Quantifying the Impact of $\text{NDVI}_{\text{soil}}$ Determination Methods and $\text{NDVI}_{\text{soil}}$ Variability on the Estimation of Fractional Vegetation Cover in Northeast China

Yanling Ding 1, Xingming Zheng 2,3,*, Kai Zhao 2,3, Xiaoping Xin 4 and Huanjun Liu 2

Received: 10 November 2015; Accepted: 25 December 2015; Published: 4 January 2016
Academic Editors: Sangram Ganguly, Compton Tucker, Parth Sarathi Roy and Prasad S. Thenkabail

Abstract: Fractional vegetation cover (FVC) is one of the most critical parameters in monitoring vegetation status. Accurate estimates of FVC are crucial to the use in land surface models. The dimidiate pixel model is the most widely used method for retrieval of FVC. The normalized difference vegetation index (NDVI) of bare soil endmember ($\text{NDVI}_{\text{soil}}$) is usually assumed to be invariant without taking into account the spatial variability of soil backgrounds. Two $\text{NDVI}_{\text{soil}}$ determining methods were compared for estimating FVC. The first method used an invariant $\text{NDVI}_{\text{soil}}$ for the Northeast China. The second method used the historical minimum NDVI along with information on soil types to estimate $\text{NDVI}_{\text{soil}}$ for each soil type. We quantified the influence of variations of $\text{NDVI}_{\text{soil}}$ derived from the second method on FVC estimation for each soil type and compared the differences in FVC estimated by these two methods. Analysis shows that the uncertainty in FVC estimation introduced by $\text{NDVI}_{\text{soil}}$ variability can exceed 0.1 (root mean square error—RMSE), with the largest errors occurring in vegetation types with low NDVI. $\text{NDVI}_{\text{soil}}$ with higher variation causes greater uncertainty on FVC. The difference between the two versions of FVC in Northeast China, is about 0.07 with an RMSE of 0.07. Validation using fine-resolution FVC reference maps shows that the second approach yields better estimates of FVC than using an invariant $\text{NDVI}_{\text{soil}}$ value. The accuracy of FVC estimates is improved from 0.1 to 0.07 (RMSE), on average, in the croplands and from 0.04 to 0.03 in the grasslands. Soil backgrounds have impacts not only on $\text{NDVI}_{\text{soil}}$ but also on other $\text{VI}_{\text{soil}}$. Further focus will be the selection of optimal vegetation indices and the modeling of the relationships between $\text{VI}_{\text{soil}}$ and soil properties for predicting $\text{VI}_{\text{soil}}$.

Keywords: fractional vegetation cover; $\text{NDVI}_{\text{soil}}$; dimidiate pixel model; HWSD; soil background

1. Introduction

The fractional vegetation cover (FVC) is the percentage of the vertical projected area of vegetation (including leaves, stems and branches) within a total area, and is an important quantitative parameter for evaluating and monitoring vegetation variation. FVC, which represents the horizontal density of live vegetation [1], first introduced by Deardorff [2], is a crucial biophysical parameter in specific models of numerical weather prediction, regional and global climate modeling and global change monitoring [3,4]. A variation in FVC by only 0.2 in a land surface model can cause a change of as much
as 100 W m⁻² in latent heat flux, while all other factors remain equal [5]. Accurate estimation of FVC is required for efficiently parameterizing global models.

Previous studies have shown that there are three basic approaches for retrieving FVC from remote sensing data: regression model, vegetation index method and pixel unmixing model [6,7]. The linear unmixing model [1] is the most widely used approach to estimate FVC, due to its ease of implementation [8–12]. Among different linear pixel unmixing models, the most widely used model is the dimidiate pixel model. The model assumes that each pixel can be decomposed into a linear combination of bare soil and full vegetation, and only contains three parameters: the normalized difference vegetation index (NDVI) value of pixels in the image, the NDVI value of bare soil (NDVI_soil), and the NDVI value of vegetation with infinite leaf area index (LAI) (NDVI_veg). However, selecting representative NDVI_soil and/or NDVI_veg can be challenging, due to variations in soil composition, grain size, and moisture content. Those cause the spectral characteristics of soil to vary and differences in vegetation species, leaf water content, etc. causing the spectral signals of vegetation to vary [13,14].

Many studies have performed substantial efforts related to the determination of NDVI_soil [1,8,9,11]. Gutman and Ignatov, for example, used the historical minimum value of a desert as NDVI_soil value, 0.04, to derive global FVC [1]. Zeng et al. [11] combined the International Geosphere-Biosphere Program (IGBP) land cover classification with 1 km NOAA AVHRR NDVI data, and employed the fifth percentile of the histogram of the maximum NDVI for the barren or sparsely vegetated category as the NDVI_soil, which was 0.05, to calculate global FVC. However, statistics from 2906 soil spectral reflectances revealed that the mean value of NDVI_soil is significantly larger, at 0.2, and is highly variable, with a standard deviation of 0.08 [5]. Therefore, it is inadequate to use an invariant value of 0.05 or 0.04, without considering the spatial variability of NDVI_soil [5,8]. Besides, the underestimation of NDVI_soil yields overestimations of FVC. However, many researchers opt to use popular published NDVI_soil values of 0.05 or 0.04 because of the extremely difficult task of specifying NDVI_soil at global or regional scales [15–18].

The specification of NDVI_soil requires information on the spatial distribution of soil reflectance. However, measurements of soil reflectance at global or regional scales are unrealistic. Montandon and Small [5] used the NDVI_soil method described in Zeng et al. [11] and 2906 reflectance spectra of soil collected from different datasets to calculate the statistically most likely FVC. Soil types with different properties determine the key physical and chemical properties of soils [19]. Soil type may be a useful tool in constraining variations of soil reflectances [9]. The Harmonized World Soil Database (HWSD) Version 1.2.1 provides global soil types with a spatial resolution of 1 km [20], providing a potential for improving FVC estimation by constraining NDVI_soil values for different soil types. Wu et al. [9] combined the HWSD and annual minimum NDVI to calculate NDVI_soil for different soil types and then estimated global FVC. The variations in soil reflectance also depend on soil properties, such as soil organic matter, texture, clay mineralogy, color and water content [13,21]. Using soil types to determine NDVI_soil takes into account the impact of soil types on NDVI_soil values but the impacts of NDVI_soil variation for a specific soil type on FVC estimation have yet to be investigated.

