Detection of Ground Materials Using Normalized Difference Indices with a Threshold: Risk and Ways to Improve

Fen Chen 1,*, Tim Van de Voorde 2, Dar Roberts 3, Haojie Zhao 1 and Jingbo Chen 4

1 School of Resources and Environment, University of Electronic Science and Technology of China, 2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu 611731, China; 202022070308@std.uestc.edu.cn
2 Department of Geography, Ghent University, Krijgslaan 281, S8, 9000 Ghent, Belgium; Tim.VandeVoorde@UGent.be
3 Department of Geography, University of California, Santa Barbara, CA 93106, USA; dar@geog.ucsb.edu
4 Aerospace Information Research Institute, Chinese Academy of Sciences, 9 Dengzhuang South Road, Haidian District, Beijing 100094, China; chenjb@aircas.ac.cn

Abstract: In this short communication, we describe the shortcomings and pitfalls of a commonly used method to detect ground materials that relies on setting thresholds for normalized difference indices. We analyze this method critically and present some experimental results on the USGS and ECOSTRESS spectral libraries and on real Sentinel-2 and Landsat-8 images. We demonstrate the risk of commission errors and provide some suggestions to reduce it.

Keywords: normalized difference vegetation index; normalized difference water index; normalized difference soil index

1. Introduction

Normalized difference indices (NDIs) are widely used in the field of remote sensing. Well known examples include the normalized difference vegetation index (NDVI) [1], the normalized difference water index (NDWI) [2,3] and the normalized difference soil index (NDSI) [4]. There are several variants of these, and some scholars have pointed out the functional equivalence among different types of indices [5,6]. Many studies directly apply thresholds to NDIs to detect specific ground materials (e.g., vegetation, soil, water, build-up areas) or use thresholds on NDIs as a preliminary step for future analysis [3,4,7–17]. This short communication describes the shortcomings and risks of such a simple binary classification method, especially when applying it in complex heterogeneous environments or large areas.

2. Analysis

An NDI is generally designed as the ratio of the difference between reflectance values in two bands and the sum of the same values:

\[ \text{NDI} = \frac{\rho_1 - \rho_2}{\rho_1 + \rho_2} \]  

(1)

where \( \rho_1 \) and \( \rho_2 \) are the reflectance at two specific bands for the specific ground material. For example, for NDVI, \( \rho_1 \) and \( \rho_2 \) denote the reflectance at the near-infrared (NIR) and red bands, respectively. For NDWI, \( \rho_1 \) and \( \rho_2 \) denote the reflectance at the green and NIR bands, respectively. A simple ratio index corresponding to an NDI is generally defined as:

\[ \text{RI} = \frac{\rho_1}{\rho_2} \]  

(2)
Many applications identify a specific material by applying a threshold to a calculated NDI. For example, if a pixel has an NDVI value larger than a specified threshold, it will be regarded as vegetation. Such simple binary classification methods have been widely used. However, as \( \text{NDI} = (R_I - 1)/(R_I + 1) \), NDI > 0 is functionally equivalent to \( R_I > T' \), where \( T \) is the specified threshold and \( T' = (1 + T)/(1 - T) \). Therefore, for a simple binary classification, there is actually no difference between using \( T \) for NDI and using \( T' = (1 + T)/(1 - T) \) for \( R_I \).

