Sub-pixel reflectance unmixing in estimating vegetation water content and dry biomass of corn and soybeans cropland using normalized difference water index (NDWI) from satellites

JINGFENG HUANG*, †, DAOYI CHEN‡ and M. H. COSH§
†Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149-1031, USA
‡Department of Engineering, The University of Liverpool, Liverpool L69 3GQ, UK
§Hydrology and Remote Sensing Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD, 20705-2350, USA

(Received 10 April 2007; in final form 1 April 2008)

Estimating vegetation cover, water content, and dry biomass from space plays a significant role in a variety of scientific fields including drought monitoring, climate modelling, and agricultural prediction. However, getting accurate and consistent measurements of vegetation is complicated very often by the contamination of the remote sensing signal by the atmosphere and soil reflectance variations at the surface. This study used Landsat TM/ETM+ and MODIS data to investigate how sub-pixel atmospheric and soil reflectance contamination can be removed from the remotely sensed vegetation growth signals. The sensitivity of spectral bands and vegetation indices to such contamination was evaluated. Combining the strengths of atmospheric models and empirical approaches, a hybrid atmospheric correction scheme was proposed. With simplicity, it can achieve reasonable accuracy in comparison with the 6S model. Insufficient vegetation coverage information and poor evaluation of fractional sub-pixel bare soil reflectance are major difficulties in sub-pixel soil reflectance unmixing. Vegetation coverage was estimated by the Normalized Difference Water Index (NDWI). Sub-pixel soil reflectance was approximated from the nearest bare soil pixel. A linear reflectance mixture model was employed to unmix sub-pixel soil reflectance from vegetation reflectance. Without sub-pixel reflectance contamination, results demonstrate the true linkage between the growth of sub-pixel vegetation and the corresponding change in satellite spectral signals. Results suggest that the sub-pixel soil reflectance contamination is particularly high when vegetation coverage is low. After unmixing, the visible and shortwave infrared reflectances decrease and the near-infrared reflectances increase. Vegetation water content and dry biomass were estimated using the unmixed vegetation indices. Superior to the NDVI and the other NDWIs, the SWIR (1650 nm) band-based NDWI showed the best overall performance. The use of the NIR (1240 nm), which is a unique band of MODIS, was also discussed.

*Corresponding author. Email: jhuang@rsmas.miami.edu
1. Introduction

As a crucial vegetation biophysical property, vegetation water content (VWC), the fraction of water within the vegetation, has wide applications in soil moisture mapping, hydrological modelling, crop water stress evaluation, forest-fire risk assessment and many other vegetation related studies. Its importance has been reported in Doraiswamy et al. (2004), Jackson et al. (2004) and Chen et al. (2005). Along with the VWC, the vegetation dry biomass (VDB) increases with vegetation respiration and photosynthesis, and quantifies the stored biomass energy in vegetation. The VDB is particularly useful for crop yield modelling, carbon sequestration evaluation, or any other biomass energy related fields. The relationships between the VDB and the crop yield have been well established (Daughtry 1988, Wiegand et al. 1991, Daughtry et al. 1992, Ruimy et al. 1994, Gamon et al. 1995).

Vegetation property estimation from satellites has been in development for decades; however, to obtain the most accurate relation of satellites’ spectral signals to the vegetation growth, researchers still face difficulties in correcting the effects from atmospheric attenuation and sub-pixel soil reflectance. These effects contaminate the reflectance values of vegetation, which are recorded as pixels in satellite images. The contaminations lead to inaccuracies of vegetation indices that are broadly used to estimate vegetation properties. Therefore, practical atmospheric and sub-pixel soil reflectance unmixing schemes are needed to ensure that vegetation properties can be accurately estimated from satellites.

When the surface reflectance of ground vegetation is remotely sensed, the atmosphere introduces errors in the vegetation reflectance on the pathways of light coming to the sensors. There are two types of atmospheric correction approaches that address this atmospheric effect: a theoretical approach using an atmospheric model, and an empirical approach with ground data measured simultaneously with the satellite overpass. Complex atmospheric models, such as MODTRAN (Kneizys et al. 1988) and 6S (Second Simulation of the Satellite Signal in the Solar Spectrum; Vermote et al. 1997, Ouaidrari and Vermote 1999), were developed to conduct atmospheric correction. Their ability heavily relies on the availability of local meteorological data, such as atmospheric water vapour, aerosol optical thickness, cloud coverage, and geometry angles of sun and satellite sensors. Frequently, these inputs are not available for specific study areas on data acquisition dates. In contrast, although the empirical approach does not have such dependency, the difficulties in managing field ground data are insurmountable on a large scale. One possible option is to use alternative satellite observations as ground data, but it is subject to the resolution requirement. Therefore, it is necessary to develop a hybrid model that will combine the advantages of both atmospheric models and empirical approaches.

Besides the atmospheric attenuation, the soil effect is a major concern in computing accurate vegetation reflectances. It is instructional to compare the bare soil reflectance to the pure vegetation reflectance to understand how a mixed reflectance differs from pure vegetation reflectance. Figure 1 shows significant difference between the reflectance of bare soil and pure green vegetation. In the visible (400–700 nm) and SWIR regions (1300–3000 nm), the soil reflectance is higher than the vegetation reflectance; in contrast, in the NIR region (800–1200 nm), the soil reflectance is lower than the vegetation reflectance. The arrows in figure 1
indicate the possible changing direction of the surface reflectance at each band from the case of the 'mixed' vegetation and soil to the case of the 'unmixed' vegetation.

Soil effect generally comes from two parts: the soil reflectance from the soil background underlying the vegetation canopy, and the soil reflectance from the sub-pixel unvegetated bare soil. The first part of soil effect has been frequently researched and eliminated. By considering the irradiance directly reflected by the vegetation and a soil-dependent component, the soil adjusted vegetation indices (SAVI) used a conceptual 'soil line' and an adjustment factor to correct this soil effect (Huete 1988, Huete et al. 1994, Liu and Huete 1995, Gao et al. 2000, Huete et al. 2002). In their study, the soil-dependent component is a product of multiplying the global irradiance and the soil reflectance, and the downward and upward global transmittance through the canopy. The use of both downward and upward transmittance implies their emphasis on the underlying soil reflectance. By assuming that the vegetation is spread homogenously over the area, the SAVI can be used for a wide range of leaf area index (LAI). Thus, it is believed that the SAVI is very valuable for correcting the soil effect from the underlying soil reflectance. Because the red radiance (630–690 nm) is easily absorbed by leaf chlorophyll content and reaches saturation earlier than the near-infrared (NIR) and shortwave infrared (SWIR) bands (Gamon et al. 1995, Chen and Brutsaert 1998), the use of red band in the SAVI is subject to the same early saturation problem as the NDVI. As a result of unvegetated bare soil, the second part of soil effect is more predominant when vegetation coverage is low, but its correction has been rarely attempted. Little progress has been made to correct this part of soil effect in practical estimation of vegetation properties using operational satellite data. The major problem lies in the insufficient information of vegetation coverage and the poor evaluation of the fractional sub-pixel soil reflectance. If vegetation coverage can be adequately
estimated, soil effect from sub-pixel bare soil reflectance can be removed through reflectance unmixing and the subsequent derived vegetation indices should become attainable. It is noteworthy that the unmixed vegetation index is complementary to the existing soil effect adjusted vegetation indices. In future work, it might be valuable to combine both approaches to correct the soil effect comprehensively.