Two NDVI_soil determining methods are used in this paper. The first method was proposed by Zeng et al. [11] without considering the variations of NDVI_soil. The second method was described in Wu et al. [9] that used the HWSD to derive NDVI_soil for each soil type. This paper aims to investigate the influence of variations of NDVI_soil derived from the second method on FVC estimation by considering the possible values of NDVI_soil for each specific soil type. This paper also assesses the differences in FVC calculations using these two NDVI_soil approaches. Combined NDVI_soil approaches and 1 km Systeme Probatoire d’Observation dela Terre-VEGETATION (SPOT-VGT), two versions of FVC over Northeast China in 2013 are estimated and compared. The accuracies of the two FVC estimations are validated against fine-resolution FVC maps over cropland and meadow steppe. The impact of soil backgrounds on FVC estimation by using the dimidiate pixel model is discussed.
2. Data and Methods

2.1. Study Area

The study area is located in Northeast China, with approximate coordinates of 38°42′N–53°35′N and 115°32′E–135°09′E (Figure 1). The geography covers Heilongjiang, Jilin and Liaoning Provinces, as well as eastern parts of the Inner Mongolia Autonomous Region. Vegetation in this region varies from temperate evergreen conifer-deciduous broad leaf mixed forests, to deciduous broadleaf forests, to woods and shrubs in the east and the north, to typical steppes in the west, with agricultural fields in the middle [22].

Both the IGBP land cover map and the HWSD were used to retrieve NDVI\textsubscript{veg} and NDVI\textsubscript{soil}. The NDVI dataset used in this study was from the SPOT-VGT sensor, for the period of 2013 [23]. The SPOT-VGT NDVI dataset was derived at a spatial resolution of 1 km at a 10-day interval. Daily 1 km NDVI for the two sampling periods below were calculated from the SPOT-VGT 1 km surface reflectance, and were corrected for atmospheric effects [24].

The sample areas (49°20′54.35″–49°21′48.83″N, 120°06′29.88″–120°07′29.42″E, 44°06′51.31″–44°07′56.94″N, and 125°21′17.68″–125°22′48.21″E.) were located in the croplands of Dehui County, Jilin Province and Hulun Buir province and Hulun Buir meadow steppe of the Inner Mongolia Autonomous Region, respectively (Figure 1). These two sites each had areas of 2 km × 2 km, corresponding to four SPOT-VGT 1 km × 1 km pixels.

![Figure 1](image)

**Figure 1.** The sampling areas of in situ fractional vegetation covers (a) and the soil types (b) in Northeast China.

2.2. Data

2.2.1. Collection of Soil Spectral Reflectances

Five-hundred-sixty-four reflectance spectra of soils were collected in Northeast China in 2002, 2005 and 2014 (Figure 2a). The corresponding soil types included Luvisols, Pheaeozems, Chernozems, Gleysols and so on. For each soil reflectance spectral, NDVI value was calculated using the red and NIR bands, convolved by the SPOT-VGT spectral response function. Statistics show that the histogram of NDVI\textsubscript{soil} follows a normal distribution, with a mean value of 0.15 and a standard deviation of 0.04. So, the range of NDVI\textsubscript{soil} is 0.07–0.22 at the \( p = 0.05 \) level (Figure 2b). The prior hypothesis is that the NDVI value in the annual minimum NDVI image, which ranges between 0.07 and 0.22, is considered as the soil area.
2.2.2. Estimation of Fine-Resolution FVC Maps

Two Landsat 8 Operational Land Imager (OLI) 30 m scene images of Dehui site, without clouds, were acquired on 17 June and 12 July 2013. The FVC field measurements were carried out on 19 June and 13 July 2013. A scene HJ1B-Charge-Coupled Device (CCD) 30 m image of Hulun Buir site, without clouds, was acquired on 11 August 2013, and ground sampling was carried out on 3 August (Table 1). All the images were atmospherically corrected using FLAASH program embedded in the ENVI 4.8 software. FLAASH incorporates the MODTRAN4 radiation transfer code to correct atmospheric effect in visible and near-infrared bands up to 3 nm [25].

The HJ-1B satellite is one of a new generation of Chinese small civilian Earth-observing satellites that were launched on 6 September 2008. The satellite carries two CCD cameras with a swath width of 700 km and a 480-h return period. The HJ-1B CCD satellite is widely used in eco-environmental monitoring. The wide-coverage multispectral CCD camera has four bands of blue, green, red and near infrared (NIR) spectral wavelengths (blue: 0.43–0.52 um, green: 0.52–0.60 um, red: 0.63–0.69 um, NIR: 0.76–0.90 um) with a spatial resolution of 30 m [26]. The image used in this study was provided by the China Center for Resources Satellite Data and Application (CRESDA).

<table>
<thead>
<tr>
<th>In-Situ Area</th>
<th>Image Data</th>
<th>Land Cover</th>
<th>Image Date</th>
<th>Ground Sampling Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dehui</td>
<td>Landsat 8 OLI</td>
<td>Croplands</td>
<td>17 June 2013, 12 July 2013</td>
<td>19 June 2013, 13 July 2013</td>
</tr>
<tr>
<td>Hulun Buir</td>
<td>HJ-1B CCD</td>
<td>Steppe</td>
<td>11 August 2013</td>
<td>3 August 2013</td>
</tr>
</tbody>
</table>

The sampling strategy is known as two-stage sampling designed by the Validation of Land European Remote Sensing Instruments (VALERI) project to collect ground parameters [27]. The 2 km × 2 km areas of Dehui and Hulun Buir were divided into four 1 km × 1 km pixels. In each pixel, three 30 m × 30 m elementary sampling units (ESUs) were considered, each equal in size to the spatial resolution of the OLI and CCD images. Subsequently, we selected two-to-five sample plots to measure FVC.