There are two shortcomings of NDIs. First, although the ratio definition can help to reduce the effect of the solar zenith angle and can suppress topographic effects, the physical reflectance magnitude is not taken into consideration. For example, the same NDWI value can be obtained using both \( (\rho_{\text{green}} = 0.045, \rho_{\text{NIR}} = 0.03) \) and \( (\rho_{\text{green}} = 0.75, \rho_{\text{NIR}} = 0.5) \). Water is nevertheless highly absorbent in NIR, and its reflectance in that spectrum is physically small. In spectral libraries, most of the standard water spectra indeed show reflectance values of less than 0.05 in NIR. Second, only two bands are involved in calculating an NDI, disregarding the reflectance at other bands. As a specific ground material generally demonstrates a characteristic reflectance curve, using only two bands cannot sufficiently capture the characteristics of the spectral curve, and this will increase the risk of misidentifying other materials as the material in question. For example, the same NDVI value is obtained using both \( (\rho_{\text{NIR}} = 0.66, \rho_{\text{red}} = 0.25, \rho_{\text{blue}} = 0.26) \) and \( (\rho_{\text{NIR}} = 0.66, \rho_{\text{red}} = 0.25, \rho_{\text{blue}} = 0.54) \). Nevertheless, vegetation is highly absorbent in blue, and thus its reflectance in this spectrum is physically small. Therefore, the latter spectral signature cannot be vegetation even though its NDVI value of 0.45 is larger than a commonly used threshold for vegetation detection. Due to these two reasons, there is a risk that many other materials may be misidentified by applying a threshold only on an NDI, especially in complex heterogeneous environments or large areas.

3. Experiments

In the literature, NDIs have generally been proposed for the Landsat series of satellites. They were derived by observing and analyzing a set of spectra of different ground materials that were collected in a few study areas. Their characteristics have been validated locally. With the development of large spectral libraries in recent years, such as the USGS spectral library [18] and ECOSTRESS [19], standard spectral datasets have become available to re-examine the effects and risks of these indices. In addition, remote sensing cloud platforms such as Google Earth Engine now provide global images to evaluate NDIs at a large spatial scale. In our experiments, we validated the NDI approach first with spectral libraries and then with Landsat-8 and Sentinel-2 images, which are two of the most widely used medium spatial resolution multispectral sensors. We validated NDIs for three main land cover classes: vegetation (NDVI), water (NDWI) and soil (NDSI). These NDIs are defined as:

\[
\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}} \tag{3}
\]

\[
\text{NDWI} = \frac{\rho_{\text{green}} - \rho_{\text{NIR}}}{\rho_{\text{green}} + \rho_{\text{NIR}}} \tag{4}
\]

\[
\text{NDSI} = \frac{\rho_{\text{MIR}} - \rho_{\text{NIR}}}{\rho_{\text{MIR}} + \rho_{\text{NIR}}} \tag{5}
\]

The same validation procedure can be performed for variants of these indices and for some other indices that have been proposed for other ground materials. We used the threshold values of 0.3 for NDVI, and 0 for NDWI and NDSI, which are commonly used or suggested in the literature [2,3,9,20,21]. In addition, we used other threshold values ranging from 0 to 0.4 in order to demonstrate changes in performance. Due to space constraints, a selection of our main results is presented in this short technical note, while the other results are provided in the Supplementary Materials.
3.1. Results on Spectral Libraries

We first validated NDVI, NDWI and NDSI using the USGS and the ECOSTRESS spectral libraries. Because NDIs were generally proposed for the Landsat satellites or similar medium resolution multispectral sensors, we used the speclib07 libraries that are resampled to broad-band Landsat-8 and Sentinel-2. We used the central wavelengths of Landsat-8 bands for the ECOSTRESS spectral library and the speclib07 library convolved to ASD standard resolution to compute corresponding narrow-band NDIs. Specifically, the central wavelengths of 480 nm, 560 nm, 655 nm, 865 nm and 1610 nm were used for the blue, green, red, NIR and MIR bands, respectively.

Regarding recall performance, most of the vegetation, water and soil spectra can be successfully identified with the corresponding NDVI, NDWI and NDSI. For example, in the ECOSTRESS library, 537 out of 544 vegetation spectra were successfully identified using NDVI with a threshold value of 0.3, 40 out of 41 soil spectra were successfully identified by NDSI with a threshold value of 0, and all 6 water spectra were successfully identified by NDWI with a threshold value of 0 (Table 1). When the threshold values for an NDI are increased, the recall decreases. For example, when the threshold value was increased to 0.2 for NDSI, only 16 out of 41 soil spectra in the ECOSTRESS library were identified, and only 33 out of 175 soil spectra in the speclib07 Landsat-8 spectral library were identified. In the ECOSTRESS library, none of the water spectra were identified when the threshold value was increased above 0.1, and only 11 out of 22 liquid spectra were identified by NDWI in the speclib07 Landsat-8 spectral library when a threshold value of 0.1 was used. Please refer to Tables S1–S12 for detailed numerical results.