Fractional vegetation cover is defined as the portion of the pixel that is covered with vegetation canopy whose soil background cannot be observed while being viewed from above. Vegetation coverage is quantified as the percentage of vegetation cover to ground area. Therefore, vegetation coverage is the descriptive parameter of ‘horizontal’ canopy expansion. Satellite derived vegetation indices, such as the well-known Normalized Difference Vegetation Index (NDVI) (Jasinski 1996, Carlson and Ripley 1997, Townshend et al. 2000), are broadly used to derive the vegetation coverage. The NDVI is the only vegetation index used in long-term historical vegetation record and is conventionally used for various vegetation properties estimation purposes. However, as indicated from previous experimental approaches and theoretical modelling (Gamon et al. 1995, Chen and Brutsaert 1998), saturation of NDVI at high LAI values may limit its use in vegetation property estimation. The SWIR (1300–2500 nm)-based index provides an alternative to the NDVI. Compared with the NDVI, it was found that the SWIR-based spectral index saturates at a higher value (Roberts et al., 1997, Chen et al. 2005). In Chen et al. (2005), the SWIR-based normalized difference water index (NDWI) saturates at a higher value of VWC than the NDVI and was suggested for the VWC estimation. The SWIR-based NDWI is more sensitive to the vegetation growth than the NDVI, particularly when vegetation coverage is high. The inclusion of the SWIR bands in vegetation coverage estimation has not been fully evaluated. Therefore, to avoid the inaccuracies caused by the early saturation phenomena of the NDVI, the NDWI should be considered for vegetation coverage estimation purposes. The saturation phenomenon of visible bands versus the SWIR bands will be discussed and the performance of the NDWI in estimating vegetation coverage will be evaluated in this study.

To demonstrate the level of sub-pixel soil reflectance contamination, the six Landsat bands (Blue 485 nm, Green 560 nm, Red 660 nm, NIR 830 nm, SWIR 1650 nm and 2215 nm) and one MODIS NIR (1240 nm) band are evaluated before and after the sub-pixel reflectance unmixing. The NIR (1240 nm) is a unique band of MODIS. The correlations between vegetation properties, namely VWC and VDB, and the unmixed vegetation indices are also investigated.

The visible NIR and SWIR based vegetation indices listed below are evaluated in this study:

\[
\text{NDWI}_{1240} = \frac{(\text{NIR}_{858\text{nm}} - \text{NIR}_{1240\text{nm}})}{(\text{NIR}_{858\text{nm}} + \text{NIR}_{1240\text{nm}})} 
\]

\[
\text{NDWI}_{1640} = \frac{(\text{NIR}_{858\text{nm}} - \text{SWIR}_{1640\text{nm}})}{(\text{NIR}_{858\text{nm}} + \text{SWIR}_{1640\text{nm}})} 
\]

\[
\text{NDWI}_{2130} = \frac{(\text{NIR}_{858\text{nm}} - \text{SWIR}_{2130\text{nm}})}{(\text{NIR}_{858\text{nm}} + \text{SWIR}_{2130\text{nm}})} 
\]

\[
\text{NDVI} = \frac{(\text{NIR}_{858\text{nm}} - \text{RED}_{648\text{nm}})}{(\text{NIR}_{858\text{nm}} + \text{RED}_{648\text{nm}})} 
\]

The subscriptions of the NDWI are consistent with the previous work of Chen et al. (2005). The NDWI_{1240} was initially proposed by Gao and Goetz (1995) and Gao
With the NIR (860 nm) as a reference band, the Red (660 nm), the NIR (1240 nm), and the SWIR (1650 nm and 2215 nm) are used as measurement bands in the NDVI, the NDWI$_{1240}$, the NDWI$_{1640}$, and the NDWI$_{2130}$ respectively.

The Soil Moisture Experiment (SMEX02) campaign was organized and conducted by the Agriculture Research Services (ARS) of the United States Department of Agriculture (USDA). The field experiments measured the LAI, VWC, dry biomass and vegetation coverage quantitatively. It provides valuable ground observations which are intensively used to derive the empirical relationships between the vegetation properties and the vegetation index. More details about the campaign can be found in Jackson (2002), Anderson (2003), and Jackson and Cosh (2003).

2. Data sources and data processing

2.1 Study area

The area of interest in this study is the Walnut Creek (WC) watershed located south of Ames, Iowa, US (figure 2). The study area is on both path 26 and 27 of row 31 of Landsat imagery, and on the tile h11v04 of MODIS imagery. It was the focus of SMEX02, which took place from mid-June to mid-July 2002. During the SMEX02 period, the VWC and VDB were increasing with time. Agricultural crops occupied 73.4% of the total area (39.5% in corn and 33.9% in soybeans), with an additional 12% as urban areas and roads, 14% as grasses and trees, and 0.6% as trace pixels of other classes (Doraiswamy et al. 2004). A supervised classification map (in 30 m resolution) which was discussed in Chen et al. (2005) was also used in this study to identify corn and soybeans fields. Vegetation properties, such as LAI, VWC, dry biomass and vegetation coverage, were measured quantitatively on the ground. Vegetation sampling data were collected for 21 corn fields and 10 soybeans fields within the Walnut Creek study region. Vegetation data were collected in two periods: DOY 166–169 and DOY 178–189. There were six sites covering both periods and the rest of the sites only covered the second period (Anderson et al. 2004). Details of the vegetation sampling methods can be found in Jackson (2002) and Anderson et al. (2004). A particular site to note is the site WC25 which is a particular site designated to a sandy hill slope where the vegetation growth was notably stunted (Anderson et al. 2004).

2.2 Landsat and MODIS datasets

Landsat and MODIS data were jointly used in this study. Over the SMEX02 period, there were good-quality images available on 5 days from Landsat: 7, 23 of June; 1, 8, and 17 of July. All of these were Landsat 7 ETM+ images except 23 June, which was Landsat 5 TM image. The TM image on 16 July was also used to investigate the incorporation of TM to ETM+ data series. The Landsat data were obtained in two versions: atmospheric uncorrected version, and 6S atmospheric corrected version. Details of the 6S atmospheric correction can be found in Chen et al. (2003). The quality MODIS data without cloud and cloud shadow contamination were available on 7, 23, and 25 June, and 1, 8, and 17 of July. Detailed MODIS data processing can be found in Chen et al. (2005) and Huang (2006).

To avoid using Landsat and MODIS interchangeably and to achieve the highest comparability of spectral bands and vegetation indices, Landsat data were used as a primary dataset for the visible, NIR (860 nm) and SWIR bands. The NIR band
which is a unique MODIS band and not available from Landsat, was used to calculate NDWI\textsubscript{1240}. The purpose of including this NIR (1240 nm) band is to evaluate the use of long wavelength NIR band in vegetation modelling. The six Landsat bands and one MODIS band were highlighted in figure 1. The resolutions of Landsat and MODIS data are 30 m and 500 m respectively. Considering the limitation from its moderate resolution, the applicability of MODIS data in this study area was discussed in Chen et al. (2005) with more explicit explanations. It is noteworthy that the relatively low resolution of MODIS may result in the

![Landsat 4-3-5 false-colour-composite (FCC) for the study area of Walnut Creek watershed in Iowa, US. (a) SMEX02 study area. (b) Large pavement area (~0.2 km x 0.2 km) (left, white circled area in (a)), and the reservoir (largest dimension from north to south is ~0.45 km) (right, black circled area in (a)). The corn and soybeans fields of SMEX02 campaign in solid rectangulars are coloured in blue and yellow respectively in (a). The reservoir and the large reflective pavement which were used in the hybrid BWAC model as reference sites are circled in black and white respectively in (a) and their dimension and shapes are shown respectively in (b). Courtesy of Google Earth.](image)

(1240 nm), which is a unique MODIS band and not available from Landsat, was used to calculate NDWI\textsubscript{1240}. The purpose of including this NIR (1240 nm) band is to evaluate the use of long wavelength NIR band in vegetation modelling. The six Landsat bands and one MODIS band were highlighted in figure 1. The resolutions of Landsat and MODIS data are 30 m and 500 m respectively. Considering the limitation from its moderate resolution, the applicability of MODIS data in this study area was discussed in Chen et al. (2005) with more explicit explanations. It is noteworthy that the relatively low resolution of MODIS may result in the
data-quality deterioration of NDWI when it is compared with the rest of the vegetation indices. Chen et al. (2005) and Huang (2006) discussed details of Landsat and MODIS data retrieval, data processing and the comparison between the Landsat and MODIS-derived results.