To estimate ground FVC, a Canon 60D (18–135 mm) digital camera was used. Digital images were acquired at each plot at the same height. We used the modified excess green index suggested by Tang et al. [28] to estimate FVC, employing a threshold of the difference between green, red and blue...
colors to distinguish vegetation against soil backgrounds and residues. Visual interpretation procedure in ArcGIS was used to test the method. The precision can reach 99% [28,29].

Wang et al. [30] suggested that empirical regression is qualified to produce fine-resolution maps because regression could provide the best fit between the regressed variables. Empirical regressions were built between in situ FVC and different vegetation indices. Soil adjusted vegetation index (SAVI) and modified soil adjusted vegetation index (MSAVI) provided the best fit with in situ FVC with the highest $R^2$ and the lowest root mean square error (RMSE). The 30 m FVC reference maps for both sites were directly linked to the regressions (Figure 3). 1 km SPOT-VGT FVC estimations will be validated using these reference maps.

![Figure 3. Regressions between in situ FVC and SAVI (a) and MSAVI (b).](image)

2.2.3. SPOT-VGT Data and Processing

SPOT-VGT NDVI subsets derived from the VGT-S products are 10-day composites at 1 km spatial resolution. The VGT-S products are generated from the VGT-P products, atmospherically corrected by a modified version of the Simple Method for the Atmospheric Correction (SMAC) code [31]. The Maximum Value Composite (MVC) technique is used in the construction of the 10-day synthesis [32]. The procedures of cloud clearing, atmospheric correction, and bi-directional composition substantially reduce the noise in reflectance and NDVI [24]. The maximum, minimum and mean NDVI for each pixel, over a one-year period, were computed and used to calculate NDVI$_{veg}$, NDVI$_{soil}$ and FVC. Daily 1 km NDVI images for the three different sampling periods were derived from the 1 km VGT-P products, through atmospheric correction.

2.3. Theory

2.3.1. NDVI$_{soil}$ Determining Methods

The commonly used dimidiate pixel model for retrieval of FVC from NDVI is described in Gutman and Ignatov [1], as follows:

$$f = \frac{NDVI - NDVI_{soil}}{NDVI_{veg} - NDVI_{soil}}$$

(1)

where $f$ is the fractional vegetation coverage of the pixel, NDVI is the NDVI of the pixel, NDVI$_{soil}$ is the NDVI of the bare soil endmember, and NDVI$_{veg}$ is the NDVI of the vegetation endmember.

We used the method described in Zeng et al. [11] to compute the NDVI$_{veg}$ for each land cover type, as defined by IGBP. First, we computed the maximum NDVI over the period of 2013 in Northeast China. Second, histograms of the maximum NDVI values for each IGBP land cover type were plotted. NDVI$_{veg}$ was selected as the 90th percentile for category 16 (barren or sparsely vegetated), and the 75th percentile was selected for all other land cover types. The determined NDVI$_{veg}$ values are given in Table 2.
NDVI\text{soil} was computed using two approaches: (1) by utilizing the 5th percentile of barren and sparsely vegetated land, as described in Zeng et al. [11], which was 0.085 in Northeast China; (2) and by the method described in Wu et al. [9], which used the HWSD and the annual minimum NDVI for the period of 2013 to determine the NDVI\text{soil} for each soil type. The NDVI\text{soil} for each group was defined as the average of the minimum NDVI values which ranged between 0.07 and 0.22, for each soil group area. The NDVI\text{soil} value determined for each soil type is given in Table 2. Different soil types have different NDVI\text{soil}. Combined NDVI\text{veg} and NDVI\text{soil}, two versions of FVC in Northeast China for the period of 2013 were calculated.

Table 2. Values of normalized difference vegetation index (NDVI)\text{veg} and NDVI\text{soil} used in this study.

<table>
<thead>
<tr>
<th>The HWSD Soil Type</th>
<th>NDVI\text{soil}</th>
<th>IGBP Land Cover Type</th>
<th>NDVI\text{veg}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Anthrosols</td>
<td>0.16 ± 0.03</td>
<td>1. Evergreen Needleleaf Forest</td>
<td>0.77</td>
</tr>
<tr>
<td>2. Arenosols</td>
<td>0.16 ± 0.03</td>
<td>3. Deciduous Needleleaf Forest</td>
<td>0.75</td>
</tr>
<tr>
<td>3. Calcisols</td>
<td>0.12 ± 0.04</td>
<td>4. Deciduous Broadleaf Forest</td>
<td>0.76</td>
</tr>
<tr>
<td>4. Cambisols</td>
<td>0.16 ± 0.03</td>
<td>5. Mixed Forest</td>
<td>0.77</td>
</tr>
<tr>
<td>5. Chernozems</td>
<td>0.16 ± 0.03</td>
<td>6. Closed Shrublands</td>
<td>0.78</td>
</tr>
<tr>
<td>6. Fluvicols</td>
<td>0.17 ± 0.02</td>
<td>7. Open Shrublands</td>
<td>0.80</td>
</tr>
<tr>
<td>7. Gleysols</td>
<td>0.13 ± 0.04</td>
<td>8. Woody Savannas</td>
<td>0.76</td>
</tr>
<tr>
<td>8. Histosols</td>
<td>0.16 ± 0.03</td>
<td>9. Savannas</td>
<td>0.77</td>
</tr>
<tr>
<td>9. Kastanozems</td>
<td>0.14 ± 0.04</td>
<td>10. Grasslands</td>
<td>0.64</td>
</tr>
<tr>
<td>10. Leptosols</td>
<td>0.15 ± 0.04</td>
<td>11. Permanent Wetlands</td>
<td>0.76</td>
</tr>
<tr>
<td>11. Luvisols</td>
<td>0.15 ± 0.04</td>
<td>12. Croplands</td>
<td>0.74</td>
</tr>
<tr>
<td>12. Phaeozems</td>
<td>0.15 ± 0.03</td>
<td>13. Urban and Built-up</td>
<td>0.74</td>
</tr>
<tr>
<td>13. Podzoluvicols</td>
<td>0.15 ± 0.03</td>
<td>14. Cropland/Natural Vegetation</td>
<td>0.66</td>
</tr>
<tr>
<td>14. Regosols</td>
<td>0.17 ± 0.04</td>
<td>15. Barren or Sparsely Vegetated</td>
<td>0.50</td>
</tr>
<tr>
<td>15. Solonchaks</td>
<td>0.15 ± 0.03</td>
<td>16. Permanent Wetlands</td>
<td>0.78</td>
</tr>
<tr>
<td>16. Solonetz</td>
<td>0.15 ± 0.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.3.2. Analysis of the Effect of Uncertainty of NDVI\text{soil} in FVC Calculation