Table 1. Number of spectra identified as specified materials with different normalized difference indices (NDIs) using various threshold values on the ECOSTRESS spectral library.

<table>
<thead>
<tr>
<th>Material</th>
<th>Number</th>
<th>Water (NDWI &gt; 0)</th>
<th>Vegetation (NDVI &gt; 0.3)</th>
<th>Soil (NDSI &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manmade</td>
<td>45</td>
<td>9</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>Mineral</td>
<td>857</td>
<td>243</td>
<td>9</td>
<td>557</td>
</tr>
<tr>
<td>NPV</td>
<td>52</td>
<td>0</td>
<td>26</td>
<td>37</td>
</tr>
<tr>
<td>Rock</td>
<td>380</td>
<td>106</td>
<td>2</td>
<td>252</td>
</tr>
<tr>
<td>Soil</td>
<td>41</td>
<td>0</td>
<td>7</td>
<td>40</td>
</tr>
<tr>
<td>Vegetation</td>
<td>544</td>
<td>0</td>
<td>537</td>
<td>5</td>
</tr>
<tr>
<td>Water</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Recall</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.04</td>
</tr>
<tr>
<td>Precision</td>
<td>0.02</td>
<td>0.92</td>
<td>0.98</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Although the recall (omission error) is generally acceptable for an NDI with an appropriate threshold, we observed that the precision (commission error) is not as high as one generally expects. For example, in the speclib07 Landsat-8 spectral library, 250 (out of 886) mineral spectra, 56 (out of 142) organic spectra and 40 (out of 278) artificial materials were misidentified as water by NDWI (Table 2). In contrast to the general expectation that land cover classes other than vegetation have low NDVI values, there are 49 artificial material spectra, 1 liquid spectrum, 16 mineral spectra, 7 organic spectra and 3 soil spectra that were misidentified by NDVI in the speclib07 Landsat-8 spectral library. These non-vegetation materials have a higher reflectance in the NIR band than in the red band (see Table S3 and Figures S1–S3).
We also observed that there are some materials that were simultaneously identified by more than one NDI. For example, in the ECOSTRESS spectral library, there are 47 rock spectra and 8 manmade spectra that were both identified as water and soil (see Table S13). For the speclib07 Landsat-8 spectral library, there are 183 mineral spectra, 14 artificial material spectra and 15 soil spectra that were identified as water and soil simultaneously (see Table S15). More detailed results are included in the Supplementary Materials.

### 3.2. Results on Spaceborne Remote Sensing Images

In the experiments that we conducted on measured Landsat-8 and Sentinel-2 satellite images, we also observed that there is an obvious commission error for some NDIs. For example, Figure 1a shows an extract of the Sentinel-2 image shown in Figure S8, which covers the Beijing area, China. The blue colored rooftops that can be seen in this area were mostly misidentified as both vegetation and soil using NDVI and NDSI (Figure 1b). However, they are manmade materials and should be classified as impervious surfaces in some urban land cover typologies. Figure 1c shows a high-resolution image of this area on Google Earth. The corresponding Landsat-8 image (Figure 1d) shows that some of these blue rooftop materials were also misidentified as both vegetation and soil. Visually, the number of misidentifications on the Landsat-8 image appears to be less than that on the Sentinel-2 image. As Landsat-8 has a coarser spatial resolution than Sentinel-2, the impact of the mixed pixel phenomenon is higher in Landsat-8 images, and this could lead to a stronger modification of the spectral characteristics of mixed pixels, making them more distinct from pure pixels. More results on Landsat-8 and Sentinel-2 images in other areas globally are included in the Supplementary Materials (see Figures S8–S16).

