https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Open Access Earth System Science Data Discussions

Normalized Difference Vegetation Index Maps of Pure Pixels over China's mainland for Estimation of Fractional Vegetation Cover

Tian Zhao[1,2], Wanjuan Song[3], Xihan Mu[1,2*], Yun Xie[4,5], Donghui Xie[1,2], Guangjian Yan[1,2]

[1]State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[2]Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[3]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
[4]State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[5]College of Arts and Sciences, Beijing Normal University, Zhuhai 519087, China

Correspondence to: Xihan Mu (muxihan@bnu.edu.cn)

Abstract. Fractional Vegetation Cover (FVC) is an important vegetation structure factor for applications in agriculture, forestry, ecology, etc. Due to its simplicity, the normalized difference vegetation index (NDVI)-based mixture model is widely used to estimate FVC from remotely sensed data. However, the accuracy and efficiency of FVC estimation require the precise calculation of two key parameters: the NDVI of fully covered vegetation and bare soil. Despite their importance, these two endmember NDVI values have not yet been produced as large-scale maps. Traditional statistical methods for obtaining endmember NDVI from satellite datasets highly rely on the assumption that a certain amount of pure pixels of vegetation and soil must be present, which is often invalid for many areas. This study generated 30 m resolution maps of endmember NDVI across China's mainland using the MultiVI algorithm incorporating multi-angle remote sensing data. The quality and accuracy of the endmember NDVI maps were evaluated using various validation data, including statistically obtained pure NDVI, soil spectra from a soil library, and field-measured FVC. The NDVI values for bare soil derived from the MultiVI algorithm were consistent with those obtained from the soil spectral library. Additionally, the FVC estimated using the MultiVI-derived endmember NDVI and the VI-based mixture model exhibited reasonable accuracy when compared to the field measurements. The root mean square deviation (RMSD) values for MultiVI FVC were below 0.13 in the Heihe, Hebei, and Three Gorges Reservoir regions of China. Furthermore, the MultiVI FVC outperformed those calculated using the statistical methods. The endmember NDVI maps provide a convenient and reliable source of key parameters for the accurate and rapid estimation of FVC at large scales. The 30 m pure NDVI maps are free access at https://zenodo.org/records/14060222 (Zhao et al., 2024).

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Open Access Earth System Science Data Discussions

1 Introduction

Fractional vegetation cover (FVC) quantitatively characterizes the horizontal density of photosynthetically active vegetation (Gutman and Ignatov, 1997). It is typically defined as the planar proportion of green vegetation to the total surface extent (Deardorff, 1978). The FVC is an essential parameter in climate and hydrologic models as it represents the spatial contribution of vegetation (Hirano et al., 2004; Gutman and Ignatov, 1998; Eriksson et al., 2006; Mölders and Olson, 2004). Accurate and high-resolution FVC products are in high demand for various studies, including climate change analysis, soil erosion assessment, land disturbance evaluation, and crop growth monitoring (Xie et al., 2011; Naqvi et al., 2013; Gan et al., 2014; Li et al., 2014; Zhang et al., 2013; Fernández-Guisuraga et al., 2021).

Remote sensing can rapidly and repeatedly observe the land surface, making estimating FVC on regional or global scales feasible. Over recent decades, tremendous efforts have been made to derive high-quality FVC from remotely sensed imagery. The published approaches for retrieving FVC can generally be summarized as follows: (i) the vegetation index (VI)-based mixture model (Gutman and Ignatov, 1998; Zeng et al., 2000; Wu et al., 2014; Mu et al., 2021; Zhao et al., 2023); (ii) spectral mixture analysis (García-Haro et al., 2005; Dimiceli et al., 2011; Guan et al., 2012); (iii) machine learning (Baret et al., 2007; Baret et al., 2013; Jia et al., 2015); and (iv) physical model (Xiao et al., 2016).

The linear mixture model is the most commonly used spectral unmixing method. It is generally utilized in surface elements evaluation such as vegetation classification, surface disturbance mapping, and evapotranspiration estimation (Li et al., 2018a; Lu et al., 2003; Cochrane and Souza Jr, 1998). When considering only two endmembers (green vegetation and bare soil), linear mixture modelling can be employed to calculate the relative abundance of live vegetation from the mixed VI, known as the VI-based mixture model. It assumes that the VI for a particular pixel originates from a linearly weighted sum of green vegetation and bare soil, with their respective areal proportions as weighting coefficients (Gitelson et al., 2002). The mixed VI (V) of the pixel is linearly decomposed by the two endmembers, *i.e.*, the VI of the fully vegetated (Vv) and bare soil pixel (Vs), to obtain the areal proportion of green vegetation as FVC:

$$FVC = \frac{V - V_S}{V_v - V_S} , \quad (1)$$

Despite being the most commonly used method for deriving FVC (Gao et al., 2020), the VI-based mixture model still requires enhancements in both accuracy and efficiency. A major limitation is the challenge of obtaining accurate values for *Vv* and *Vs* on a large scale. The normalized difference vegetation index (NDVI) is the primary VI to derive FVC due to its strong correlation with vegetation structural parameters (Gutman and Ignatov, 1998). The two endmember NDVI values (hereafter referred to as pure NDVI values) are often assigned a priori, and the main methods for extracting these values, along with other pure VIs are summarized in Table 1.