Table 2 shows the mean NDVI\text{soil} values, with large standard deviations for each soil type. In order to investigate the impact of NDVI\text{soil} variability of each soil type on FVC calculation, a possible FVC was calculated using the same NDVI\text{veg} as in Table 2. However, we used the values derived from the HWSD for each soil type to vary NDVI\text{soil} that ranging between 0.07 and 0.22 for each pixel and satisfied the dimidiate pixel model condition of NDVI\text{soil} ≤ NDVI\text{pixel}.

Assuming that there are \( n \) NDVI\text{soil} values for each soil type, this yields \( n \) NDVI\text{soil} for each pixel in each soil group area, and \( n \) FVC values. The FVC as calculated by the mean of the \( n \) FVC is referred to as \( f^* \) [5]:

\[
f^* = \frac{\sum_{i=1}^{n} f_{i}}{n} = \frac{\sum_{i=1}^{n} \left( \frac{NDVI_{pixel} - NDVI_{soil,i}}{NDVI_{veg} - NDVI_{soil,i}} \right)}{n}
\] (2)

Then, the error in FVC estimation due to the variability of NDVI\text{soil} is quantified, hereafter referred to as \( \Delta f^* \). It was the difference between the FVC values, respectively computed from Equations (1) and (2) for each pixel:

\[
\Delta f^* = f^* - f
\] (3)

The uncertainties in FVC estimation due to possible \( n \) NDVI\text{soil} values for each pixel were estimated by computing the standard deviation of the resulting \( n f_i \) values:

\[
\sigma = \sqrt{\frac{\sum (f_i - f)^2}{n}}
\] (4)
In order to understand how the error ($\Delta f^*$) and the uncertainty ($\sigma$) vary throughout Northeast China, we calculated the corresponding $f$, $f^*$, $\Delta f^*$, and $\sigma$ values, using the parameters in Table 2 and Equations (1)–(4).

3. Results

3.1. Influence of $\text{NDVI}_{\text{soil}}$ Variability on FVC Calculation

In order to investigate the impact of inherent variation of $\text{NDVI}_{\text{soil}}$ for each soil type on FVC estimation, we consider the possible $\text{NDVI}_{\text{soil}}$ values for each soil type. The possible $\text{NDVI}_{\text{soil}}$ values used in Equation (2) are in the range of 0.07 to 0.22. The $\text{NDVI}_{\text{soil}}$ values listed in Table 2 are the means of the $n$ $\text{NDVI}_{\text{soil}}$, and $f$ is computed using these mean values. $f^*$ is averaged by $n$ possible $f_i$ for each pixel. $f_i$ is calculated using $\text{NDVI}_{\text{soil},i}$. The difference $\Delta f^*$ between $f^*$ and $f$ is caused by the variability of $\text{NDVI}_{\text{soil}}$. For pixels with $\text{NDVI}_{\text{pixel}} < \text{NDVI}_{\text{soil}}, f$ is set to zero, indicating no vegetation covers.

Figure 4 shows the $\Delta f^*$ for the six main land cover types, with seven different soil backgrounds in Northeast China. Over deciduous broadleaf forests (Figure 4a), the majority of pixel NDVI values are higher than 0.22, and the absolute $\Delta f^*$ decreases as the pixel NDVI increases. Over other land cover types with pixel NDVI ranging from 0 to 0.8, $\Delta f^*$ presents two peak values (Figure 4b–f). $\Delta f^*$ displays similar variations over each land type, based on specific soil type. Vegetation types and the values of NDVI show little influence on $\Delta f^*$, and differences in $\Delta f^*$ exist among the eight soil types.

$\Delta f^*$ reaches its highest value when the pixel NDVI is equal to the mean value of $\text{NDVI}_{\text{soil}}$. The $\Delta f^*$ caused by Fluvvisols reaches 0.06 over grasslands and croplands/natural vegetation, as shown in Figure 4d,f. $f^*$ overestimates $f$ with $\Delta f^* > 0$ over areas with Arensols, Cambisols, Chernozems and Fluvvisols backgrounds. The derived $\text{NDVI}_{\text{soil}}$ values for these soils are much larger than those for Gleysols, and the $\text{NDVI}_{\text{soil}}$ values for Gleysols is much lower than that of others. $f$ overestimates $f^*$ when the pixel NDVI is greater than 0.15, over areas with Gleysols. The $\Delta f^*$ caused by Gleysols could reach $-0.04$ over grasslands and croplands/natural vegetation, as shown in Figure 4d,f, respectively. With the growth of vegetation, the $\Delta f^*$ caused by the $\text{NDVI}_{\text{soil}}$ variability approaches zero. $\text{NDVI}_{\text{soil}}$ shows variation because of differences in soil types. Various $\text{NDVI}_{\text{soil}}$ yields errors on FVC estimation, especially when the pixel NDVI is at a low value.

![Figure 4](image-url)

**Figure 4.** Relationship between $\Delta f^*$ and pixel NDVI over deciduous broadleaf forest (a); mixed forest (b); woody savanna; (c); grasslands (d); croplands (e) and croplands/natural vegetation (f), with different soil backgrounds.
3.2. Uncertainty on FVC Estimation

For the variability of NDVI\textsubscript{soil}, f is calculated using the mean value of nd NDVI\textsubscript{soil}, accompanied by an uncertainty σ, as defined in Equation (4). The uncertainties σ in FVC estimation over the six land cover types with different soil backgrounds are shown in Figure 5. Results illustrate that the σ displays similar variations over the six land types. The σ for the seven soil backgrounds also shows similar variations, with differences in magnitude. It increases with the increasing NDVI\textsubscript{pixel} when NDVI\textsubscript{pixel} is about 0.2, then decreased. The σ reaches the maximum value when pixel NDVI is about 0.2. The σ caused by the Luvisols is the greatest, due to its high variance (std = 0.04 in Table 2), and the σ of Fluvic soils is the lowest, due to its low variance (std = 0.02 in Table 2).