**Table 2.** Number of spectra that are identified as specified materials with different NDIs using various threshold values on the USGS spectral library (spectra resampled to Landsat-8 bands).

<table>
<thead>
<tr>
<th>Material</th>
<th>Number</th>
<th>Water (NDWI &gt; 0)</th>
<th>Vegetation (NDVI &gt; 0.3)</th>
<th>Soil (NDSI &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial</td>
<td>278</td>
<td>40</td>
<td>49</td>
<td>142</td>
</tr>
<tr>
<td>Coatings</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Liquids</td>
<td>22</td>
<td>15</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Minerals</td>
<td>886</td>
<td>250</td>
<td>16</td>
<td>615</td>
</tr>
<tr>
<td>Organic</td>
<td>142</td>
<td>56</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Soils</td>
<td>175</td>
<td>21</td>
<td>3</td>
<td>118</td>
</tr>
<tr>
<td>Vegetation</td>
<td>285</td>
<td>2</td>
<td>212</td>
<td>123</td>
</tr>
</tbody>
</table>

Recall   | 0.68   | 0.74            | 0.67                    | 0.67           |

Precision| 0.04   | 0.74            | 0.12                    | 0.12           |

![Figure 1. Cont.](image-url)
Figure 1. (a) True-color (red–green–blue) composite of Sentinel-2 image; (b) pixels that were both identified as vegetation and soil in (a); (c) true-color (red–green–blue) composite of Google Earth image; (d) true-color (red–green–blue) composite of Landsat-8 image; (e) pixels that were both identified as vegetation and soil in (d). Note that all the images in this figure are zoomed to be the same size for an easy comparison.

Shadow has a negative effect on NDIs in real-world applications, especially in urban and high mountain areas. For example, Figure 2a shows a Sentinel-2 image of Manhattan, New York. Because there are many dense tall buildings in this Central Business District area, shadows are omnipresent in the scene. Figure 2b shows the pixels that were identified as both soil and water using NDSI and NDWI. After careful visual examination of high-resolution images on Google Earth, soil and water are very rare in these shaded areas, and most of these should be classified as impervious surfaces and vegetation.

Cloud cover is another negative factor for using an NDI in real-world applications. We observed that clouds are often misidentified as liquid water by NDWI. For example, Figure 3a shows a Sentinel-2 scene in the Himalayas. Note that this scene had the least cloud cover among all available Sentinel-2 Level-2A images in this area. Figure 3b shows the pixels that were identified as water. Some clouds and glaciers were misidentified as liquid water by NDWI. Figure S17 shows another scene that covers the same area but with much more cloud cover.

Figure 2. (a) True-color (red–green–blue) composite of Sentinel-2 image; (b) pixels that were both identified as soil and water in (a); (c) true-color (red–green–blue) composite of Google Earth image. Note that all the images in this figure are zoomed to be the same size for an easy comparison.
4. Discussion

From the experimental results on spectral libraries presented above, we can deduce that the omission error is not high when an appropriate threshold value is used for an NDI. In contrast, the risk of commission error is generally high for an NDI. The low omission error (or high recall) is due to the fact that an NDI generally uses the two bands that represent the maximum and minimum reflectance values of a certain land cover type. Among all possible ratio combinations of any two bands, the value of $\rho_{\text{max}} / \rho_{\text{min}}$ is the most sensitive, where $\rho_{\text{max}}$ and $\rho_{\text{min}}$ represent the maximum and minimum reflectance bands. It should be noted that using a higher threshold value can decrease the commission error, but the omission error will increase at the same time.

