Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans

Thomas J. Jackson\textsuperscript{a,}\ast, Daoyi Chen\textsuperscript{b}, Michael Cosh\textsuperscript{a}, Fuqin Li\textsuperscript{a}, Martha Anderson\textsuperscript{c}, Charles Walthall\textsuperscript{a}, Paul Doriaswamy\textsuperscript{a}, E. Ray Hunt\textsuperscript{a}

\textsuperscript{a}USDA ARS Hydrology and Remote Sensing Laboratory, 104 Building 007 BARC-West, Beltsville, MD 20705, USA
\textsuperscript{b}Manchester Centre for Civil and Construction Engineering, University of Manchester, Manchester, UK
\textsuperscript{c}University of Wisconsin, Madison, WI 53705, USA

Received 5 May 2003; received in revised form 23 September 2003; accepted 21 October 2003

Abstract

Information about vegetation water content (VWC) has widespread utility in agriculture, forestry, and hydrology. It is also useful in retrieving soil moisture from microwave remote sensing observations. Providing a VWC estimate allows us to control a degree of freedom in the soil moisture retrieval process. However, these must be available in a timely fashion in order to be of value to routine applications, especially soil moisture retrieval. As part of the Soil Moisture Experiments 2002 (SMEX02), the potential of using satellite spectral reflectance measurements to map and monitor VWC for corn and soybean canopies was evaluated. Landsat Thematic Mapper and Enhanced Thematic Mapper Plus data and ground-based VWC measurements were used to establish relationships based on remotely sensed indices. The two indices studied were the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI). The NDVI saturated during the study period while the NDWI continued to reflect changes in VWC. NDWI was found to be superior based upon a quantitative analysis of bias and standard error. The method developed was used to map daily VWC for the watershed over the 1-month experiment period. It was also extended to a larger regional domain. In order to develop more robust and operational methods, we need to look at how we can utilize the MODIS instruments on the Terra and Aqua platforms, which can provide daily temporal coverage.

Keywords: Vegetation water content; Landsat; NDWI

1. Introduction

Vegetation water content (VWC) is an important parameter in agricultural and forestry applications. The VWC could possibly provide information for agriculture that can be used to infer water stress for irrigation decisions, aid in yield estimation (Penuelas et al., 1993) and assessment of drought conditions (Tucker, 1980). The principle application to forestry is determining fire susceptibility (Pyne et al., 1996). The VWC is also used in retrieving soil moisture from microwave remote sensing observations (Jackson & Schmugge, 1991), which is the focus of our investigation.

As part of a large-scale field campaign called the Soil Moisture Experiments 2002 (SMEX02), we explored the potential of using remote sensing to map and monitor VWC for corn and soybean canopies. Our goals were to provide high-quality VWC estimates for the numerous microwave remote sensing soil moisture and evapotranspiration studies conducted during SMEX02, and to contribute to the development of robust methods for providing near real-time information appropriate for global satellite applications.

2. Vegetation indices and vegetation water content

A number of investigators have explored the potential of using reflectance data to estimate the VWC. The physical definitions of VWC vary from water volume per leaf or ground area (equivalent water thickness) to water mass per mass of vegetation dry matter. The problem has been approached at leaf, plant and canopy scales. Leaf scale functions are well defined whereas canopies are more
difficult to characterize from first principles (Ceccato et al., 2001, 2002a; Hunt, 1991; Hunt & Rock, 1989; Knipling, 1970; Tucker, 1979, 1980). Some of this work has focused on optimal selection of bands based upon simulation or high-spectral resolution data (Aldakheel & Danson, 1997; Datt, 1999; Gao, 1996; Gao & Goetz, 1995; Hunt & Rock, 1989; Penuelas et al., 1993; Roberts et al., 1997; Rollin & Milton, 1998; Serrano et al., 2000; Sims & Gamon, 2002; Tucker, 1980; Ustin et al., 1998). Few studies have attempted VWC retrieval using operational satellite data (Ceccato et al., 2002b; Zarco-Tejada et al., 2003). Our objective is the development of a quasi-operational method using currently available satellite data. Therefore, we focused on canopy scales and available satellite bands.