Table 1 Brief summary of the primary methods for determining the two pure VI values (*Vv* and *Vs* with NDVI as the default VI).

Methods	*Vv*	*Vs*	Reference

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Independent field spectral measurements of pure vegetation and bare soil pixels	0.71	0.16	(Wang and Qi, 2008)
		*The values are of Modified Soil-Adjusted Vegetation Index (MSAVI).	
Visual interpretation to identify pure pixels from high-resolution remotely sensed images	0.78	0.03	(Jiao et al., 2014)
The endmember extraction algorithm, *e.g.*, the pure pixel index (PPI) method	0.941	0.068	(Jia et al., 2017)
The cumulative maximum and minimum, or cumulative percentages of NDVI values derived from remotely sensed datasets within a specific area or time series	***Vv* determination for different land types:** The 90th percentile of the annual maximum NDVI for different land types (shrubland, barren, sparsely vegetated), the 75th percentile (other) ***Vv* values for each land type:** 0.49 (grassland) 0.6 (open shrubland) 0.68 (mixed forest) 0.7 (broadleaf and deciduous forests)	***Vs* determination for all land types:** The 5th percentile of the annual maximum NDVI for barren and sparsely vegetated land areas ***Vs* values for all land types:** 0.05	(Zeng et al., 2000)
	***Vv* determination for different land types:** The 98th percentile of the monthly maximum NDVI over 5 years for different land types ***Vv* values for each land type:** 0.752 (crop, grass, desert, shrub) 0.816 (mixed woodland, forest) 0.824 (broadleaf deciduous)	***Vs* determination for all land types:** The 2nd percentile of the monthly maximum NDVI over 5 years for desert and semi-desert ***Vs* values for all land types:** 0.048	(Oleson et al., 2000)

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Open Access Earth System Science Data Discussions

	***Vv* determination for different biomes:** The 97th percentile of the historical maximum over 20 years for different biomes ***Vv* values for each biome:** 0.52 (arid) 0.72 (seasonal) 0.67 (evergreen)	***Vs* determination for different biomes:** The 3rd percentile of the historical minimum over 20 years for different biomes ***Vs* values for each biome:** 0.03 (arid) 0.04 (seasonal) 0.05 (evergreen)	(Matsui et al., 2005)

Traditional methods for determining pure NDVI values have significant limitations. Collecting *Vv* and *Vs* from ground truth data or high-resolution remotely sensed imagery by extracting pure pixel values is time-consuming and often limited by data availability. These methods become ineffective when data are unavailable or there are no pure pixels in the usable dataset. The statistical method infers the two pure NDVI values from the data within a specific area or time series. It typically assumes that pixels with the lowest NDVI values represent bare soil, while those with the highest values represent pure vegetation (Gao et al., 2020). However, in arid and semiarid regions with few fully vegetated pixels, or evergreen forests with limited bare soil pixels, the statistically extracted endmember values can be significantly inaccurate (Song et al., 2017).

Additionally, statistical methods typically assign a single value of *Vv* or *Vs* for a specific region or land type. However, endmember values can vary significantly from pixel to pixel due to differences in species composition, vegetation health, moisture levels, and other factors (Jensen, 2000). In many studies, a fixed value of *Vs* is adopted for various soil types, as shown in Table 1, which overlooks the spatial variations in soil moisture, texture, mineralogy, organic matters, and other characteristics (Yang and Yang, 2006; Zeng et al., 2000). The NDVI values of soil samples exhibit considerable variation, ranging from 0 to 0.4, with a mean value of 0.2, which is significantly higher than the commonly used *Vs* value (of approximately 0.05) (Montandon and Small, 2008). Notably, the accuracy of FVC estimation is highly sensitive to variations of *Vs*, particularly in sparse-vegetated areas (Asrar et al., 1984; Montandon and Small, 2008). Underestimating *Vs* may lead to an overestimation of FVC, with errors reaching up to 20% in grassland and shrubland regions (Montandon and Small, 2008; Ding et al., 2016). It has been demonstrated that locally derived pure NDVI values provide higher accuracy for FVC estimation compared to using a fixed global *Vs* value (Donohue et al., 2014; Montandon and Small, 2008). Therefore, pixel-wise maps of endmember NDVI are essential for effectively addressing the spatial variability of plant and soil reflectance. However, such products are currently unavailable.

