC in Iowa. If VWC_{soy/corn} <= 0, VWC is set equal to 0.0 kg/m².

<table>
<thead>
<tr>
<th>Crop</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Iowa</strong></td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>VWC_{corn} = 10.184 * NDWI + 1.0026</td>
</tr>
<tr>
<td>Soy</td>
<td>VWC_{soy} = 3.3909 * NDWI + 1.342</td>
</tr>
<tr>
<td><strong>Manitoba</strong></td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>VWC_{corn} = 10.3103 * NDWI – 0.97944</td>
</tr>
<tr>
<td>Soy</td>
<td>VWC_{soy} = 1.61879 * NDWI + 0.10125</td>
</tr>
<tr>
<td>Canola</td>
<td>VWC_{canola} = 6.85295 * NDWI – 0.69693</td>
</tr>
<tr>
<td>Oat</td>
<td>VWC_{oat} = 7.44985 * NDWI – 1.033</td>
</tr>
<tr>
<td>Wheat</td>
<td>VWC_{wheat} = 4.111 * NDWI + 0.46821</td>
</tr>
<tr>
<td>Black beans</td>
<td>VWC_{black beans} = 6.83 * NDWI – 0.794</td>
</tr>
</tbody>
</table>
the linearity of the relationship.6,7 These experiments had similar data collection protocols and designs to SMAPVEX16 and captured transitional soybean and corn plants after emergence to the beginning of the reproductive stages of each crop. These equations can then be used for any classified pixel in the domain for an observed or interpolated NDWI.

For Manitoba, crop growth was sufficiently advanced that bare soil conditions were not present, so NDWI values were typically greater from the beginning. NDWI was paired with physically sampled VWC, to establish correlations per crop type for both study areas (Fig. 3). Once the regression equations were established between the crop type and NDWI values, they could be applied to the daily NDWI images, and examples are shown in Fig. 4. Once this was complete, the root mean squared error values (RMSE) for each crop type in each domain could be calculated and are shown in Table 4.

These RMSE values are consistent with but larger than values observed during the SMEX02 campaign;6 however, that campaign occurred earlier in the year, and the crop biomass was generally less in total value. In addition, 2016 had a higher maximum VWC compared to 2002 and 2005 (see Fig. 3).

5 Discussion

VWC mapping as it relates to soil moisture retrievals is most important during the growing seasons for agricultural domains as this captures the temporal change in the VWC signal. Previous analysis has shown that NDWI is a better estimator of VWC than NDVI because
of the saturation level, but those studies were limited in biomass, usually occurring during the vegetation stages of agricultural crops.\textsuperscript{6,21,22} The SMAPVEX16 campaign was specifically designed to capture the peak biomass of agricultural crops to provide the most complete analysis of corn and soybean growth conducted during large-scale field experiments on soil moisture.

Figure 5 shows a comparison between VWC from field sampling and from remotely sensed NDWI in Iowa from the SMEX02/SMEX05/SMAPVEX16 experiments. It is observed that both crops have saturated NDWI from remote sensing at 1 and \( 4 \text{ kg/m}^2 \) for soybean and corn, respectively, while the measured VWC increased to 8 \( \text{ kg/m}^2 \) for soybean and 9 \( \text{ kg/m}^2 \) for corn (Fig. 5). Remotely sensed VWC of 1 and 4 \( \text{ kg/m}^2 \) are approximately the mass of water in the stems and reproductive organs in another study.\textsuperscript{23} Saturation of remotely sensed NDWI is related to maximum leaf area index and foliar vegetation cover, similar to NDVI. Saturation of NDWI was not observed in the Manitoba study site, because the availability of the aircraft in the 2016 campaign limited observations only during the earlier vegetative growth stages.

VWC is the sum of foliar and stem water contents, where NDWI is sensing the changes in foliar water content.\textsuperscript{8,23–25} Stem water content may be estimated by different methods: by empirical relationships with leaf area index,\textsuperscript{8,24} by land cover maps,\textsuperscript{25} or by plant height.\textsuperscript{23} Whereas separately estimating foliar and stem water contents may not improve the accuracy of

\begin{table}[h]
\centering
\caption{Root mean squared errors for the crop volumetric water content for Iowa and Manitoba study domains.}
\begin{tabular}{lcc}
\hline
Crop & Iowa & Manitoba \\
\hline
Corn & 1.37 & 0.59 \\
Soybean & 1.10 & 0.25 \\
Canola & — & 0.88 \\
Wheat & — & 0.59 \\
Oat & — & 0.74 \\
Black beans & & 0.31 \\
\hline
\end{tabular}
\end{table}

\begin{center}
\textbf{Fig. 5} Comparison of sampled VWC versus estimated water content from remotely sensed NDWI.
\end{center}
estimated VWC compared to measured VWC, it presents a reasonable method for extrapolating VWC from NDWI to other crops and vegetation cover types without resorting to major field experiments.