2.3 Landsat TM/ETM+ data incorporation

The use of the Landsat data is often limited due to its poor temporal coverage. It can be overcome to some extent if TM and ETM+ images can be incorporated. The incorporation method assumes that there is a simple reflectance ratio between a TM image and an ETM+ image on the same or close date (Chen et al. 2003, Jackson et al. 2004, Chen et al. 2005, Huang 2006). This hypothesis was first proposed in Chen et al. (2003) by comparing a TM image on 16 July 2002 to an ETM+ image on 17 July 2002. The ratio between the reflectance of crop fields on TM (16 July) and the reflectance of the same crop fields on ETM+ (17 July) were obtained. It is further hypothesized that the ratio can also be extended to other days. The time series of the Landsat TM/ETM+ has smooth development over time after they are incorporated together (Chen et al. 2003, Chen et al. 2005, Huang 2006). Huang (2006) advanced the investigation by exploring the sensitivity of this ratio to two vegetation species, corn and soybeans, both of which are dominant crops within the Great Plains of the US. The ratios were obtained at both regional level and field level for comparison purposes. Tables 1 and 2 list all the ratios for comparison. The results in table 1 indicate that this ratio only varies slightly from the corn field to the soybean field. The regional averaged ratio and the field averaged ratio are also comparable to each other. As observed in table 1, this ratio is not species dependent;

Table 1. Surface reflectance at seven bands of TM and ETM+, and incorporation ratios from TM to ETM+: (a) regional level.

<table>
<thead>
<tr>
<th>Band</th>
<th>TM, 16 July</th>
<th>ETM+, 17 July</th>
<th>Incorporation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3186</td>
<td>0.3179</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>0.1249</td>
<td>0.2502</td>
<td>2.00</td>
</tr>
<tr>
<td>3</td>
<td>0.1204</td>
<td>0.2028</td>
<td>1.68</td>
</tr>
<tr>
<td>4</td>
<td>0.4551</td>
<td>0.4481</td>
<td>0.98</td>
</tr>
<tr>
<td>5</td>
<td>0.3529</td>
<td>0.3860</td>
<td>1.09</td>
</tr>
<tr>
<td>7</td>
<td>0.1161</td>
<td>0.1945</td>
<td>1.68</td>
</tr>
</tbody>
</table>

Table 2. Surface reflectance at seven bands of TM and ETM+, and incorporation ratios from TM to ETM+: (b) crop and soybean field-average level.

<table>
<thead>
<tr>
<th>Reflectance at corn sites</th>
<th>Reflectance at soybean sites</th>
<th>Incorporation ratio average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band</td>
<td>TM, 16 July</td>
<td>ETM+, 17 July</td>
</tr>
<tr>
<td>1</td>
<td>0.3065</td>
<td>0.2999</td>
</tr>
<tr>
<td>2</td>
<td>0.1151</td>
<td>0.2273</td>
</tr>
<tr>
<td>3</td>
<td>0.1025</td>
<td>0.1704</td>
</tr>
<tr>
<td>4</td>
<td>0.5055</td>
<td>0.4762</td>
</tr>
<tr>
<td>5</td>
<td>0.2797</td>
<td>0.3032</td>
</tr>
<tr>
<td>7</td>
<td>0.0693</td>
<td>0.1282</td>
</tr>
</tbody>
</table>
however, it is wavelength dependent. The fields’ averaged ratio shown in bold in table 2 was finally adopted to incorporate the TM (23 June) data to the ETM+ data series.

3. Methodology

3.1 Atmospheric correction

In addition to the 6S atmospheric model, the atmospheric correction of Landsat TM/ETM+ was also conducted using a newly proposed hybrid atmospheric correction algorithm, namely Black/White Atmospheric Correction (BWAC) model. The hybrid model combines the strengths of atmospheric modelling and empirical approaches. Similar to an empirical atmospheric correction, the hybrid model uses ground reference points. It is based on three theories: first, deep clean water has the lowest possible reflectance (appears black in Landsat 4-3-5 false colour composite in figure 2(a)); in contrast, highly reflective materials such as pavement or rock outcrops have the highest possible reflectance in the region (appears white in Landsat 4–3-5 false colour composite in figure 2(a)). The former and the latter represent the low and the high limits of the reflectance range in the studied region respectively. Second, if the radiometers are static, the reflectances of the reference points do not have significant temporal changes when their condition is persistent. Therefore, if atmospheric correction can be done successfully by an atmospheric model on one clear day which has the most data availability, the atmospherically corrected reflectance of these two reflectance limits can be extended to the other days. Lastly, the atmospheric influence under a homogenous local atmospheric condition is linear. Therefore, reflectance of the other pixels on any other days can be linearly stretched between the two limits. Thus, it is vital that such water body and highly reflective areas are carefully chosen as reference points.

Using an atmospheric model to determine the reference reflectance may still introduce numerical uncertainties. The most accurate method is to directly measure the reference sites using a field spectrometer. However, such a practice requires large-scale field deployment which is not feasible or efficient in most cases. Therefore, intentionally the hybrid BWAC approach is solely based on satellite data. It combines the strength of theoretical approach in calculating reference reflectance accurately and the strength of empirical approach with its simplicity. It avoids the difficulties in retrieving complicated weather parameters while maintaining reasonable accuracy in atmospheric correction.

In the BWAC scheme, the atmospheric corrected reflectance for any pixel on the data acquisition day can be calculated using the following equation:

\[
R_{\text{corrected}} = \frac{R - R_B}{R_W - R_B} (R_{W0} - R_{B0}) + R_{B0},
\]

where \(R_{B0}\) and \(R_{W0}\) are the reference reflectances which are the atmospheric corrected low and high limits respectively calculated in an atmospheric model; \(R_B\) and \(R_W\) are the raw reflectances of the two ground reference points respectively; \(R\) is the raw reflectance of the data acquisition pixel on the data acquisition day; \(R_{\text{corrected}}\) is the final atmospheric corrected surface reflectance for the corresponding pixel.

A step-by-step procedure is summarized as follows:

1. Reference point selection: select reference points which have the best possible potential to represent the low and high limits of the reflectance range in the region.
2. Reference date determination: decide which date is the most suitable for the atmospheric model. A date with the most availability of model inputs is preferred so that the numerical uncertainties in the model can be minimized.
3. Reference reflectance calculation: use an atmospheric model to calculate the atmospherically corrected reflectance of the reference points.
4. Atmospheric correction by the BWAC model: There are five inputs: two reference reflectance ($R_{B0}$, $R_{W0}$) on the reference date, and three raw reflectance ($R_B$, $R_W$, $R$) of the two reference points and data acquisition pixel on the data acquisition date. By linearly stretching the raw reflectance range $[R_B, R, R_W]$ to the atmospherically corrected reflectance range $[R_{B0}, R_{corrected}, R_{W0}]$ using equation (5), the atmospherically corrected reflectance of the data acquisition pixel on the data acquisition date ($R_{corrected}$) can be calculated.