Recent studies have proposed an alternative method that uses multi-angle datasets from the Moderate Resolution Imaging Spectroradiometer (MODIS) to derive pixel-wise *Vv* and *Vs* (MultiVI algorithm) (Mu et al., 2018; Song et al., 2022a). The discrepancies in directional NDVI values observed from multiple viewing geometries imply vegetation structural information and soil characteristics (Chen et al., 2005; Diner et al., 1999; Deering, 1999; Verrelst et al., 2008). The MultiVI algorithm utilizes the variations from two large viewing angles to establish equations for simultaneously retrieving *Vv* and *Vs*, without

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

assuming invariant endmember values within a scene or biome. Feasibility analysis indicated its potential to estimate high-quality, high-resolution FVC products (Song et al., 2022b). The *Vv* and *Vs* derived from the MultiVI algorithm have been used to generate 30-m / 15-day FVC products, which are reported to possess satisfactory accuracy (Zhao et al., 2023). However, the precision of the *Vv* and *Vs* maps still requires evaluation, and strategies to optimize the MultiVI algorithm for large-scale *Vv* and *Vs* mapping are essential. Providing high-quality pure NDVI values for the VI-based mixture model will significantly improve the accuracy and efficiency of FVC estimation.

This study aims to generate and validate 30 m resolution pixel-wise *Vv* and *Vs* maps across China's mainland. These datasets can be flexibly applied to accurately calculate FVC at various resolutions on regional or national scales. The *Vv* and *Vs* maps were derived from MODIS reflectance data using the MultiVI algorithm. Subsequently, the 500 m MODIS *Vv* and *Vs* were downscaled to 30 m resolution based on land cover types (hereafter referred to as MultiVI *Vv* and *Vs*). Traditional statistical methods were employed to extract *Vv* and *Vs* from Landsat data (hereafter referred to as statistical *Vv* and *Vs*). The two sets of pure NDVI values were then compared and analysed. The generated *Vs* maps were validated using soil NDVI values calculated from a soil spectral library. Finally, the FVC values derived from the MultiVI algorithm and the statistical method were validated against field-measured FVC obtained from various experimental sites across China's mainland.

2 Datasets

2.1 Satellite data for *Vv* and *Vs* calculation

2.1.1 Terra/Aqua MODIS BRDF products

The MODIS Bidirectional Reflectance Distribution Function (BRDF) / Albedo Model Parameters product (MCD43A1) and its corresponding quality assessment product (MCD43A2) (https://lpdaac.usgs.gov/products/mcd43a1v006/) are the primary datasets used to derive pure NDVI values using the MultiVI algorithm. These datasets are produced daily using atmospherically corrected, cloud-cleared input data from the Terra and Aqua satellites over a 16-day period at 500 m resolution. The BRDF characterizes surface anisotropic scattering as a function of illumination and viewing angles. The MCD43A1 product contains three sets of model weighting parameters, *i.e.*, the RossThick kernel (volume-scattering kernel), LiSparseR kernel (geometric-optical kernel), and isotropic kernel parameters. These parameters can be used with the semiempirical linear kernel-driven model, known as the semiempirical RossThick-LiSparse Reciprocal (RTLSR) to calculate surface reflectance (SR) for any required viewing and illumination directions (Roujean et al., 1992; Schaaf et al., 2002). All MCD43A1 data obtained in 2014 over China's mainland were used to reconstruct the ground surface reflectance of red and near-infrared (NIR) bands at viewing zenith angles (VZAs) of 55° and 60°. These reflectance values were subsequently used to generate directional NDVI for the MultiVI algorithm. The quality assessment data from MCD43A2 were applied to exclude clouds, snow, water, and low-quality pixels.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

2.1.2 GlobeLand30 datasets

A global land cover dataset with a 30 m resolution, known as GlobeLand30, was used to downscale the 500 m resolution pure NDVI values to 30 m. The GlobeLand30 products are available globally in three versions: 2000, 2010, and 2020, with the 2020 version being adopted for this study (https://doi.org/10.12041/geodata.140236667788805.ver1.db, last accessed 16 September 2024). These products were developed and updated using cloudless or minimally cloudy multispectral images from Landsat, HJ-1, and GF-1 (Chen et al., 2014; Chen et al., 2015). Validation based on over 230,000 points indicated a total accuracy of 85.72% for the 2020 version of GlobeLand30, with a Kappa coefficient of 0.82. The GlobeLand30 products define bare land as having an FVC of lower than 10%, which is stricter than the criteria used by other land cover products. This criterion helps minimize the misclassification of sparse shrubland or grassland as bare land (Liu et al., 2021). The GlobeLand30 product categorizes land cover into ten classes: six vegetation classes (*i.e.*, cultivated land, forest, grassland, shrubland, wetland, and tundra) and four non-vegetation classes (*i.e.*, artificial surfaces, bare land, water bodies, and perennial snow and ice) (Jun et al., 2014). The classes of wetlands, water bodies, and perennial snow and ice were grouped during the downscaling of *Vv* and *Vs*, as their pure NDVI values are generally similar (below 0).

2.2 Validation data

2.2.1 Statistical *Vv* and *Vs*

The Landsat 8 data were used to obtain *Vv* and *Vs* using statistical methods on the Google Earth Engine (GEE) platform. These statistical *Vv* and *Vs* maps were subsequently compared to the pure NDVI values derived from the MultiVI algorithm. The Landsat 8 Collection 2 Surface Reflectance (SR) products with a resolution of 30 m provided atmospherically corrected SR data (https://www.usgs.gov/landsat-missions/landsat-collection-2-level-2-science-products). The time-series Landsat 8 SR images from 2013 to 2015 were analysed to derive statistical *Vv* and *Vs* for the year 2014. Pixels identified as cloud, cloud shadow, water, and snow in the Landsat 8 images were excluded using the corresponding quality assessment data.