Tucker (1979) used the Normalized Difference Vegetation Index (NDVI), developed by Rouse et al. (1973), to estimate leaf water content and other physiological variables for grasses

$$\text{NDVI} = \frac{R_{\text{NIR}} - R_{\text{RED}}}{R_{\text{NIR}} + R_{\text{RED}}}$$  \hspace{1cm} (1)$$

where $R_{\text{NIR}}$ is the reflectance or radiance in a near infrared channel (0.78–0.90 $\mu$m TM/ETM+ and 0.72–1.1 $\mu$m AVHRR) and $R_{\text{RED}}$ is the reflectance or radiance in a visible channel (0.63–0.69 $\mu$m TM/ETM+ and 0.58–0.68 $\mu$m AVHRR). Tucker (1979) recognized that the relationship between NDVI and VWC was based on correlations of both quantities with the amount of vegetation. An advantage to using NDVI for VWC estimation is that NDVI is a routinely produced product, available globally typically every 10 days using satellite instruments such as the Advanced Very High Resolution Radiometer (AVHRR).

It is well known that the NDVI has limited capability for estimating VWC because it is affected by other variables. NDVI is, in fact, often referred to as a greenness index. Ceccato et al. (2002b) summarize the limitations of using the NDVI as follows:

- Each plant species has its own relationship of chlorophyll content and VWC;
- A decrease in chlorophyll content does not imply a decrease in VWC;
- A decrease in VWC does not imply a decrease in chlorophyll content.

Additionally, NDVI saturates at intermediate values of leaf area index (LAI), therefore it is not responsive to the full range of canopy VWC. However, for specific canopy types within specific regions and when supported by ground sampling it is possible to establish useful VWC functions based on NDVI. Jackson et al. (2002) used this approach for grasslands during previous microwave soil moisture remote sensing investigations because use of NDVI increased the accuracy of soil moisture predictions. Even with these limitations in many situations, the only available option has been to use the NDVI because it was the only information available.

Gao (1996) developed the Normalized Difference Water Index (NDWI) for determination of VWC based on physical principles. Hardisky et al. (1983) developed the similar Normalized Difference Infrared Index for the Landsat Thematic Mapper concluded that this index was related to VWC. These indices are of the general form

$$\text{NDWI} = \frac{R_{\text{NIR}} - R_{\text{SWIR}}}{R_{\text{NIR}} + R_{\text{SWIR}}}$$  \hspace{1cm} (2)$$

where $R_{\text{SWIR}}$ is the reflectance or radiance in a short wave infrared wavelength channel (1.2–2.5 $\mu$m). For Landsat TM/ETM+, $R_{\text{NIR}}$ and $R_{\text{SWIR}}$ correspond to bands 4 (0.78–0.90 $\mu$m) and 5 (1.55–1.75 $\mu$m), respectively. One reason that the NDWI may not have received much attention until recently is that the infrequent temporal coverage of TM and ETM+ make it difficult to estimate VWC for various applications. Classic operational instruments such as the AVHRR did not include a SWIR band. However, new satellite sensors such as the Moderate Resolution Imaging Sensor (MODIS) on NASA’s Terra and Aqua satellites now make such data routinely available. Gao (1996) recommended the use of a SWIR band centered at 1.24 $\mu$m, now available on MODIS, for NDWI because this band has similar atmospheric transmittance as the NIR band.

Ceccato et al. (2002a,b) developed and tested an index similar to the NDWI using data from the SPOT-vegetation sensor. They concluded that the SWIR channel was critical to estimating VWC and that the NIR channel was needed to account for variation of leaf internal structure and dry matter content variations. They and many others have also concluded that indices contrasting the SWIR channel with the NIR channel were sensitive to the mass or volume of water and not to the fractional percentage of water.

We utilize Landsat TM and ETM+ imagery for the investigation presented here because of the high spatial resolution (30 m). The higher spatial resolution facilitates coordination of ground data collection with the imagery, thus assuring homogeneity of vegetation conditions within pixels (i.e., no mixtures). Our goal is to eventually adapt this work to the coarser resolution of MODIS, with its 500-m SWIR spatial resolution.

3. SMEX02 data description

SMEX02 was a soil moisture and evapotranspiration remote sensing experiment conducted in Iowa between late June and mid-July 2002 (hydrolab.arsusda.gov/SMEX02 for the experiment plan). Three of the data sets developed during SMEX02 are used for this investigation. These are
described in the following sections along with brief descriptions of the data processing.