3.2 Reflectance unmixing

If vegetation coverage is known, soil reflectance in the satellite signal can be extracted and removed through a reflectance unmixing process. The unmixed vegetation reflectance is more attributable to vegetation properties, such as vegetation water content and dry biomass, while soil reflectance contamination is minimized.

The linear mixture model discussed by Jasinski (1996) and Townshend et al. (2000) provided a method of removing sub-pixel soil effect. The model assumes that the reflectance ‘seen’ at the satellite sensor is the additive sum (or linear mixture) of individual elements:

$$f_V R_V + f_S R_S = R,$$

where $f_V$ and $f_S$ are the fractional vegetation and soil coverages respectively, $R_V$ and $R_S$ represent the reflectances of pure vegetation and bare soil; $R$ is the ‘mixed’ reflectance ‘seen’ at satellite. The constraints of equation (6) are:

$$f_V + f_S = 1, f_V \geq 0, f_S \geq 0.$$

Equations (6) and (7) are wavelength-dependent. They are only true if accurate end-members have been chosen and there are no other sub-pixel materials outside the particular vegetation type and soil in the pixel (Tompkins et al. 1997, Dennison and Roberts 2003). The linear mixture model is based on the assumption that each cover type contributes linearly to pixel reflectance, and nonlinear interactions between end-members are negligible (Asner and Lobell 2000). Outstanding applications of the linear reflectance mixture model can be found in Gilabert et al. (2000), Maas (2000), Hu et al. (2004), and Shanmugama et al. (2006).

Since the average scale of crop fields (~800 m) is much larger than the Landsat pixel resolution (30 m), it is reasonable to assume the end-members in each crop pixel (pixels identified as corn or soybeans fields) are either corn and soil, or soybeans and soil. For each site in a crop field, the surface reflectance at the nearest bare soil pixel in the same crop field was used as the sub-pixel bare soil reflectance for that site. Bare soil pixels can be identified using the ‘soil strip’ in the NIR-Red spectral scatter plot (Baret et al. 1993).
In equation (6), the fractional sub-pixel soil reflectance $f_s R_s$ can be extracted from the mixed reflectance $R$. The resultant $R - f_s R_s$ is then considered as the fractional sub-pixel vegetation reflectance. Therefore, the reflectance $R_v$ can be calculated as

$$R_v = \frac{R - f_s R_s}{f_v},$$

where $R_v$ is considered as ‘pure’ reflectance from the vegetation itself, like ideally measuring a vegetation area without any sub-pixel bare soil contamination in spectral signals.

A step-by-step procedure is summarized as follows:

1. Vegetation coverage evaluation: $f_v$ and $f_s$ can be estimated by an empirical relationship between vegetation coverage and vegetation index.
2. Bare soil reflectance approximation: in this study, the bare soil reflectance $R_s$ is approximated as the reflectance from the nearest bare soil pixel.
3. Vegetation reflectance calculation: use equation (8) to unmix sub-pixel bare soil reflectance and to calculate vegetation reflectance.

4. Results

4.1 Atmospheric corrected results from BWAC

To implement the BWAC scheme in this study, a deep-water reservoir and a large section of pavement at the edge of Ames were identified as reference sites in the SMEX02 area. The two sites are highlighted in figure 2(a). The true colour high-resolution photographs are also provided in figure 2(b) to show their identities and dimensions. The reference reflectances of the reference sites were calculated by the 6S model on 1 July 2002. These reference reflectances were then applied to the other dates following the prescribed steps in §3.1.

Taking the atmospheric correction on 8 July 2002 as an example, figure 3 compares the results from the BWAC scheme and the 6S model. The corrected spectral reflectances from two approaches were remarkably comparable for both corn and soybeans. This comparability is particularly true for the SWIR bands, which are the focus of this study. This shows the potential of the BWAC as a simple, effective alternative to the 6S and MODTRAN models for atmospheric correction. The visible reflectance derived from the BWAC scheme, however, appeared relatively low compared to that from the 6S model. The corrected reflectance of the blue band was unrealistically negative. The reason is believed to be the imperfect selection of the reservoir as one reference site. In figure 2(b), the true colour of the reservoir appeared green (as a result of algae) rather than deep blue. Therefore, the crop reflectance at the blue band can sometimes be even lower than the water body and consequently the corrected reflectance of the crop at the blue band is negative. When a perfect deep water body is available as a reference site, the BWAC scheme can be extended to visible bands.

The BWAC scheme is a good alternative to atmospheric models when the availability of input variables is poor. Because of the sufficient input data availability and desired better accuracy in visible bands, the 6S atmospherically corrected spectral reflectances were used for the subsequent soil reflectance unmixing and vegetation index calculations. Considering the comparability of
reflectance from the BWAC scheme and the 6S model in figure 3, the results in the following sections should not have changed significantly if the BWAC-derived reflectance had been used instead.

4.2 Sub-pixel soil reflectance unmixing

4.2.1 Vegetation coverage mapping. In figure 4, the vegetation coverage is plotted against the NDVI and the NDWIs. A notable feature in the relationship between the NDVI and vegetation coverage is that corn samples are clustered at high NDVI values. It indicates the NDVI saturation, a situation where there was no increase in NDVI value while vegetation coverage increased from 80% to 100%. This is due to the use of the red band as measurement band in the NDVI, because the red band is sensitive to the chlorophyll contents which quickly saturate in early vegetation growth stages (Gamon et al. 1995, Gao 1996). Despite its high $R^2$ value (0.73), this saturation phenomenon suggests an improper use of the NDVI to predict high vegetation coverage via simple linear models. In contrast, the NDWI$_{1640}$ and NDWI$_{2130}$ achieved comparable $R^2$ values of 0.670 and 0.676 respectively without significant early saturations. When vegetation coverage is high (80–100%), the SWIR-based NDWIs are more sensitive to vegetation coverage growth in comparison with the NDVI and the NDWI$_{1240}$. It is attributable to the use of SWIR (1640 nm, 2130 nm) bands in these indices because these bands are sensitive to water content and biomass which grow as vegetation coverage increases.

Another interesting feature for corn and soybeans data in SMEX02 is that the relationship between vegetation indices (NDVI, NDWI$_{1240}$, NDWI$_{1640}$, and NDWI$_{2130}$) and vegetation coverage seems crop independent. Figure 4(b) shows a relationship between vegetation coverage and the NDWI$_{1640}$ using both corn and soybeans data. The NDWI$_{1640}$ is chosen because it achieves the best linear correlation coefficient when both corn and soybeans data are plotted together. The vegetation coverage of both corn and soybeans grew in similar paces as the NDWI$_{1640}$ developed. To test this species-independency of the relationship, vegetation coverage of corn and soybeans derived separately from different linear
regression models in figure 4(a) were compared with that derived from the identical NDWI\textsubscript{1640} model in figure 4(b). The difference between the results was less than 10%. The implication is that the vegetation coverage can be mapped via an identical model for corn and soybeans without using a classification map for this study. This

Figure 4. Linear estimation of vegetation coverage using different vegetation indices: (a) NDVI and NDWIs for corn; (b) NDWI\textsubscript{1640} for both corn and soybeans.
vegetation species independency is particularly valuable when a classification map (which usually requires field visits) is inaccessible. Whether such independency can be extended to other vegetation species requires further verification.