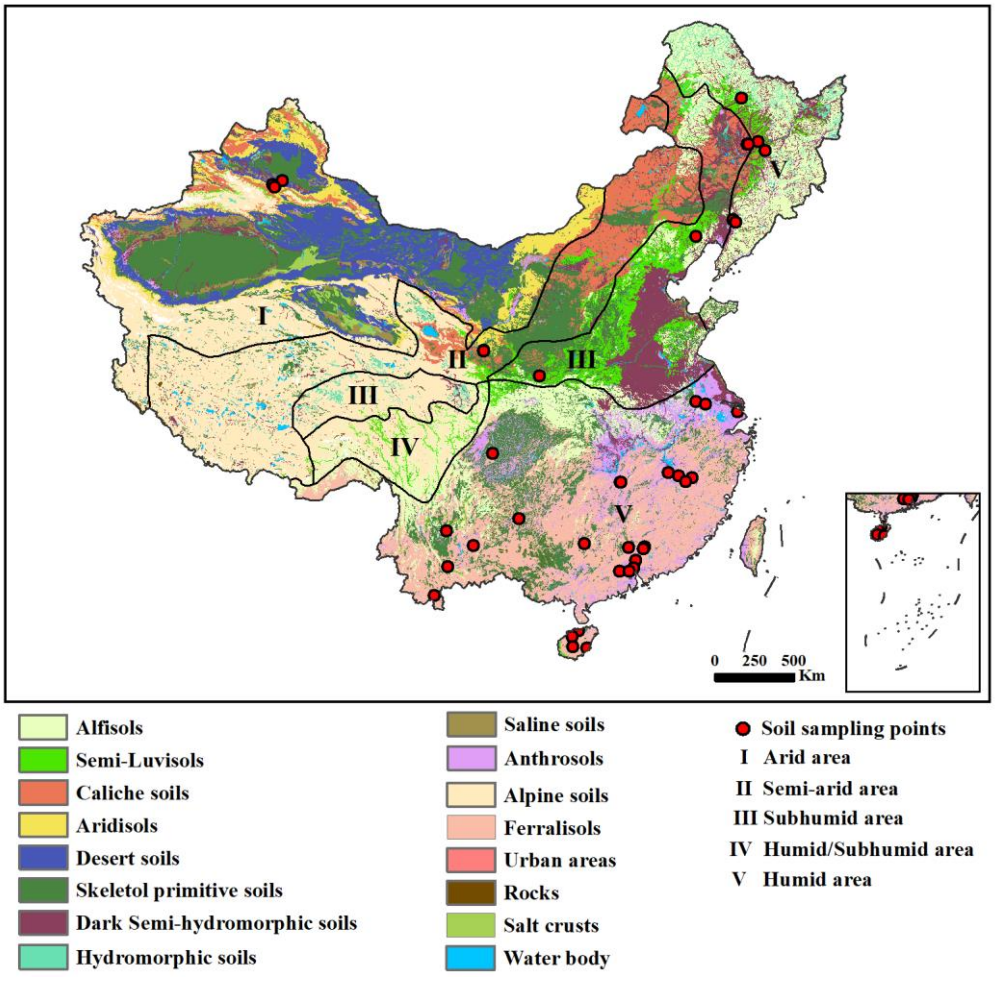
2.2.2 Soil NDVI from the soil spectral library

The soil spectra measured at the Soil and Plant Spectral Diagnostic Laboratory of the World Agroforestry Center (ICRAF) were used to calculate soil NDVI values (https://files.isric.org/public/other/), which were subsequently compared with the retrieved *Vs* values. The soil spectral library that includes soil samples collected from 58 countries was developed by ICRAF in collaboration with the International Soil Reference and Information Centre (ISRIC). This soil spectral library provides laboratory-measured soil spectra, along with attribute data such as geographical coordinates, horizon, and physical and chemical properties. Approximately 20 g of air-dried soil samples were passed through a 2 mm sieve, and placed into 7.4 cm diameter Duran glass Petri dishes for measurement (Garrity and Bindraban, 2006). Spectral measurements were conducted using a FieldSpec FR spectroradiometer (Analytical Spectral Devices, Boulder, CO) at wavelengths ranging from 0.35 to 2.5 μm, with 1 nm intervals (Garrity and Bindraban, 2006).

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Among the soil samples in the ICRAF soil spectral library, 247 were collected in China's mainland (Figure 1). The retrieved *Vs* values were constrained to be no lower than 0, consistent with the findings of most studies (Montandon and Small, 2008; Ding et al., 2016). Consequently, those samples with NDVI values below 0 were excluded, resulting in a total of 228 samples available for validation. These samples were categorized into eight soil types (Table 2) based on their collection locations and the soil type map of China's mainland (https://www.resdc.cn/data.aspx?DATAID=145). Additionally, Figure 1 illustrates the ecological and geographical zones of China's mainland, highlighting the humid and arid areas according to moisture conditions (https://www.resdc.cn/data.aspx?DATAID=125).