![Figure 5](https://example.com/figure5.png)

Figure 5. The uncertainty σ in FVC estimations over deciduous broadleaf forests (a); mixed forests (b); woody savannas (c); grasslands (d); croplands (e) and croplands/natural vegetation (f), with different soil backgrounds.

Statistics of uncertainty σ for the six land cover types with different soil backgrounds are shown in Table 3. For deciduous and mixed forests with the seven soil backgrounds, the means and maximums of σ are lower than those of the other land cover types. The uncertainty of the FVC estimation caused by Luvisols can reach a maximum of 0.076, with a mean of 0.063 over the grasslands. The uncertainty of the FVC calculation of croplands/natural vegetation is also great. Plotting the seasonal cycle for the six land cover types reveals that the peaks of grasslands and croplands/natural vegetation NDVI are lower than those of the other land cover types throughout a typical year. The NDVI values of forests are large, so the uncertainty remains small. The spatial distributions of σ and bias throughout Northeast China are shown in Figure 6c,d, which are influenced by the spatial distribution of NDVI\textsubscript{soil} (Figure 6b). In all, the variability of NDVI\textsubscript{soil} introduces great uncertainty for FVC estimations. NDVI\textsubscript{soil} with higher variation causes greater uncertainty. The accuracy of the results of the dimidiate pixel model will be improved by parameterizing NDVI\textsubscript{soil} instead of using an invariant NDVI\textsubscript{soil}. 

Table 3. For deciduous and mixed forests with the seven soil backgrounds, the means and maximums variation causes greater uncertainty. The accuracy of the results of the dimidiate pixel model will be variations, with differences in magnitude. It increases with the increasing NDVI

2016 improved by parameterizing NDVI

the variability of NDVI

shown in Figure 6c,d, which are influenced by the spatial distribution of NDVI

the uncertainty remains small. The spatial distributions of six land cover types reveals that the peaks of grasslands and croplands/natural vegetation NDVI are

of the FVC calculation of croplands/natural vegetation is also great. Plotting the seasonal cycle for the six land cover types reveals that the peaks of grasslands and croplands/natural vegetation NDVI are

σ

by an uncertainty

3.2. Uncertainty on FVC Estimation

For the variability of NDVI \textsubscript{soil}, f is calculated using the mean value of \( n \) NDVI \textsubscript{soil}, accompanied by an uncertainty \( \sigma \), as defined in Equation (4). The uncertainties \( \sigma \) in FVC estimation over the six land cover types with different soil backgrounds are shown in Figure 5. Results illustrate that the \( \sigma \) displays similar variations over the six land types. The \( \sigma \) for the seven soil backgrounds also shows similar variations, with differences in magnitude. It increases with the increasing NDVI \textsubscript{pixel} when NDVI \textsubscript{pixel} is about 0.2, then decreased. The \( \sigma \) reaches the maximum value when pixel NDVI is about 0.2. The \( \sigma \) caused by the Luvisols is the greatest, due to its high variance (std = 0.04 in Table 2), and the \( \sigma \) of Fluvic soils is the lowest, due to its low variance (std = 0.02 in Table 2).

![Figure 5](https://example.com/figure5.png)

Figure 5. The uncertainty \( \sigma \) in FVC estimations over deciduous broadleaf forests (a); mixed forests (b); woody savannas (c); grasslands (d); croplands (e) and croplands/natural vegetation (f), with different soil backgrounds.

Statistics of uncertainty \( \sigma \) for the six land cover types with different soil backgrounds are shown in Table 3. For deciduous and mixed forests with the seven soil backgrounds, the means and maximums of \( \sigma \) are lower than those of the other land cover types. The uncertainty of the FVC estimation caused by Luvisols can reach a maximum of 0.076, with a mean of 0.063 over the grasslands. The uncertainty of the FVC calculation of croplands/natural vegetation is also great. Plotting the seasonal cycle for the six land cover types reveals that the peaks of grasslands and croplands/natural vegetation NDVI are lower than other land cover types throughout a typical year. The NDVI values of forests are large, so the uncertainty remains small. The spatial distributions of \( \sigma \) and bias throughout Northeast China are shown in Figure 6c,d, which are influenced by the spatial distribution of NDVI \textsubscript{soil} (Figure 6b). In all, the variability of NDVI \textsubscript{soil} introduces great uncertainty for FVC estimations. NDVI \textsubscript{soil} with higher variation causes greater uncertainty. The accuracy of the results of the dimidiate pixel model will be improved by parameterizing NDVI \textsubscript{soil} instead of using an invariant NDVI \textsubscript{soil}. 

Table 3. For deciduous and mixed forests with the seven soil backgrounds, the means and maximums variation causes greater uncertainty. The accuracy of the results of the dimidiate pixel model will be variations, with differences in magnitude. It increases with the increasing NDVI

2016 improved by parameterizing NDVI

the variability of NDVI

shown in Figure 6c,d, which are influenced by the spatial distribution of NDVI

the uncertainty remains small. The spatial distributions of six land cover types reveals that the peaks of grasslands and croplands/natural vegetation NDVI are

of the FVC calculation of croplands/natural vegetation is also great. Plotting the seasonal cycle for the six land cover types reveals that the peaks of grasslands and croplands/natural vegetation NDVI are

σ

by an uncertainty

3.2. Uncertainty on FVC Estimation

For the variability of NDVI \textsubscript{soil}, f is calculated using the mean value of \( n \) NDVI \textsubscript{soil}, accompanied by an uncertainty \( \sigma \), as defined in Equation (4). The uncertainties \( \sigma \) in FVC estimation over the six land cover types with different soil backgrounds are shown in Figure 5. Results illustrate that the \( \sigma \) displays similar variations over the six land types. The \( \sigma \) for the seven soil backgrounds also shows similar variations, with differences in magnitude. It increases with the increasing NDVI \textsubscript{pixel} when NDVI \textsubscript{pixel} is about 0.2, then decreased. The \( \sigma \) reaches the maximum value when pixel NDVI is about 0.2. The \( \sigma \) caused by the Luvisols is the greatest, due to its high variance (std = 0.04 in Table 2), and the \( \sigma \) of Fluvic soils is the lowest, due to its low variance (std = 0.02 in Table 2).