Putting a threshold on an NDI is equal to an orthographic projection for $n$-dimensional spectral data ($n > 2$ denotes the number of available bands) into a two-dimensional space that is spanned by the two used bands and using a linear decision boundary that passes through the origin to separate the specified material from the other ground materials. For example, Figure 4 shows the scatter plot of all the spectra in USGS spectral library. Non-water and water spectra are plotted as small black dots and large green dots, respectively. It can be seen that the threshold value (NDWI = 0), which corresponds to the slope of the blue line in Figure 4, plays the function of a linear decision boundary. All of the spectra above the line will be regarded as water, and the other spectra below the line will be regarded as non-water. Using a linear decision boundary that passes the origin in a two-dimensional space to separate a material of interest from the other materials could be too strict to avoid commission errors in complex heterogeneous environments. Designing an indicator that uses non-linear decision boundaries is a way to improve performance of an NDI. For example, Figure 5 shows an elliptical decision boundary for the water spectra in USGS spectral library. Compared with the linear decision boundary of NDWI = 0, the water detection precision increases from 0.03 to 0.54 while achieving the identical 100% recall by using this elliptical decision boundary. Figure S18 shows another example of a parabolic decision boundary for these spectra. Taking all available bands into consideration should be considered as another way to improve the commission error. Some researchers have proposed to include a tasseled cap transformation in the design, which uses a sensor-dependent linear combination of all available bands [22]. For large-scale areas, however, a content-dependent transformation such as principal component analysis is not recommended because of regional differences.
Using inequality constraints can help to capture spectral curve characteristics. Chen et al. used several inequality judgment conditions (e.g., $\rho_{\text{red}} > \rho_{\text{SWIR}}$) for water detection and achieved good performance. This approach is more meaningful than on strong reflection bands.

Non-linear decision boundary for all the water and non-water spectra in the USGS spectral library.

Setting a physically meaningful magnitude threshold on specified absorbent bands is another feasible improvement. It can effectively remove other materials. Chen et al. used a reflectance magnitude threshold on the SWIR band for water detection and successfully decreased the commission error [23]. Dozier used two reflectance magnitude thresholds on the Landsat TM1 and TM5 bands in addition to a normalized difference snow index in order to distinguish snow from clouds and shaded areas [24]. In Figure 4, most of the non-water spectra that are above the decision line are successfully removed after setting a threshold value of 0.05 on the SWIR band. Only the spectra that are both above the blue line and on the left of the yellow vertical line will be regarded as water. We believe that similar physically based magnitude threshold values on specific bands could also be derived for other NDIs. For example, in both the USGS and ECOSTRESS spectral libraries, almost all the vegetation spectra have a reflectance magnitude lower than 0.25 in the blue band (see Figures S4 and S5). As vegetation has a characteristic high reflectance in the NIR band, setting a threshold on NIR (e.g., $\rho_{\text{NIR}} > T$) seems intuitive. However, due to the effect of shadows, the range of the reflectance magnitude of vegetation in the NIR band is generally large. For example, the NIR reflectance magnitude of vegetation directly exposed to sunlight could be larger than 0.6 while the reflectance of shaded vegetation could be smaller than 0.05. Therefore, setting a threshold on strong absorption bands is more meaningful than on strong reflection bands.

Using inequality constraints can help to capture spectral curve characteristics. Chen et al. used several inequality judgment conditions (e.g., $\rho_{\text{VIS}} > \rho_{\text{SWIR}}$, which denotes ($\rho_{\text{red}} > \rho_{\text{SWIR}}$)\&($\rho_{\text{green}} > \rho_{\text{SWIR}}$)\&($\rho_{\text{blue}} > \rho_{\text{SWIR}}$)) for water detection and achieved good performance.
good results [23]. Similarly, one could use $\rho_{\text{VIS}} < \rho_{\text{NIR}}$ for vegetation in addition to only putting a threshold value on NDVI. These intuitive inequality conditions can describe the relative magnitude differences between two characteristic bands rather well, and they are parameter-free.