The NDVI values of the soil samples in the ICRAF soil spectral library range from 0 to 0.3, with the majority concentrated around 0.1. The mean NDVI value for all soil samples in China's mainland is 0.08, with a standard deviation of 0.05. Desert soils exhibit the lowest NDVI values, with a maximum of 0.0391. This soil type is mainly found in the arid regions of northwestern China (Figure 1). In contrast, Alfisols have the highest NDVI values, reaching a maximum of 0.2732. These soils are primarily distributed in the humid regions of northeastern and southwestern China (Figure 1).



https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Figure 1: The spatial distribution of soil types in China's mainland. The red circle represents the locations of soil samples in the ICRAF soil spectral library. The ecological and geographical zones of China's mainland are numbered using Roman numerals.

Table 2 NDVI of the soil samples collected from the soil spectral library.

Soil type	Number of samples	Mean	Maximum	Minimum	Standard deviation
Alfisols	10	0.1580	0.2732	0.0554	0.0697
Semi-Luvisols	22	0.1091	0.2062	0.0507	0.0392
Caliche soils	12	0.1077	0.2057	0.0438	0.0612
Desert soils	10	0.0282	0.0391	0.0118	0.0113
Skeletol primitive soils	28	0.0811	0.1629	0.0156	0.0468
Dark Semi-hydromorphic soils	9	0.1424	0.2035	0.0567	0.0489
Anthrosols	41	0.0578	0.1098	0.0013	0.0309
Ferralisols	96	0.0635	0.1803	0.0005	0.0371

2.2.3 Field-measured FVC

The field-measured FVC was collected from three sites: (i) the Hebei watershed in humid northeastern China, (ii) the Heihe River Basin in arid northwestern China, and (iii) the Three Gorges Reservoir Area in humid southwestern China. All field measurements were conducted over a complete vegetation growth cycle, with data collected approximately every 15 days. Figure 2 shows the locations of the these sites.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

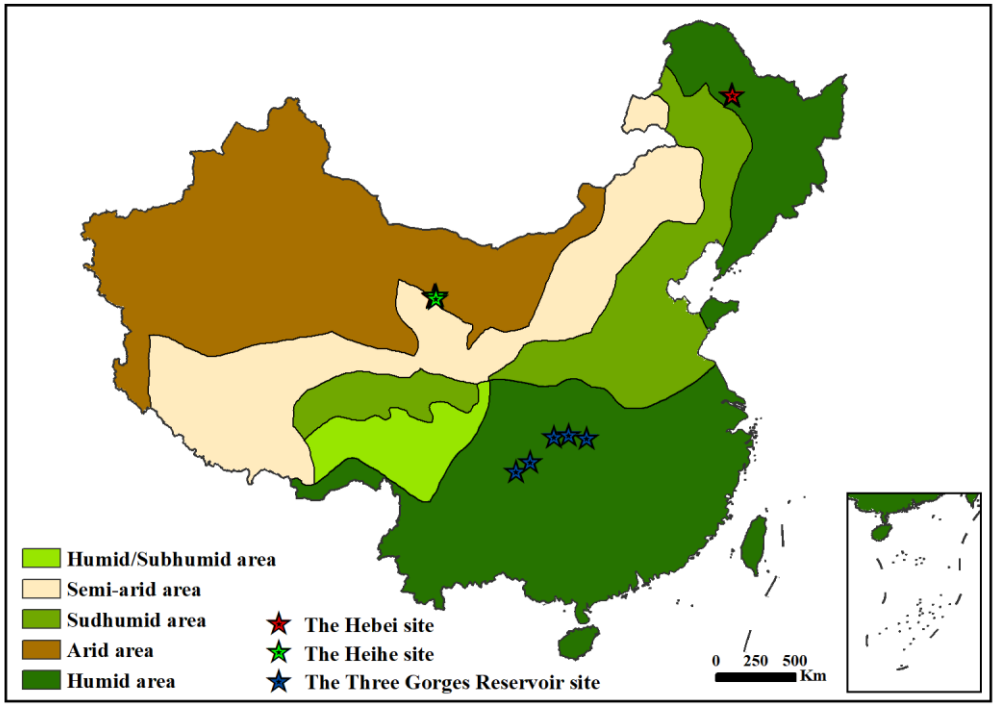


Figure 2: The ecological and geographical zones of China's mainland. The red, green, and blue pentagrams symbolize the Hebei site, the Heihe site and the Three Gorges Reservoir site, respectively.