![Figure 5](https://example.com/figure5.png)

Figure 5. The uncertainty \( \sigma \) in FVC estimations over deciduous broadleaf forests (a); mixed forests (b); woody savannas (c); grasslands (d); croplands (e) and croplands/natural vegetation (f), with different soil backgrounds.

Statistics of uncertainty \( \sigma \) for the six land cover types with different soil backgrounds are shown in Table 3. For deciduous and mixed forests with the seven soil backgrounds, the means and maximums of \( \sigma \) are lower than those of the other land cover types. The uncertainty of the FVC estimation caused by Luvisols can reach a maximum of 0.076, with a mean of 0.063 over the grasslands. The uncertainty of the FVC calculation of croplands/natural vegetation is also great. Plotting the seasonal cycle for the six land cover types reveals that the peaks of grasslands and croplands/natural vegetation NDVI are lower than other land cover types throughout a typical year. The NDVI values of forests are large, so the uncertainty remains small. The spatial distributions of \( \sigma \) and bias throughout Northeast China are shown in Figure 6c,d, which are influenced by the spatial distribution of NDVI \textsubscript{soil} (Figure 6b). In all, the variability of NDVI \textsubscript{soil} introduces great uncertainty for FVC estimations. NDVI \textsubscript{soil} with higher variation causes greater uncertainty. The accuracy of the results of the dimidiate pixel model will be improved by parameterizing NDVI \textsubscript{soil} instead of using an invariant NDVI \textsubscript{soil}. 

Table 3. For deciduous and mixed forests with the seven soil backgrounds, the means and maximums variation causes greater uncertainty. The accuracy of the results of the dimidiate pixel model will be variations, with differences in magnitude. It increases with the increasing NDVI
with high NDVI, the RMSE is about 0.07. Over grasslands and croplands/natural vegetation with low values, two versions of the yearly mean FVC over Northeast China for the period of 2013 are calculated, using the invariant NDVI soil (0.085) and pairs of NDVI soil, which are given in Table 2. Statistics of the root mean square error (RMSE) and bias of the two versions of FVC for the main land cover types are calculated and shown in Table 4. Large discrepancies between two retrievals are observed. Over forest with high NDVI, the RMSE is about 0.07. Over grasslands and croplands/natural vegetation with low NDVI, the RMSE is 0.08. The accuracy of FVC remote sensing products in the campaign of validation is required to be 0.05 [33].

Although the invariant NDVI soil method ignores the variability of NDVI soil, many researchers use this method anyway. In order to compare differences in FVC estimation caused by different NDVI soil values, two versions of the yearly mean FVC over Northeast China for the period of 2013 are calculated, using the invariant NDVI soil (0.085) and pairs of NDVI soil, which are given in Table 2. Statistics of the root mean square error (RMSE) and bias of the two versions of FVC for the main land cover types are calculated and shown in Table 4. Large discrepancies between two retrievals are observed. Over forest with high NDVI, the RMSE is about 0.07. Over grasslands and croplands/natural vegetation with low NDVI, the RMSE is 0.08. The accuracy of FVC remote sensing products in the campaign of validation is required to be 0.05 [33].

The invariant NDVI soil method generates systematically higher FVC estimates than the second approach with biases between 0.06 and 0.08. This is because the NDVI soil mean value computed from the soil database is about two times higher than the invariant NDVI soil, at 0.15 versus 0.085, respectively, confirming that the invariant NDVI soil method tends to underestimate NDVI soil [5]. This yields

### Table 3. Statistics of $\sigma$ over the main land cover types with different soil backgrounds.

<table>
<thead>
<tr>
<th>Std (σ)</th>
<th>Luvisols Mean, Max</th>
<th>Phaeozems Mean, Max</th>
<th>Arenosols Mean, Max</th>
<th>Cambisols Mean, Max</th>
<th>Chernozems Mean, Max</th>
<th>Flubvisols Mean, Max</th>
<th>Gleysols Mean, Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deciduous broadleaf forest</td>
<td>0.043, 0.062</td>
<td>0.039, 0.051</td>
<td>0.033, 0.049</td>
<td>0.035, 0.050</td>
<td>0.036, 0.045</td>
<td>0.025, 0.029</td>
<td>0.039, 0.052</td>
</tr>
<tr>
<td>Mixed forest</td>
<td>0.044, 0.062</td>
<td>0.039, 0.050</td>
<td>0.038, 0.048</td>
<td>0.034, 0.049</td>
<td>0.036, 0.045</td>
<td>0.021, 0.031</td>
<td>0.040, 0.053</td>
</tr>
<tr>
<td>Woody savannas</td>
<td>0.049, 0.062</td>
<td>0.043, 0.051</td>
<td>0.041, 0.049</td>
<td>0.043, 0.050</td>
<td>0.040, 0.046</td>
<td>0.025, 0.031</td>
<td>0.044, 0.054</td>
</tr>
<tr>
<td>Grasslands</td>
<td>0.063, 0.076</td>
<td>0.054, 0.061</td>
<td>0.055, 0.060</td>
<td>0.055, 0.061</td>
<td>0.049, 0.056</td>
<td>0.033, 0.038</td>
<td>0.057, 0.065</td>
</tr>
<tr>
<td>Croplands</td>
<td>0.052, 0.064</td>
<td>0.045, 0.052</td>
<td>0.044, 0.051</td>
<td>0.042, 0.051</td>
<td>0.042, 0.047</td>
<td>0.027, 0.032</td>
<td>0.048, 0.055</td>
</tr>
<tr>
<td>Cropland/Natural vegetation</td>
<td>0.052, 0.073</td>
<td>0.049, 0.059</td>
<td>0.052, 0.057</td>
<td>0.047, 0.059</td>
<td>0.044, 0.054</td>
<td>0.031, 0.037</td>
<td>0.053, 0.063</td>
</tr>
</tbody>
</table>

**Figure 6.** Spatial distributions of yearly mean NDVI (a); the NDVI$_{\text{soil}}$ (b); the bias (c); and the standard deviation (d) throughout the Northeast China.