The NDWI$_{1640}$ regression model in figure 4(b) was used to map vegetation coverage. The vegetation coverage map for DOY 182 (1 July 2002) is shown in figure 5(a) in comparison with the classification map of figure 5(b). On 1 July, corn fields were more vegetated (appear white in figure 5(a)) than soybeans fields (appear black or grey in figure 5(a)). It is also worthwhile noting that although the identical vegetation coverage regression model is designed for corn and soybeans, the forest in SMEX02 area also appears white in figure 5(a). Because the forest sites were indeed largely vegetated on 1 July, it is plausible that the vegetation species independency of the relationship between vegetation coverage and vegetation index can be extended to at least certain vegetation types. Unfortunately, there was no vegetation cover survey for forest sites in SMEX02 for further verification.

Figure 5. (a) NDWI$_{1640}$ estimated vegetation coverage (%) map for 1 July 2002. (b) Supervised classification map on 1 July 2002, 30 m. Green: corn; red: soybeans; navy blue: urban and roads; yellow: grasses; sea green: forest. The estimated vegetation coverage in (a) is more meaningful for corn and soybeans classes than the other classes in (b).
A series of vegetation coverage maps were generated for all satellite data acquisition dates. The vegetation coverage information obtained from these maps was used as data input in the sub-pixel soil reflectance unmixing model.

4.2.2 **Bare soil reflectance approximation.** Besides vegetation coverage, sub-pixel soil reflectance is another important input to build the linear mixture model. There are various factors influencing soil reflectance, such as soil moisture, soil type, soil texture, and vegetation roots fraction (Lobell and Asner 2002). To avoid the effects from these factors, the reflectance at the nearest bare soil pixel provides the best approximation of the sub-pixel soil reflectance. The main reason is that the conditions of those influential factors in the nearby pixels are closest to their conditions in the desired pixel. Bare soil pixels are identified by the ‘soil strip’ in the NIR-Red spectral scatter plot (figure 6). For each site, the surface reflectance at the nearest bare soil pixel was used as the sub-pixel soil reflectance for that site.

This approach has several significant advantages: first, the data were acquired on the same date under the same weather conditions. It avoided temporal variation of external forcing which may change soil moisture. Second, if the soil pixels are fairly close to the crop pixels, spatial variation of soil moisture is minimal and soil type is consistent. If the bare soil pixel has soil properties significantly different from the desired sub-pixel, the approach can be problematic. Therefore, the bare soil pixel was restrained to be located in the same crop field in this study to keep the mismatching error to a minimum. A good example is one SMEX02 site (WC25) which was assigned to a sandy hill slope. Because the soil type of all the rest sites is clay, the sub-pixel soil reflectance of this site is not accurately matched by the close bare soil pixel.

![Figure 6. ‘Soil strip’ in the NIR-Red spectral scatter plot used to identify bare soil pixels in the corn and soybeans fields.](image)

2088  
*J. Huang et al.*

Downloaded By: [University of Miami] At: 17:54 30 April 2009
4.2.3 Results of soil reflectance unmixing. The temporal variations of surface reflectance at seven spectral bands before and after the sub-pixel soil reflectance unmixing are plotted in figure 7. The unmixed reflectances are significantly different from their original counterpart. The changes in curves can be physically explained by the physical facts in figure 1. For example, for the visible bands (485 nm, 560 nm, 660 nm) and the SWIR bands (1650 nm, 2215 nm), the bare soil reflectance is higher than the pure vegetation reflectance. After unmixing, the reflectances at these bands decrease to be closer to the vegetation reflectance. The decreases are particularly noticeable during the early days (DOY 174–182) when vegetation coverage was low. In contrast, for NIR bands (830 nm, 1240 nm), the bare soil reflectance is lower than the pure vegetation reflectance. Thus, the unmixed reflectances increase to approach the pure vegetation reflectance in figure 7. It is also more noticeable at the early stages of vegetation growth when bare soil is the majority of the fields. At later stages of the crop’s growth (8–17 July), the soil effect significantly decreases because of the increasing vegetation cover. On 17 July, the reflectances at each band before and after unmixing are fairly close to each other. The consistency of these changes with the spectral signatures of vegetation and soil in figure 1 evidences the effectiveness of this sub-pixel soil reflectance unmixing scheme.

Without the sub-pixel bare soil reflectance contamination, the changes in unmixed reflectances in the satellite signal represent the true response to the ground vegetation growth. In figure 7, after unmixing, the decreasing trend in the temporal variation curves of the SWIR bands (1650 nm, 2215 nm) persisted. Since VWC was increasing throughout the SMEX02 period, such persistence shows their water absorption features of SWIR bands. In other words, they are more correlated to the VWC growth than other bands due to their water absorption features. In the period when the VWC growth was dominant in the vegetation growth, the SWIR (1650 nm) had the most significant decreasing percentage in surface reflectance. Similar decreasing tendency is also observed for the NIR (1240 nm) band due to its water absorption behaviour which is relatively weaker than the SWIR bands.

To summarize the sub-pixel soil reflectance contamination quantitatively, table 3 lists the percentage of the reflectance change with respect to the mixed reflectance. Such a measure indicates how sensitive the accuracy of vegetation reflectance at each band is to the level of sub-pixel soil reflectance contamination. The higher the percentage, the easier it is for the vegetation reflectance at this band to be affected by soil reflectance. For corn, the soil effects at seven bands ranged from 4.1% to 32.6% on DOY 174. This range was much less on DOY 198, which is only from 0.3% to 5.8%. Similarly for soybeans, these ranges were 6.9–81.9% on DOY 174 and 0.9–14.5% on DOY 198. Quantitatively, this again verified that the sub-pixel soil reflectance contamination is more significant during the early stages of vegetation growth than the later stages.

Since the sub-pixel soil reflectance unmixing produces higher reflectance at the reference band (i.e. NIR, 830 nm) and lower reflectance at the measurement bands (i.e. red 648 nm, NIR 1240 nm, SWIR 1650 nm, SWIR 2215 nm in NDVI, NDWI_{1240}, NDWI_{1640} and NDWI_{2130} respectively) in the vegetation indices, it is worthwhile to determine how much the sub-pixel soil reflectance contamination may influence the vegetation indices quantitatively. Figure 8 presents the temporal variation curves of the four vegetation indices before and after the sub-pixel soil reflectance unmixing. In general, all the vegetation indices increase after the sub-pixel soil reflectance unmixing. The percentages of the vegetation indices changes in the sub-pixel soil reflectance unmixing...
due to the sub-pixel soil reflectance contamination are summarized in Table 4.
Because the changes of vegetation indices are all positive, the negative percentages
of NDWI$_{1240}$ in the early days are due to its original negative value. The percentages
describe how sensitive a vegetation index can be to the sub-pixel soil reflectance
contamination. The changes attributed to the contamination range from 0.9%
(NDVI on DOY 198 when fields were mostly fully vegetated) to 181.8% (NDWI$_{1240}$
on DOY 174, primarily due to its low absolute value and the large unvegetated
fields). The results indicate the importance and necessity of conducting sub-pixel soil
reflectance unmixing for vegetation indices. Otherwise, the changes in vegetation
indices directly calculated from satellite signals cannot be taken as pure responses to
vegetation growth on the ground.