The Hebei watershed is located in a typical black soil region of China and is mainly covered by crops. Eight sampling plots were distributed in relatively flat and uniform areas, comprising five forest plots, two grassland plots, and one cropland plot. The field measurements were conducted between 15 April and 30 October, 2022. An unmanned aerial vehicle (UAV) was used to capture FVC images of the cropland and grassland, achieving a resolution of higher than 1.5 cm and a plot size of 100 m × 100 m. A digital camera was used to acquire FVC for the forests, with a plot size of 30 m. The camera was mounted on a long pole at a height of 1.5 to 2 meters above the ground and took vertical photographs at regular intervals along two diagonal lines within each sampling plot (Mu et al., 2013; Li et al., 2012). Images were captured from top to bottom and bottom to top at each step to document the coverage of understory vegetation (f_{up}) and overstory canopy (f_{down}), respectively. The FVC was calculated as the weighted sum of the f_{up} and f_{down} using the following Eq. (2):

$$FVC = f_{up} + (1 - f_{up}) * f_{down}\ , \quad (2)$$

The study area in the Heihe River Basin comprises 72% cropland, 24% residential land, and 4% woodland, indicating significant surface heterogeneity (Mu et al., 2015). The sampling plots, each measuring 10 m × 10 m, were exclusively located in cropland areas within the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) sites, where corn was the predominant crop. The FVC measurements were conducted using a digital camera from 15 May to 14 September, 2012, encompassing the entire growing season.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

The Three Gorges Reservoir Area is situated in the mid-upper reaches of the Yangtze River. Five representative small river basins were selected for vegetation monitoring within this region. Seasonal trajectories of vegetation cover were recorded using UAVs every two weeks from 15 August, 2021, to 1 August, 2022. Each small river basin contained a sampling plot of approximately 100 m × 100 m, distributed in a relatively flat and homogeneous area. The primary vegetation types included orchards, forests, and shrubland. The UAV images had a resolution of higher than 1.5 cm.

A Half-Gaussian Fitting algorithm (HAGFVC) was used to calculate FVC from UAV-acquired images, resulting in a minimal mean bias error (MBE) and root mean square error (RMSE) of less than 0.04 (Li et al., 2018b). The digital images captured by hand-held cameras were processed using a shadow-resistant algorithm (SHAR-LABFVC) to extract FVC, achieving an RMSE of approximately 0.025 (Song et al., 2015).

3 Methods

3.1 MultiVI algorithm for retrieving the pure NDVI maps

3.1.1 Theory

The MultiVI algorithm uses multi-angle remotely sensed observations to retrieve *Vv* and *Vs*. It defines the directional vegetation cover, $F(\theta)$, which represents the FVC at the VZA θ. A nonlinear coefficient, *k*, is introduced in the VI-based mixture model. This coefficient helps mitigate the saturation effect of NDVI in dense vegetation and accommodates the slight nonlinear relationship between FVC and NDVI, as shown in Eq. (3) (Xiao and Moody, 2005; Jiapaer et al., 2011; Choudhury et al., 1994):

$$F(\theta) = \left(\frac{V(\theta)-V_S}{V_v-V_S}\right)^k, \quad (3)$$

where $V(\theta)$ is the NDVI observed at VZA θ. The directional gap fraction model can be expressed as Eq. (4) (Nilson, 1971):

$$P(\theta) = e^{-G(\theta)\cdot\Omega(\theta)\cdot LAI/\cos\theta}, \quad (4)$$

Here, $P(\theta)$ denotes the directional gap fraction, $G(\theta)$ is the mean projection of unit foliage area (Goel and Strebel, 1984), $\Omega(\theta)$ is the clumping index, and LAI represents the leaf area index. The directional FVC and gap fraction exhibit a complementary relationship, such that the sum of $F(\theta)$ and $P(\theta)$ equals 1. Therefore, Eq. (3) and (4) can be combined as follows:

$$1-\left(\frac{V(\theta)-V_S}{V_v-V_S}\right)^k = e^{-G(\theta)\cdot\Omega(\theta)\cdot LAI/\cos\theta}, \quad (5)$$

The *G(θ)* in Eq. (5) is often assumed to be constant at large VZAs around 57.5°, despite variations in leaf angle distributions (Leblanc et al., 1999; He et al., 2011; Song et al., 2017; Mu et al., 2018; Weiss et al., 2004; Roujean et al., 1992). Furthermore, the variation of $G(\theta)\cdot\Omega(\theta)$ is significantly smaller than the angular variation of cosθ at large VZAs (Mu et al., 2018). Since the LAI is also independent of VZA, $G(\theta)\cdot\Omega(\theta)\cdot LAI$ can be assumed to be invariant at large VZAs around 57.5°. The angular

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

effects of *Vv* and *Vs* are negligible (Escadafal and Huete, 1991; Mu et al., 2018). By using pairs of observations at large VZAs around 57.5° and eliminating angle-invariant parameters, Eq. (5) can be reorganized as:

$$\left(1-\left(\frac{V(\theta_i)-V_s}{V_v-V_s}\right)^k\right)^{cos\theta_i}=\left(1-\left(\frac{V(\theta_j)-V_s}{V_v-V_s}\right)^k\right)^{cos\theta_j}, \quad (6)$$

where the subscripts "*i*" and "*j*" represent pairs of large VZAs around 57.5°. The combination of 55° and 60° in the forward viewing directions was identified as the optimal angular configuration. This selection is attributed to its minimal influence on $G(\theta)$ and the high quality of angular remote sensing observations (Mu et al., 2018). These angles were used to formulate equations for estimating *Vv* and *Vs*:

$$\left(1-\left(\frac{V(55°)-V_s}{V_v-V_s}\right)^k\right)^{cos55^\circ}=\left(1-\left(\frac{V(60°)-V_s}{V_v-V_s}\right)^k\right)^{cos60^\circ}, \quad (7)$$

The unknown parameters *Vv*, *Vs*, and *k* for a given pixel can be derived using at least three pairs of angular observations at VZAs of 55° and 60° to solve Eq. (7).