#### 3.3. Comparison of FVC Derived from Two NDVI$_{\text{soil}}$ Methods

Although the invariant NDVI$_{\text{soil}}$ method ignores the variability of NDVI$_{\text{soil}}$, many researchers use this method anyway. In order to compare differences in FVC estimation caused by different NDVI$_{\text{soil}}$ values, two versions of the yearly mean FVC over Northeast China for the period of 2013 are calculated, using the invariant NDVI$_{\text{soil}}$ (0.085) and pairs of NDVI$_{\text{soil}}$, which are given in Table 2. Statistics of the root mean square error (RMSE) and bias of the two versions of FVC for the main land cover types are calculated and shown in Table 4. Large discrepancies between two retrievals are observed. Over forest with high NDVI, the RMSE is about 0.07. Over grasslands and croplands/natural vegetation with low NDVI, the RMSE is 0.08. The accuracy of FVC remote sensing products in the campaign of validation is required to be 0.05 [33].

The invariant NDVI$_{\text{soil}}$ method generates systematically higher FVC estimates than the second approach with biases between 0.06 and 0.08. This is because the NDVI$_{\text{soil}}$ mean value computed from the soil database is about two times higher than the invariant NDVI$_{\text{soil}}$, at 0.15 versus 0.085, respectively, confirming that the invariant NDVI$_{\text{soil}}$ method tends to underestimate NDVI$_{\text{soil}}$ [5]. This yields
an overestimation of FVC, because NDVI_{soil} appears in the equation’s numerator and denominator. FVC is an essential parameter in many hydroclimatic applications, due to its important contributions to climate models [12]. Accurate estimation of FVC is crucial for proper application of the models. It is important to accurately retrieve NDVI_{soil} values when using the dimidiate pixel model to estimate FVC. In the next section, the accuracies of these two methods will be validated using fine-resolution FVC reference maps.

Table 4. The root mean square error (RMSE) and bias derived from the two fractional vegetation cover (FVC) versions for the main land cover types in Northeast China.

<table>
<thead>
<tr>
<th></th>
<th>Deciduous Forest</th>
<th>Mixed Forest</th>
<th>Woody Savannas</th>
<th>Grasslands</th>
<th>Croplands</th>
<th>Cropland/Natural Vegetation</th>
<th>Over All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Deciduous Forest</td>
<td>0.43</td>
<td>0.41</td>
<td>0.33</td>
<td>0.27</td>
<td>0.31</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>FVC1</td>
<td>0.37</td>
<td>0.35</td>
<td>0.26</td>
<td>0.19</td>
<td>0.24</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>FVC2</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Note: * FVC1 and FVC2 are respectively derived using the invariant NDVI_{soil} and NDVI_{soil} in Table 2.

3.4. Validation of FVC Estimations Using 30 m FVC Reference Maps

High resolution imagery is a useful tool in scaling a point to a larger area, when using in situ measurements to validate the accuracy of FVC estimated from satellite image [34]. Thirty meter FVC reference maps are derived from Landsat 8 OLI 30 m images and a HJ-1B CCD 30 image, using the empirical functions displayed in Figure 3. The 1 km FVC values are computed using SPOT-VGT 1 km NDVI and three NDVI_{soil} methods below. Then, the FVC calculated by the three NDVI_{soil} methods are compared with the 1 km FVC aggregated from the 30 m reference maps.

NDVI_{soil} are calculated using three approaches: (1) using in situ NDVI_{soil} measurements made at the cropland Dehui sites, which is 0.21; (2) using the invariant NDVI_{soil} method, which equals to 0.085; and (3) using the values computed from the historical minimum NDVI at each pixel, along with the soil database shown in Table 2, which is 0.14 for croplands at Dehui site, and is 0.13 for grasslands at Hulun Buir site. Figure 7 shows how different NDVI_{soil} approaches affect FVC estimates over the croplands and grasslands. The FVC computed using the third method yields much better estimates than the invariant method over croplands and meadow steppe, compared with the aggregated FVC and the FVC derived from the first method.

At Dehui site (Figure 7a,b), the in situ NDVI_{soil} resampled using SPOT-VGT spectral response functions is 0.21. The NDVI_{soil} computed from the third method is 0.14. Both values are much larger than the invariant NDVI_{soil}. The invariant NDVI_{soil} method underestimates NDVI_{soil}, resulting in the largest RMSE which is 0.14 and overestimations at both study areas (Table 5). The in situ NDVI_{soil}
method shows the highest accuracies. However, $\text{NDVI}_{\text{soil}}$ varies significantly between soils, it is unrealistic to measure soil spectral reflectance all over the study area. FVC estimates using the third method are more accurate than FVC estimated by the invariant method. In these three examples, the estimated accuracies (RMSE) are improved from 0.14 to 0.09 and from 0.06 to 0.05 at Dehui site on 17 June 2013 and 13 July 2013, respectively. At Hulun Buir site, the accuracy is improved from 0.04 to 0.03. Using the information on soil types to take into account NDVI$_{\text{soil}}$ variations is an effective approach for improving FVC estimation accuracy.

Table 5. The estimated accuracies of the two versions of FVC validated by aggregated FVC reference maps.