### 4.3 Vegetation properties estimation

Empirically the sub-pixel soil reflectance contaminated vegetation indices are widely
used to estimate VWC and VDB. For example, Chen et al. (2005) used the SWIR-
based NDWI to predict VWC with reasonable accuracy. Theoretically, however,
such vegetation indices contain soil reflectance information which has nothing to do
with vegetation properties. Without removing the sub-pixel soil reflectance, the
vegetation index distorts the true relationship between the vegetation index and the
VWC in modelling. The VWC estimation using the original Landsat- and MODIS-
derived NDWIs (without sub-pixel soil reflectance unmixing) was fully discussed in
Chen et al. (2005). The study demonstrated that the SWIR based NDWI have great
potential in VWC modelling although the soil effects were not yet corrected. How
well the sub-pixel soil reflectance unmixed vegetation indices (NDVI, NDWI$_{1240}$,
NDWI$_{1640}$, and NDWI$_{2130}$) work with VWC and dry biomass estimation are
presented in this study. Although the unmixed vegetation indices are not necessarily
better than the traditionally used vegetation indices in empirical approaches to
estimate vegetation contents, those who are interested to compare them can read

Table 3. Percentage of surface reflectance change attributable to sub-pixel soil reflectance
contamination over total surface reflectance at seven spectral bands: ($R_{\text{unmixed}} - R_{\text{mixed}}$) / $R_{\text{mixed}} \times 100$.

<table>
<thead>
<tr>
<th>%</th>
<th>485 nm</th>
<th>560 nm</th>
<th>660 nm</th>
<th>830 nm</th>
<th>1240 nm</th>
<th>1640 nm</th>
<th>2130 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Corn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>−18.75</td>
<td>−20.66</td>
<td>−32.61</td>
<td>15.92</td>
<td>4.14</td>
<td>−7.86</td>
<td>−25.04</td>
</tr>
<tr>
<td>182</td>
<td>−31.44</td>
<td>−28.43</td>
<td>−57.82</td>
<td>5.03</td>
<td>1.09</td>
<td>−7.51</td>
<td>−25.11</td>
</tr>
<tr>
<td>189</td>
<td>−2.02</td>
<td>−2.07</td>
<td>−5.39</td>
<td>0.72</td>
<td>0.05</td>
<td>−1.51</td>
<td>−6.37</td>
</tr>
<tr>
<td>198</td>
<td>−2.93</td>
<td>−2.60</td>
<td>−5.76</td>
<td>0.56</td>
<td>0.34</td>
<td>−1.48</td>
<td>−5.00</td>
</tr>
<tr>
<td>(b) Soybeans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>−55.71</td>
<td>−81.87</td>
<td>−76.58</td>
<td>22.65</td>
<td>8.32</td>
<td>−6.89</td>
<td>−10.55</td>
</tr>
<tr>
<td>182</td>
<td>−26.50</td>
<td>−54.59</td>
<td>−57.46</td>
<td>17.20</td>
<td>8.84</td>
<td>−15.97</td>
<td>−30.69</td>
</tr>
<tr>
<td>189</td>
<td>−9.75</td>
<td>−6.94</td>
<td>−15.55</td>
<td>10.35</td>
<td>2.02</td>
<td>−5.67</td>
<td>−6.11</td>
</tr>
<tr>
<td>198</td>
<td>−7.55</td>
<td>−4.62</td>
<td>−14.53</td>
<td>5.88</td>
<td>1.70</td>
<td>2.42</td>
<td>−0.94</td>
</tr>
</tbody>
</table>

Figure 7. Temporal variations of the six Landsat bands and one MODIS NIR (1240 nm)
band before and after the sub-pixel soil reflectance unmixing: (a) corn, SMEX02; (b)
soybeans, SMEX02. ‘sc’ denotes ‘soil effect corrected’. All unmixed bands are in solid lines
and all original bands are in dashed lines.
Chen et al. (2005) to obtain more details of the empirical VWC estimation using the traditional NDVI and the NDWIs.

4.3.1 VWC estimation. The SWIR-based NDWIs, after sub-pixel soil reflectance unmixing, are tested for the VWC estimation purposes. Figure 9 and figure 10 present the results for corn and soybeans respectively. Since vegetation data were collected in two time periods (DOY 166–169 for round 1, and DOY 178–189 for round 2), data are plotted with two rounds clearly indicated. Due to the cloud and cloud shadow contamination on MODIS image during DOY 166–169, meaningful MODIS surface reflectances were retrieved only for three out of six sites. Additionally, because Landsat data were only available for limited days (5 days in

Figure 8. Temporal variations of four vegetation indices before and after the sub-pixel soil reflectance unmixing: (a) corn, SMEX02; (b) soybeans, SMEX02. 'sc' denotes 'soil effect corrected'. All unmixed vegetation indices are in solid lines and all original vegetation indices are in dashed lines.
total), the time series of vegetation indices are extrapolated to those ground measurement data acquisition days. A cubic polynomial method is used in the data extrapolation.

Four data points were excluded from linear regression models. Three points were from the site WC25 which is a particular site designated to a sandy hill slope where the vegetation growth was notably stunted (Anderson et al. 2004). In figure 9, the NDWI\textsubscript{1640} and NDWI\textsubscript{2130} for site WC25 are generally higher than the rest data when the VWC levels are comparable. It is because the soil type of this site is sand and that of other crop fields is clay (Crosson and Laymon 2003). The soil type of the nearby bare soil pixel, which is used to obtain the sub-pixel bare soil reflectance for the WC25, is clay. Thus, the mismatching of sub-pixel soil type and soil reflectance resulted in the abnormal behaviours of the biased data points from the site WC25. It consequently affected the accuracy of the reflectance unmixing for the site WC25. This signifies that, as a significant factor to influence soil reflectance, soil type information is very important in conducting sub-pixel soil reflectance unmixing. One more data point was identified as data for the site WC31 on DOY 186. It had vegetation coverage of 63\% while it had been 90\% on an earlier date of DOY 179. It may either indicate a measurement error or a corn growth cycle different from the other sites.

Two kinds of VWC variables are considered in their correlations to vegetation indices. One is the ‘total areal VWC (kg/m\textsuperscript{2})’ which is defined as the weight of the vegetation water content per unit area referring to the entire pixel. The total VWC in weight (kg) within the pixel can be calculated as the areal VWC (kg/m\textsuperscript{2}) multiplying the total area of the pixel; because the total area of a pixel is temporally constant, the areal VWC (in kg/m\textsuperscript{2}) is as indicative as the total VWC in weight (in kg). Therefore, it includes information of both horizontal expansion and vertical clumping of vegetation growth. The second VWC variable is defined as ‘vegetated areal VWC (kg/m\textsuperscript{2})’ which is calculated by referring to the vegetated part of the pixel only. This variable only counts for the vertical clumping of the vegetation growth. The relationship between the two VWC variables is: ‘vegetated areal VWC’ = ‘total areal VWC’ / Vegetation Coverage.

We believe that both the horizontal and vertical growth of vegetation contribute to the change in vegetation reflectance. However, there is little evidence to

<table>
<thead>
<tr>
<th>%</th>
<th>NDVI</th>
<th>NDWI\textsubscript{1240}</th>
<th>NDWI\textsubscript{1640}</th>
<th>NDWI\textsubscript{2130}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Corn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>12.25</td>
<td>-181.63</td>
<td>46.05</td>
<td>22.48</td>
</tr>
<tr>
<td>182</td>
<td>9.03</td>
<td>38.02</td>
<td>13.40</td>
<td>11.53</td>
</tr>
<tr>
<td>189</td>
<td>0.83</td>
<td>2.81</td>
<td>1.74</td>
<td>1.71</td>
</tr>
<tr>
<td>198</td>
<td>0.89</td>
<td>1.19</td>
<td>1.59</td>
<td>1.41</td>
</tr>
<tr>
<td>(b) Soybeans</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>67.92</td>
<td>-44.11</td>
<td>-209.61</td>
<td>48.59</td>
</tr>
<tr>
<td>182</td>
<td>28.22</td>
<td>-64.02</td>
<td>1055.90</td>
<td>72.16</td>
</tr>
<tr>
<td>189</td>
<td>5.20</td>
<td>66.44</td>
<td>35.64</td>
<td>10.73</td>
</tr>
<tr>
<td>198</td>
<td>2.61</td>
<td>13.02</td>
<td>3.89</td>
<td>2.73</td>
</tr>
</tbody>
</table>

For all cases, VI\textsubscript{unmixed} - VI\textsubscript{mixed} are positive. A negative value for this percentage indicates a negative VI\textsubscript{mixed}.
(a)

(b)
quantitatively compare the two parts of contributions in the literature. The physical theoretical basis can be understood in the following two cases:

1. Case 1: If the vegetation is only growing horizontally, its local vegetated areal VWC does not increase; however, vegetation coverage and total VWC weight (in kg) will increase. In this case, it is the horizontal expansion of the vegetation that contributes to the change in vegetation reflectance within the pixel. Physically, one reasonable mechanism is that the neighbouring effect can complicate light pathways and consequently increase VWC absorption and reduce vegetation reflectance.