6 Conclusions

This study demonstrates a methodology to achieve daily estimates of VWC for agricultural domains in the northern plains of North America with reasonable accuracy. Regression relationships were developed for the major crops in two experimental watersheds in Iowa and Manitoba, and magnitudes of the errors were generally <25% of the total plant water content, which is typical of such studies. In particular for soybeans in Iowa, this study captured fields that had very high crop biomass. The high resolution (30 m) images can be used for modeling and for remote-sensing retrievals from aircraft platforms to help solve the complex issues of signal downscaling for agricultural domains where important physical processes occur at scales smaller than a remote-sensing pixel (~9 km for SMAP).

Acknowledgments

USDA is an equal-opportunity employer and provider. A partial contribution to this work was carried out in Jet Propulsion Laboratory, California Institute of Technology under contract with National Aeronautics and Space Administration.

References


Michael H. Cosh is a research hydrologist at the US Department of Agriculture’s Agricultural Research Service in the Hydrology and Remote Sensing Laboratory. He received his BA degree in engineering from St. Francis University, his BA degree in civil engineering from Pennsylvania State University, his MSc and PhD degrees in civil and environmental engineering from Cornell University.

William A. White is a support scientist at the US Department of Agriculture’s Agricultural Research Service in the Hydrology and Remote Sensing Laboratory. He received his BS degree in environmental science and policy, and his MPS degree in geospatial information sciences, from the University of Maryland.

Andreas Collander is a research scientist at the Jet Propulsion Laboratory, California Institute of Technology. He received his MSc and PhD degrees in electrical engineering from Aalto University, Finland. His research is focused on developing remote-sensing methods to answer science and societally important questions regarding Earth’s hydrological cycle.

Thomas J. Jackson is a retired hydrologist from the US Department of Agriculture’s Agricultural Research Service. He received his BS, MS, and PhD degrees from the University of Maryland.

John H. Prueger is a research soil scientist at the US Department of Agriculture’s Agricultural Research Service in the National Laboratory for Agriculture and Environment.

Brian K. Hornbuckle received his ScB degree in electrical engineering Brown University, his MA degree in secondary education from the University of Mississippi, and his MSE and PhD...
degrees in electrical engineering and atmospheric science from the University of Michigan. Since 2003, he has been a faculty member in the Department of Agronomy, Iowa State University of Science and Technology, Ames, where he is currently a professor of agricultural meteorology.

**E. Raymond Hunt** is a research physical scientist at the USDA Agricultural Research Service’s Hydrology and Remote Sensing Laboratory located at the Beltsville Agricultural Research Center. He received his BS degree from Ohio University (Botany) and his PhD from the University of Michigan (Botany). His current research uses remote sensing with unmanned aircraft systems for precision agriculture. Previous research used remote sensing to estimate plant water content and to detect invasive plant species.

**Heather McNairn** is a research scientist with Agriculture and Agri-Food Canada, Science and Technology Branch. She received her bachelor of environmental studies degree from the University of Waterloo, her master of science degree in soil science from the University of Guelph, and her PhD in geography from the Université Laval.

**Jarrett Powers** is a manager with Agriculture and Agri-Food Canada’s Science and Technology Branch in Winnipeg, Manitoba. He received his BSc degree at the University of Manitoba in 1989 and a Certificate in GIS Studies from Simon Fraser University in 2001.

**Victoria A. Walker** is a NASA Earth and Space Science fellow and a PhD candidate in agricultural meteorology at Iowa State University. She received her MS degree in agricultural meteorology from Iowa State University and her BS degree in applied meteorology from Embry–Riddle Aeronautical University, Prescott campus.

**Paul Bullock** is a professor and formerly head, Department of Soil Science, University of Manitoba. He received his BSc agriculture and his MSc soil science degrees from the University of Saskatchewan and his PhD Centre for Resource and Environmental Studies, Australian National University. He did global operational crop yield forecasting and remote sensing with the Canadian Wheat Board and Noetix Research Inc. His research includes weather-based crop modeling and production risk assessment as well as soil moisture modeling and measurement.