3.1. Landsat TM reflectance data

All data collected by the Landsat 5 and 7 satellites for an extended time period centered on the SMEX02 field campaigns were evaluated. The study area was located within the overlap region of two Landsat paths, and thus, resulted in a greater probability for data sets than the usual 16-day temporal coverage (Path 26 Row 31 and Path 27 Row 31). Table 1 summarizes the scenes used for the analyses. Cloud fraction was very important during scene selection. Two areas of the landscape were outlined for study. One area surrounded the Walnut Creek Watershed where ground observations were obtained, and where the higher resolution aircraft microwave remote sensing missions were focused. The other landscape area was a larger Region that was targeted for analysis of data from the Aqua Advanced Microwave Scanning Radiometer (AMSR) as well as higher resolution aircraft observations. Table 2 describes these landscape areas. Based upon Table 1, we concluded that an adequate number of Landsat ETM+ scenes were available for the Walnut Creek Watershed and a marginal number for the Region. Fig. 1 shows the false color composite images (Bands 2, 3 and 4) of the Watershed area on June 6, June 23, July 1, July 8, and July 17.

The Landsat ETM+ data were processed to provide at-surface apparent reflectance data. Reflectances are apparent since it was not possible to know all of the atmospheric properties at the time of the satellite observation and how these properties varied spatially. However, it is important to reduce the atmospheric effects due to known factors prior to the formation of indices.

Sun photometer data http://aeronet.gsfc.nasa.gov collected from a site within the watershed study area was used to drive the Second Simulation of the Satellite Signal in the Solar Spectrum model (6S; Vermote et al., 1997) for atmospheric corrections and surface reflectance calculations. Landsat 7 ETM+ data were used to correct Landsat 5 TM data sets using the derived reflectances of invariant targets. During this time period, there were problems with the Landsat 5 TM calibration, thus creating some uncertainty with these images. The July 16 Landsat 5 and July 17 Landsat 7 data sets were selected as base images to compare differences between the two satellites and to develop any adjustments necessary for the Landsat 5 data. A basic assumption made was that the surface reflectance was the same on both of these dates and that relationships between bands 3, 4 and 5 for the two satellites were consistent during the entire study period. After establishing the adjustments for the July 16 Landsat 5, the same adjustments were applied to the June 23 Landsat 5 data set.

The resulting data set consists of apparent reflectance values for channels 3, 4 and 5 on June 6, June 23, July 1, July 8, July 16 and July 17. These data were validated via comparisons with limited ground based reflectance observations obtained during SMEX02 and also with the results from the MODTRAN4 program (Berk et al., 1998) using atmospheric profiles of pressure, temperature, and water vapor from radiosonde profiles within the watershed during SMEX02.

3.2. Land cover classification analysis

Land cover is important when using NDVI and NDWI to estimate VWC because the relationships vary by vegetation type. As part of SMEX02, a land cover database was developed at a 30-m spatial resolution.

Landsat TM data from three dates (May 14th, July 1st, July 17th) and two ground surveys (June and July) were used to produce a land cover map for the SMEX02 time period. Bands 3, 4, 5, and 7 were extracted for each date. The study area was split into two pieces due to clouds on one date. The North and South portions of the area were cloud free on all three dates, while the center portion (approximately at equal latitude with the City of Ames) was cloud covered on July 17th. This resulted in twelve channels of data for the North–South and eight channels for the middle.

Land cover classification was performed using a supervised procedure. Training sites were chosen from a drive-by survey conducted during the field campaign. The resulting classes included: Unclassified, Alfalfa, Corn, Grass, Soybean, Trees, Urban, and Water. A map of roads provided by

| Table 2 |
| SMEX02 study area coordinates |

<table>
<thead>
<tr>
<th>Area</th>
<th>Longitude (degrees)</th>
<th>Latitude (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watershed</td>
<td>93.832437 W</td>
<td>42.037725 N</td>
</tr>
<tr>
<td></td>
<td>93.392070 W</td>
<td>42.040074 N</td>
</tr>
<tr>
<td></td>
<td>93.830295 W</td>
<td>41.872919 N</td>
</tr>
<tr>
<td></td>
<td>93.391061 W</td>
<td>41.875254 N</td>
</tr>
<tr>
<td>Region</td>
<td>93.816924 W</td>
<td>42.729216 N</td>
</tr>
<tr>
<td></td>
<td>93.163691 W</td>
<td>42.732922 N</td>
</tr>
<tr>
<td></td>
<td>93.827997 W</td>
<td>41.694599 N</td>
</tr>
<tr>
<td></td>
<td>93.161040 W</td>
<td>41.697469 N</td>
</tr>
</tbody>
</table>

WGS84 Ellipsoid—ITRF and CTRS realization
the Iowa Department of Transportation was digitized such that each road was 60 m wide. This image was merged with the classified image creating a new land cover class of road network.