<table>
<thead>
<tr>
<th></th>
<th>Dehui, on 17 June 2013</th>
<th>Dehui, on 13 July 2013</th>
<th>Hulun Buir</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>In situ NDVI$_{\text{soil}}$</td>
<td>0.04</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Invariant NDVI$_{\text{soil}}$</td>
<td>0.14</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>NDVI$_{\text{soil}}$ based on the HWSD</td>
<td>0.09</td>
<td>0.02</td>
<td>0.05</td>
</tr>
</tbody>
</table>

4. Discussion

4.1. Impact of Soil Backgrounds on Canopy Vegetation Indices

The results in Section 3 show that the biases and uncertainties on FVC estimation are greater over areas with low NDVI, such as grasslands and croplands/natural vegetation, than over other land covers with high NDVI. Validation results also show that the FVC estimates are more accurate at high vegetation cover than at relatively low cover, i.e., RMSE = 0.09 and 0.05, respectively, which can be explained by the sensitivity of canopy NDVI to soil backgrounds. Huete et al. [35] found that the sensitivity of vegetation indices to soil backgrounds was greatest in canopies with intermediate levels of vegetation cover. Soil influences on the spectra of incompletely covered canopy are partly due to differences in red and NIR flux transferring the overlying canopy [36]. A significant amount of NIR flux is scattered and transmitted by canopy towards the soil surface. The soil reflects parts of the scattered and transmitted flux back toward the sensor. In red band, red light is mostly absorbed by leaf layers. The soil only reflects the irradiance from the sun and sky through canopy gaps. NDVI, intrinsically calculated from red and NIR reflectance, is sensitive to FVC, but it is also sensitive to soil backgrounds.

Estimations of FVC from NDVI suffer from the interference of soil backgrounds. Therefore, some vegetation indices have been developed to reduce soil background effects, such as soil adjusted vegetation index (SAVI) [37], modified soil adjusted vegetation index (MSAVI) [38], enhanced vegetation index (EVI) [39] and so on, which have been widely used to monitor vegetation. Compared to NDVI, EVI was found to be more suitable for monitoring vegetation greenness due to minimal atmospheric and soil effects [40]. However, EVI is limited to sensor systems designed with a blue band as well as the red and near-infrared bands [41]. Gitelson et al. [42] suggested that the two-band enhanced vegetation index without a blue band (EVI2) was accurate in estimating FVC. Other indices have been developed to estimate FVC, such as the scaled difference vegetation index (SDVI) and the visible atmospherically resistant index (VARI) [43,44]. SDVI is a suitable approach for retrieval of FVC over heterogeneous surfaces, and VARI is minimally sensitive to atmospheric effects. All of the above indices are possible alternatives to NDVI in the dimidiate pixel model for estimating FVC. The impact of soil backgrounds on the values of these vegetation indices of bare soil (VI$_{\text{soil}}$) need to be investigated.

4.2. Impact of Soil Backgrounds on VI$_{\text{soil}}$

Above mentioned soil-adjusted vegetation indices considerably reduce soil effects on canopy reflectance. However, calculations from the 564 soil reflectance spectra illustrate that EVI$_{\text{soil}}$ and
Validation results reveal that the second method of using soil type information to estimate NDVI\textsubscript{soil} has great differences compared to different soil types. NDVI\textsubscript{soil} with larger variations introduce greater uncertainties in FVC estimation that can reach 0.08. The largest uncertainties and errors occur in vegetation with low NDVI. Compared to the FVC derived from the second method, the invariant NDVI\textsubscript{soil} method yields overestimation of FVC that is about 0.08 in Northeast China. Validation results reveal that the second method of using soil type information to estimate NDVI\textsubscript{soil} yields better estimates of FVC than using the invariant NDVI\textsubscript{soil} value. The FVC estimate accuracies are improved from an average of 0.1 to 0.07 (RMSE) over the croplands and 0.04 to 0.03 over the grasslands. Soil backgrounds have impacts not only on NDVI\textsubscript{soil} but also on other VI\textsubscript{soil} values. More advances will be acquired by selecting optimal vegetation indices sensitive to FVC and non-sensitive to soil backgrounds and modeling the relationships between VI\textsubscript{soil} and soil properties. Soil type is an effective method for considering NDVI\textsubscript{soil} variability.

5. Conclusions

Two NDVI\textsubscript{soil} determining methods were used to estimate FVC in Northeast China. The first method used an invariant NDVI\textsubscript{soil} to estimate FVC and the second method considered the variations of NDVI\textsubscript{soil} by using soil type information to derive NDVI\textsubscript{soil} for each soil type. The accuracies of FVC derived from these two methods were compared and the impact of the variations of NDVI\textsubscript{soil} derived from the second method on FVC estimations was quantified. Analysis shows that NDVI\textsubscript{soil} has great differences compared to different soil types. NDVI\textsubscript{soil} with larger variations introduces greater uncertainties in FVC estimation that can reach 0.08. The largest uncertainties and errors occur in vegetation with low NDVI. Compared to the FVC derived from the second method, the invariant NDVI\textsubscript{soil} method yields overestimation of FVC that is about 0.08 in Northeast China. Validation results reveal that the second method of using soil type information to estimate NDVI\textsubscript{soil} yields better estimates of FVC than using the invariant NDVI\textsubscript{soil} value. The FVC estimate accuracies are improved from an average of 0.1 to 0.07 (RMSE) over the croplands and 0.04 to 0.03 over the grasslands. Soil backgrounds have impacts not only on NDVI\textsubscript{soil} but also on other VI\textsubscript{soil}. More advances will be acquired by selecting optimal vegetation indices sensitive to FVC and non-sensitive to soil backgrounds and modeling the relationships between VI\textsubscript{soil} and soil properties. Soil type is an effective method for considering NDVI\textsubscript{soil} variability.

Acknowledgments: This study was supported by the National Natural Science Foundation of China (41501408). The authors sincerely appreciate the use of in situ FVC at the Hulun Buir site provided by the Virtual Laboratory of Hulunber Grassland Remote Sensing. The authors would like to thank the reviewers and the editor for precise and excellent comments that greatly improved this work.
**Author Contributions:** Yanling Ding, Xingming Zheng and Kai Zhao designed the study. Yanling Ding wrote the paper. Xiaoping Xin provided *in situ* FVC collected in Hulun Buir. Huanjun Liu provided part of soil reflectance spectra collected in Northeast China.

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

**References**


28. Tang, L.; Tian, L.; Stewart, B.L. Classification of broadleaf and grass weeds using Gabor wavelets and artificial neural network. Trans. ASAE 2003, 46, 1247–1254. [CrossRef]


47. Stenberg, B. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma* **2010**, *158*, 15–22. [CrossRef]