A threshold value of 0 is generally used or suggested for some NDIs, such as NDWI. From the NDI definition of (1), it can be seen that using this value is equal to simply implementing the inequality condition of $\rho_1 < \rho_2$, because the denominator $\rho_1 + \rho_2$ is always positive as it is the sum of two physical reflectance values. Some content-dependent automatic thresholding methods have been proposed in the literature, but some of them, such as Otsu’s method [25], only perform well if the image histogram has a bimodal distribution. As the general aim of using an NDI with a threshold is to quickly detect a specific ground material in a large area, a computationally fast and content-independent threshold method is preferable in many applications.

It is also noteworthy that there is a difference between material detection and quality evaluation. An NDI can be used to evaluate the quality of some ground materials, such as using NDVI to assess canopy characteristics, e.g., leaf area and biomass. For quality evaluation, the presence of the material in question is usually known, and one generally intends to use a feature that is sensitive to the change of reflectance at specific bands in order to indicate the difference from a state that is deemed normal. However, for material detection at large spatial scales, the analyst generally does not know whether the materials of interest are present, or how many there are, or where they are. Therefore, one generally intends to use a feature that is insensitive or robust to the change of reflectance in order to detect all types of materials in question (including both normal and abnormal ones). As we have discussed above, an NDI is generally designed as a sensitive indicator; therefore, the risk of commission error is always present when applying it for detection. Although in this communication we are only concerned with the detection problem, NDI values (without thresholding) can still be used to indicate the quality of the materials involved, to derive biophysical properties or be used as a derived feature for classification in addition to the original spectral reflectance signals.

There are three phenomena that should be taken into consideration when using an NDI, especially in complex heterogeneous environments or large areas. The first one is the mixed pixel phenomenon. Mixed pixels are always present in a scene, especially in medium or low spatial resolution images. In addition, some ground materials could have an intimate mixture with each other, e.g., soils covered by sparse grasses or shrubs. In general, the commission error is more obvious in high spatial resolution images than in medium and low spatial resolution images. Shadow is also a noteworthy phenomenon. It can be cast by tall buildings, tall mountains, trees or clouds. It reduces the difference in reflectance magnitude between different bands and thus reduces the signal to noise ratio of the recorded reflectance signal. It can, therefore, cause many misidentifications [26]. The negative impact of shadows on accuracy is also higher on high spatial resolution images, as shade can be modeled more easily in images with a coarser resolution. Clouds also have a negative impact, and although cloud-free images are always preferred, they are simply unavailable in some areas.

Although we only evaluated three NDIs for three main ground classes (water, soil, vegetation) in our experiments, we believe that the same conclusions can be drawn for other land cover types and for indices that were specifically proposed for other ground materials, such as the normalized difference snow index [24,27] or vegetation index built-up index [28]. Some results on such indices are included in the Supplementary Materials. They also show relatively high commission errors.

5. Conclusions

The aim of this short technical communication was not to criticize the use of NDIs, which have been widely and successfully used in innumerable applications. Rather, its sole purpose was to draw attention to the pitfalls when thresholds are applied to such
indices in order to detect ground materials. There is a high risk of misidentifying many land cover types, especially in complex environments or extensive areas. Although the reported error may be small in some specific study areas, a slight over-estimation could be present in some reported results in the literature. With the advent of remote sensing cloud platforms, fast large-scale monitoring (even at continental and global scale) using simple and efficient algorithms is increasingly being preferred. We believe, however, that the risk of over-estimation should not be rashly ignored. As the remote sensing community has become larger and more multidisciplinary, we hope this paper can act as a warning for those new researchers who might be planning on using NDIs in their studies.

**Supplementary Materials:** The following are available online at https://www.mdpi.com/2072-4292/13/3/450/s1.

**Author Contributions:** Conceptualization, F.C. and T.V.d.V.; methodology, F.C., T.V.d.V. and D.R.; software, F.C. and H.Z.; validation, H.Z.; resources, J.C.; writing—original draft preparation, F.C., T.V.d.V. and D.R.; writing—review and editing, F.C., T.V.d.V. and D.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Key Research and Development Program of China under Grant 2017YFC0821900.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