3.3. Vegetation water content data

VWC was measured several times in 31 fields as part of a comprehensive vegetation sampling effort during SMEX02. Four rounds of data were collected during SMEX02, however not every field was sampled during each set. Twelve sites containing flux towers were sampled four times during the experiment; the remaining 19 sites were sampled twice. For each field-date combination, three locations in the field were visually pre-selected from airborne digital imagery to represent average, minimum and maximum canopy conditions. Each sampling location was 10 rows across, and 12 m long. Samples were collected from every 2nd row, yielding 5 sets of observations per location × 3 locations per field = 15 data sets per field per sampling date. When multi-date sampling was conducted within a field, each time a sampling location was revisited, samples were collected in a 3 by 5-m row sub-area adjacent to the previously sampled sub-area (to avoid decimating a sample location).

Above ground biomass was removed and wet and dry weights were used to compute VWC. For this investigation, we averaged all samples (all plots) within a field on a given date and used this single value.

4. Development and application of VWC mapping functions

4.1. Overview of approach

The general approach used in this investigation is as follows. Temporal functions were developed for corn and soybean VWC. Temporal functions were then developed for corn and soybean NDVI and NDWI using Landsat TM and ETM+ data. For each crop type, VWC was linked to NDVI or NDWI using the date. Using these VWC functions, images were generated for each Landsat TM scene (cloud-free portions). Finally, pixel-by-pixel temporal interpolation was performed to generate daily VWC images for the Watershed and Region with the available scenes.

4.2. VWC temporal relationships

An important characteristic for consideration during the analysis was that each round of sampling was performed over a period of several days. All crops were rapidly growing during this time. Therefore, it is impossible to
directly relate the Landsat TM/ETM+ data collected on a single day to VWC obtained over a week. Our solution to this problem was to use the VWC samples to establish a growth curve for each crop type. The procedure is as follows:

- For each field on each day it was sampled, all samples collected were averaged to generate a field VWC estimate;
- For each field, the available estimates were used to interpolate and extrapolate VWC values for each day of the SMEX02 study period for that field;
- For each day of the study period, the estimates for each of the 31 fields were used to compute an average corn and an average soybean VWC for that date;
- A functional relationship was established to predict VWC as a function of date for corn and soybeans.

Fig. 2 shows the results obtained for corn and soybeans for the SMEX02 study period. This figure also includes the standard deviation, which illustrates the variability of VWC for each crop as a function of VWC level.

4.3. NDVI and NDWI temporal relationships

Landsat data were available for five dates spanning the SMEX02 study and ground sampling campaigns. Temporal

Fig. 2. Vegetation water content as a function of day of year during SMEX02 per crop type. Also plotted are the ±1 standard deviation error bars for each crop per day.

Fig. 3. Vegetation indices, NDVI and NDWI, as a function of day of year (DOY) during SMEX02. Note the saturation of some indices after DOY 185.

Fig. 4. Vegetation water content of corn and soybean versus the (a) NDVI and (b) NDWI during SMEX02.
relationships for NDVI and NDWI were generated as follows:

- For each image, the relevant bands were used to compute NDVI and NDWI;
- All 31 sampling fields were identified on each image;
- For each field on each date, an average NDVI and NDWI was computed for the three vegetation sampling sites;
- For each date, an average NDVI and NDWI was computed for corn and for soybeans (averaging over all of the sampling sites);
- NDVI and NDWI values for corn and soybeans at an early date (June 10) when VWC was expected to be close to zero were specified to equal 0.05 kg m$^{-2}$;
- A functional relationship was established to predict NDVI or NDWI as a function of date for corn and soybeans.

Fig. 3 shows the NDVI and NDWI relationships obtained for corn and soybeans. We observed that for corn the NDVI saturates during the later period of the study while NDWI continues to change. This is an expected, yet very important limitation of using NDVI. The VWC is still increasing during this period while the NDVI no longer responds to the changes.

4.4. VWC as a function of NDVI and NDWI

The final step in developing the VWC functions was to relate the VWC to either NDVI or NDWI. This was
accomplished using the temporal functions. By matching dates, we obtained the relationships shown in Fig. 4a (NDVI) and Fig. 4b (NDWI). A functional form was then fit to each set of data to obtain the following equations:

\[
VWC = VWC = 192.64\text{NDVI}^5 - 417.46\text{NDVI}^4 \\
\quad + 347.96\text{NDVI}^3 - 138.93\text{NDVI}^2 \\
\quad + 30.699\text{NDVI} - 2.822 \quad (3)
\]