2. Case 2: If the vegetation only grows vertically, leaf layers will be stacked without vegetation coverage growth. In this case, the change in vegetation reflectance is only attributable to the change in local vegetated areal VWC. This is easier to understand because more leaf layers result in more VWC absorption and less vegetation reflectance.

Therefore, it is intriguing to link both VWC properties to the unmixed vegetation indices to see the comparability of the two parts of contributions to the change in vegetation reflectance.

Figures 9 and 10 illustrate the relationships between the unmixed vegetation indices and the two kinds of VWC variables, for corn and soybeans respectively. Part (a) refers to the total areal VWC, and part (b) refers to the vegetated areal VWC. The best-fitted linear correlations with minimum root-mean-square-error (RSME) are presented. The parameters of the two-tailed t-test on the correlation significance are summarized in table 5.

In the total areal VWC estimations in figures 9(a) and 10(a), NDVI exhibits the saturation problem when the total areal VWC exceeded 2.5 kg/m² for corn or 0.4 kg/m² for soybeans. The data points cluster at high NDVI values of 0.85–0.9. In comparison with the NDWI, the NDVI obtains the lowest coefficient of determination ($R^2$) and highest RMSE in the total areal VWC estimation models in figures 9(a) and 10(a): $R^2 = 0.210$ and RMSE = 0.841 kg/m² for corn, and $R^2 = 0.080$ and RMSE = 0.285 kg/m² for soybeans. For corn VWC estimation, the NDWI$_{1640}$ has relatively higher $R^2$ value and lower RMSE than the NDWI$_{1240}$ and NDWI$_{2130}$ in linear models: 0.616 versus 0.310 and 0.475 in $R^2$, 0.679 kg/m² versus 0.794 kg/m² and 0.757 kg/m² in RMSE; for soybeans, however, the NDWI$_{1240}$ seems superior over the NDWI$_{1640}$ and NDWI$_{2130}$ instead: 0.481 versus 0.311 and 0.308 in $R^2$, 0.214 kg/m² versus 0.247 kg/m² and 0.247 kg/m² in RMSE. Seen as $p$ value and critical correlation coefficient at 95% confidence level, the significance of the correlations is also listed in table 5. The results of soybeans may have low reliability because of the difficulties in ground measurements as a result of the various planting practices and low vegetation properties amounts.

In the vegetated areal VWC estimation in figures 9(b) and 10(b), the general conclusions on the superiority of vegetation indices are the same as those from part (a). The $R^2$ values in linear correlations deteriorate in general when vegetation indices are correlated to the vegetated areal VWC instead of the total areal VWC.

Figure 9. Relationship between corn VWC and the sub-pixel reflectance unmixed NDVI and NDWI (upper left: VWC vs NDVI; upper right: VWC vs NDWI$_{1240}$; lower left: VWC vs NDWI$_{1640}$; lower right: VWC vs NDWI$_{2130}$). The outlier data of site WC25 and site WC31 were indicated by the pointed arrows. (a) Unmixed VI and total areal VWC. (b) Unmixed VI and vegetated areal VWC.
For example, for NDWI1640 in the case of corn, the $R^2$ value decreased from 0.616 to 0.540; for NDWI1240 in the case of soybeans, the $R^2$ value decreased from 0.481 to 0.140. But both correlations were still significant at the 95% confidence level (table 5). This indicates that in addition to the vertical clumping of vegetation leaf layers, the horizontal expansion of the vegetation canopy also plays a role in changing vegetation reflectance.

4.3.2 VDB estimation. Similar to the VWC estimation, the VDB of corn and soybeans are also estimated using the unmixed vegetation indices. Both the total areal VDB and the vegetated areal VDB are tested. Using the ground measurements in SMEX02, figure 11 demonstrates that correlation between the VWC and the VDB growths are significant at 99% confidence level. The $R^2$ values are 0.889 and 0.9679 for corn and soybeans respectively. Therefore, the behaviours of the vegetation indices in the VDB linear estimation models are similar to those observed in the VWC estimation.

The saturation of NDVI is also observed in the biomass estimation. As seen in table 5, it achieved low $R^2$ values of 0.306 (0.067) and high RMSE of 0.176 kg/m$^2$ (0.074 kg/m$^2$) in linear models for corn (soybeans). The representative results from the NDWI1240 and NDWI1640 are presented in figure 12 with linear VDB estimation models. The $R^2$ values in the linear models of the NDWI1640 are 0.606 for corn and 0.315 for soybean. In comparison, the $R^2$ values in the linear models of the NDWI1240 are 0.392 for corn and 0.552 for soybeans. The superiority of NDWI1640 in the case of corn and of NDWI1240 in the case of soybeans are also seen when the vegetated areal VDB substitutes the total areal VDB.

In the case of soybeans, the NDWI1240 achieves the best correlation in both VWC and VDB estimations; however, the following factors limit its utilization: the large uncertainties of soybeans datasets, the moderate MODIS data resolution, and the strong sensitivity to sub-pixel soil reflectance contamination (figure 8 and table 4). Therefore, the NDWI1640 is more preferable in practice.

5. Discussions and conclusions

With support from the ground measurement of SMEX02, we used Landsat TM/ETM+ and Terra-MODIS data to explore their feasibility to estimate vegetation properties, such as vegetation coverage, VWC and dry biomass. An atmospheric correction scheme and a sub-pixel soil reflectance unmixing approach were proposed.

As a hybrid approach to correct the atmospheric effect, the BWAC scheme was introduced to atmospherically correct the Landsat images. This hybrid model combines the strength of the theoretical atmospheric model in calculating reference reflectance accurately and the strength of the empirical approach in avoiding the model’s dependency on various meteorological inputs. It can achieve comparable accuracy to the 6S atmospheric model. It particularly performs well for the SWIR bands. Its less satisfactory performance with visible bands seems to be associated with the selection of the reference objects.

Figure 10. Relationship between soybeans VWC and the sub-pixel soil reflectance unmixed NDVI and NDWI (upper left: VWC vs NDVI; upper right: VWC vs NDWI1240; lower left: VWC vs NDWI1640; lower right: VWC vs NDWI2130).
Table 5. Parameters of linear correlation models and correlation significance tests for the relationship between vegetation properties (VWC, VDB) and vegetation indices.