3.1.2 Implementation

The observed NDVI used to solve Eq. (7) should exhibit distinct variations to ensure stable results (Mu et al., 2018). This study used observations from different periods to construct appropriate equations for each pixel. The *Vv* and *Vs* were independently retrieved to enhance accuracy by applying high and low NDVI values, respectively. The historical minimum and maximum NDVI values for each pixel in 2014 were used as statistical boundaries. This approach ensured that the derived *Vs* values remained below the historical minimum, while the derived *Vv* values exceeded the historical maximum. Furthermore, empirical boundaries for *Vs* ([0.01, 0.3]) and *Vv* ([0.6, 1.0]) were applied to constrain the retrieval, preventing the upper limit of *Vs* from exceeding 0.3 and the lower limit of *Vv* from falling below 0.6 (Montandon and Small, 2008). This flexible approach integrates statistical boundaries for *Vv* and *Vs*, while allowing for reasonable intra-variability within each land cover type.

Figure 3 illustrates the steps for using the MultiVI algorithm to derive 500 m pure NDVI values from the MODIS BRDF products:

(1) Daily directional NDVI values at VZA of 55° and 60° were calculated for each pixel throughout the year using the semiempirical RTLSR model;

(2) NDVI pairs at VZA of 55° and 60° were ranked in ascending order based on the values at 55° VZA for the entire year;

(3) The annual NDVI value sequence was divided into two groups based on the 10th percentile. The NDVI values below the 10th percentile were used to retrieve *Vs* using Eq. (7), while the remaining 90% were used to retrieve *Vv*;

(4) The 25th, 50th, 75th, and 100th percentiles of the NDVI pairs in each group were selected to construct inversion equations (Eq. (7)) for each pixel. The unknown parameters *Vv*, *Vs*, and *k* were numerically solved using the least squares method;

(5) For a small number of invalid pixels due to limited observations, gap filling was performed based on the MODIS land cover data (MCD12Q1). The mean values of *Vv* and *Vs* corresponding to the same land cover type were used to fill these gaps.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