**NDVI soybeans**

\[
VWC = 7.63\text{NDVI}^4 - 11.41\text{NDVI}^3 + 6.87\text{NDVI}^2 \\
\quad - 1.24\text{NDVI} + 0.13 \quad (4)
\]

**NDWI corn**

\[
VWC = 9.82\text{NDWI} + 0.05 \quad (5)
\]

**NDWI soybeans**

\[
VWC = 1.44\text{NDWI}^2 + 1.36\text{NDWI} + 0.34 \quad (6)
\]

### 4.5. Watershed daily VWC and retrieval performance evaluation

Eqs. (3)–(6) were applied to each of the Landsat scenes for the Watershed area. The land cover database described previously was used to determine which equation to use. In addition, several additional procedures were used:

- VWC values were assigned as fixed values, based upon previous investigations (Jackson et al., 2002) to pixels with the following land cover types: Alfalfa (0.5 kg m\(^{-2}\)), Grass (0.5 kg m\(^{-2}\)), Trees (10.0 kg m\(^{-2}\)), Roads (0 kg m\(^{-2}\)), Urban (0 kg m\(^{-2}\)), Water (0 kg m\(^{-2}\)), Unclassified (0 kg m\(^{-2}\)).

Images of the VWC maps for the five Landsat dates based on NDWI are shown in Fig. 5. Daily VWC images were then generated by interpolating between these five images on a pixel by pixel basis using a Hermite cubic interpolation scheme.

The overall procedure was then quantitatively evaluated by comparing the predicted VWC for each field on the day it was sampled to the ground-based estimate. These results are summarized in Table 3. The bias and root mean square error (rmse) values for both corn and soybean fields were greater for NDVI than for NDWI. Errors for corn were greater than soybean. However, this is reasonable when considering the much larger range of observed VWC values for corn.

### 4.6. Regional daily VWC

There was cloud cover over the Region study area on several of the Landsat TM dates that prevented us from using the exact same procedure developed for the Watershed area. Instead of utilizing all five scenes, only the June 23, June 1, and June 17 images were used. The regional study area spanned two separate rows on the same TM/ETM+ path. Therefore, to capture the entire study region, two scenes were composited. The same equations were used to calculate VWC from NDWI (Eqs. (5) and (6)) as were used for the watershed study area. Linear interpolation was used to generate daily regional scenes. Using the same watershed vegetation sampling data, statistics were calculated to verify the accuracy of this new interpolation scheme. Soybean values showed reasonable agreement with the field data, having a bias of \(-0.05\) kg m\(^{-2}\) and rmse equal to 0.196 kg m\(^{-2}\). Errors within the cornfield sampling were greater however, with a bias of \(-0.262\) kg m\(^{-2}\) and rmse of 0.650 kg m\(^{-2}\).

### 5. Discussion and conclusions

Both NDVI- and NDWI-based methods for VWC estimation were considered and NDWI was found to be superior based upon a quantitative analysis of bias and standard error. The NDWI method was used to successfully map daily VWC for the watershed over the 1-month experiment period and then extended to a larger regional area.

The concept of using vegetation indices to predict the VWC has been around for several decades but has never been widely applied. One application requiring knowledge of VWC is prediction of water stress for irrigation scheduling and crop yield modeling, but physiological responses to drought stress happen over a very small interval of VWC and sensors may not be able to detect the change (Hunt & Rock, 1989 and references therein). Another application of VWC is determination of fire fuel moisture. Recently, VWC has been used to estimate LAI, because red/near-infrared vegetation indices saturate at intermediate levels of LAI, but LAI estimation depends on canopy structure. As noted previously, the temporal coverage of Landsat limits the capability to detect changes in VWC and limits all of these potential applications.

Although we were fortunate to have two satellites, path overlap and benign cloud conditions for SMEX02, limitations in image frequency degrades the relationships that can be developed between vegetation indices and VWC. In order to develop more robust methods and operational
applications it will be necessary to examine the utilization of the MODIS instruments on the Terra and Aqua platforms, which can provide daily coverage. Differences of spectral bands and coarser spatial resolution will present challenges to this extension. However, the alternative infrared band available on MODIS, 1.24 μm, may offer improved capabilities and extend the range of detectable VWC before saturation (Gao, 1996).

Acknowledgements

The authors wish to thank the NASA EOS Aqua AMSR, Terrestrial Hydrology, and Global Water and Energy Cycle programs for support of SMEX02 and this research. We also wish to thank our cooperators and participants in SMEX02. The authors thank Brent Holben (NASA/GSFC) for the loan of the AERONET sun photometer used during the experiment, processing and quality control analysis of the sun photometer data.

References