<table>
<thead>
<tr>
<th></th>
<th>VI vs total areal VWC</th>
<th></th>
<th>VI vs vegetated areal VWC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>$R^2$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>NDVI</td>
<td>7.355</td>
<td>-2.978</td>
<td>0.210</td>
<td>0.841</td>
</tr>
<tr>
<td>NDWI$_{1240}$</td>
<td>25.29</td>
<td>1.310</td>
<td>0.310</td>
<td>0.794</td>
</tr>
<tr>
<td>NDWI$_{1640}$</td>
<td><strong>7.708</strong></td>
<td><strong>0.264</strong></td>
<td><strong>0.616</strong></td>
<td><strong>0.679</strong></td>
</tr>
<tr>
<td>NDWI$_{2130}$</td>
<td>10.148</td>
<td>-4.107</td>
<td>0.475</td>
<td>0.757</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>VI vs total areal VDB</th>
<th></th>
<th>VI vs vegetated areal VDB</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>$R^2$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>NDVI</td>
<td>1.223</td>
<td>-0.531</td>
<td>0.306</td>
<td>0.176</td>
</tr>
<tr>
<td>NDWI$_{1240}$</td>
<td>4.830</td>
<td>0.130</td>
<td>0.392</td>
<td>0.155</td>
</tr>
<tr>
<td>NDWI$_{1640}$</td>
<td><strong>1.473</strong></td>
<td><strong>-0.072</strong></td>
<td><strong>0.606</strong></td>
<td><strong>0.132</strong></td>
</tr>
<tr>
<td>NDWI$_{2130}$</td>
<td>2.069</td>
<td>-1.004</td>
<td>0.533</td>
<td>0.144</td>
</tr>
</tbody>
</table>

At the 95% confidence level, $N=50$ and $R_c^2=0.077$ for corn ($N=47$ and $R_c^2=0.082$ in case of MODIS NDWI1240)

(a) Corn, SMEX02

<table>
<thead>
<tr>
<th></th>
<th>VI vs total areal VWC</th>
<th></th>
<th>VI vs vegetated areal VWC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>$R^2$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.891</td>
<td>-0.297</td>
<td>0.080</td>
<td>0.285</td>
</tr>
<tr>
<td>NDWI$_{1240}$</td>
<td><strong>3.302</strong></td>
<td><strong>0.632</strong></td>
<td><strong>0.481</strong></td>
<td><strong>0.214</strong></td>
</tr>
<tr>
<td>NDWI$_{1640}$</td>
<td>0.847</td>
<td>0.331</td>
<td>0.311</td>
<td>0.247</td>
</tr>
<tr>
<td>NDWI$_{2130}$</td>
<td>0.972</td>
<td>-0.010</td>
<td>0.308</td>
<td>0.247</td>
</tr>
</tbody>
</table>
Table 5. (Continued.)

(b) Soybeans, SMEX02

<table>
<thead>
<tr>
<th></th>
<th>VI vs total areal VDB</th>
<th></th>
<th>VI vs vegetated areal VDB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>$R^2$</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.209</td>
<td>-0.067</td>
<td>0.067</td>
</tr>
<tr>
<td>NDWI$_{1240}$</td>
<td>0.906</td>
<td>0.158</td>
<td>0.552</td>
</tr>
<tr>
<td>NDWI$_{1640}$</td>
<td>0.219</td>
<td>0.077</td>
<td>0.315</td>
</tr>
<tr>
<td>NDWI$_{2130}$</td>
<td>0.247</td>
<td>-0.009</td>
<td>0.302</td>
</tr>
</tbody>
</table>

$N=33$ and $R^c_2=0.12$ for soybeans; $R_c$ is the critical correlation coefficient.

Model: $Y=AX + B$; $R$: correlation coefficient; $\sigma$: root-mean-square error; $p$-value: probability for a non-directional $t$-test on the significance of correlation coefficient. The parts shown in bold are the best vegetation indices which can be employed for corn/soybeans VWC and VDB estimation respectively.
The insufficient vegetation coverage information and the difficulty of approximating sub-pixel soil reflectance are main problems in sub-pixel soil reflectance unmixing. Using water-absorbent bands, the SWIR based NDWI appeared superior over the NDVI in vegetation coverage mapping. The sub-pixel bare soil reflectance was best approximated from the nearest bare soil pixel, identified by the 'soil strip' in the NIR-Red spectral scatter plot. This approach minimizes the mismatching of soil moisture, soil type, soil structure, and vegetation root fraction.

With the estimated vegetation coverage and the approximated bare soil reflectance, a simple linear reflectance mixture model was adopted to unmix sub-pixel soil reflectance from vegetation reflectance. The changes to the spectral bands

![Figure 11](https://example.com/figure11.png)

Figure 11. Linear relationships between VDB and VWC for (a) corn and (b) soybeans.

![Figure 12](https://example.com/figure12.png)

Figure 12. Relationships between vegetation dry biomass and the sub-pixel reflectance unmixed NDWI (upper left: VDB vs NDWI_{1240} for corn; upper right: VDB vs NDWI_{1640} for corn; lower left: VDB vs NDWI_{1240} for soybeans; lower right: VDB vs NDWI_{1640} for soybeans). (a) Unmixed VI and total areal VDB; (b) unmixed VI and vegetated areal VDB.
Sub-pixel reflectance unmixing in VWC and VDB estimation

(a)

(b)
due to unmixing are consistent with the physical truth of the difference between pure vegetation reflectance and bare soil reflectance. The fraction of the sub-pixel soil reflectance contamination was evaluated quantitatively. The significance of bare soil contamination is reduced as vegetation coverage increases. The contamination is most influential during the early vegetation growing stage, and can be negligible for fields in high vegetation cover.

The sub-pixel soil reflectance unmixed vegetation indices (NDVI, NDWI\textsubscript{1240}, NDWI\textsubscript{1640}, and NDWI\textsubscript{2130}) were calculated and compared for linear VWC and VDB estimation purposes. Without sub-pixel soil reflectance contamination, the relationships between the vegetation indices and vegetation properties reflect the true linkage between the changes of ground vegetation and the corresponding satellite spectral signals. Results indicate that the NDWI\textsubscript{1640} achieved the best linear correlation with the VWC and VDB for corn, and the NDWI\textsubscript{1240} was the best for soybeans. Taking some limitations of NDWI\textsubscript{1240} into account, the NDWI\textsubscript{1640} is overall preferable. It is in part attributable to the moderate water absorption feature of the SWIR (1640 nm), which corresponds better to the VWC growth. Results are believed applicable to crops and fields similar to those in SMEX02. Although in-depth investigations into other vegetation species are needed, results to date showed great potential of the water-absorbent-bands-based NDWI in evaluating various vegetation properties in satellite era.

Acknowledgements
The authors would like to extend thanks to Dr. T. Jackson, Dr. M. Anderson, and their colleagues at USDA for conducting intensive ground measurements in SMEX02 campaign and for making the data publicly available. The authors are also grateful for the valuable comments from the anonymous reviewers who helped to improve the manuscript significantly. The Overseas Research Scholarship from Universities UK to support Jingfeng Huang’s study at the University of Manchester is also appreciated.

References

ANDERSON, M., 2003, SMEX02 Watershed Vegetation Sampling Data, Walnut Creek, Iowa (Boulder, CO: National Snow and Ice Data Center).


CHEN, D., JACKSON, T.J., LI, F., COSH, M.H., WALTHALL, C. and ANDERSON, M., 2003, Estimation of vegetation water content for corn and soybeans with a Normalized

Cros, W. and Laymon, C., 2003, SMEX02 Soil Moisture and Temperature Profiles, Walnut Creek, Iowa (Boulder, CO: National Snow and Ice Data Center).


Huang, J., 2006, Vegetation properties relationships from spectral bands and vegetation indices from operational satellites. PhD dissertation, The University of Manchester, Manchester, UK.


JACKSON, T. and COSH, M., 2003, SMEX02 Watershed Vitel Network Soil Moisture Data, Walnut Creek, Iowa (Boulder, CO: National Snow and Ice Data Center).