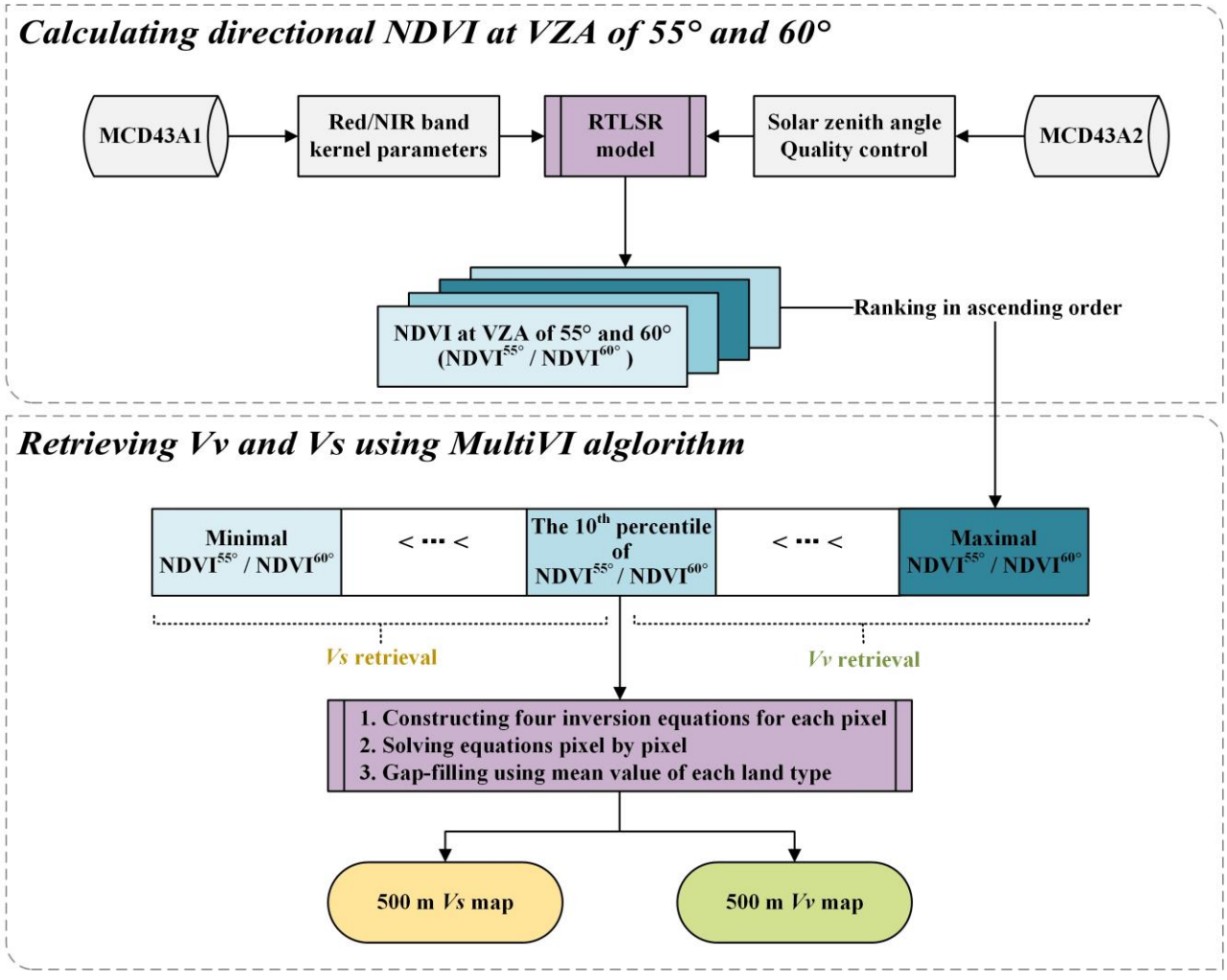


Figure 3: The scheme of the MultiVI algorithm to derive 500 m pure NDVI values from MODIS BRDF products.

3.2 Spectral unmixing for downscaling the 500 m *Vv* and *Vs*

The 500 m *Vv* and *Vs* were downscaled to 30 m resolution using spectral unmixing theory to facilitate fine-scale FVC estimation. It was assumed that 30 m pixels of the same land cover type shared identical *Vv* and *Vs* values. The 500 m *Vv* and *Vs* were considered as linear combinations of the 30 m *Vv* and *Vs* values, with weights determined by the areal proportion of each land cover type, as shown in Eq. (8):

$$\begin{cases} V_{v,500} = \sum_{i=1}^{n} P_i V_{v,30,i} \\ V_{s,500} = \sum_{i=1}^{n} P_i V_{s,30,i} \end{cases}, \quad (8)$$

where $V_{v,500}$ and $V_{s,500}$ represent the *Vv* and *Vs* at 500 m resolution within a MODIS pixel, respectively; P_i signifies the proportion of the *i*th land cover type within the 500 m MODIS pixel, indicating its area-weighted contribution; and $V_{v,30,i}$ and

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

255 $V_{s,30,i}$ denote the *Vv* and *Vs* at 30 m resolution for land cover type *i*, respectively. Here, i (1,2,⋯,n) represents each land cover type.

Figure 4 illustrates the method for downscaling 500 m pure NDVI values to 30 m resolution. The downscaling process utilized seven land cover types from the GlobeLand30 product, specifically four vegetation classes (cultivated land, forest, grassland, shrubland, and tundra), a grouped water surface category (wetland, water body, and permanent snow and ice), bare land, and
260 artificial surfaces. A 3 × 3 sliding window with a step size of one MODIS pixel was employed to construct the unmixing equations (Eq. (8)). The derived $V_{v,30,i}$ and $V_{s,30,i}$ values were then assigned as the *Vv* and *Vs* for all 30 m pixels of land cover type *i* within the central MODIS pixel area.

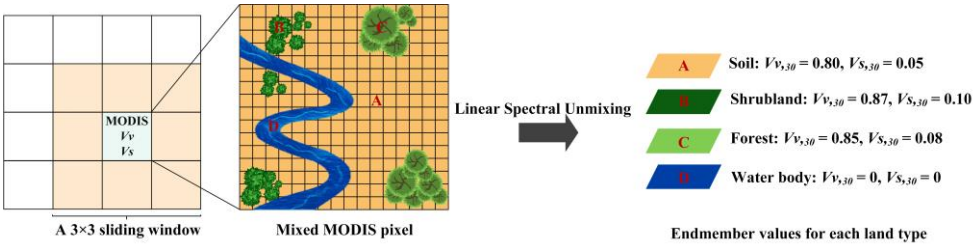


Figure 4: The approach for downscaling the 500 m MultiVI *Vv* and *Vs* to 30 m resolution based on the GlobeLand30 product.

265 **3.3 Assessment and validation**

3.3.1 Comparison with statistical methods

Statistical methods were used to generate 30 m statistical *Vv* and *Vs* in 2014 across China's mainland using Landsat 8 data, which were then compared with the MultiVI *Vv* and *Vs*. The statistical method utilized Landsat 8 data from 2013 to 2015, processed on the Google Earth Engine (GEE) platform. Initially, the pixel-wise maximum and minimum NDVI values over
270 the three years were selected as the preliminary estimates for *Vv* and *Vs*, respectively. These initial values were then smoothed using an 11 × 11 pixel (330 m × 330 m) moving window to reduce spatial heterogeneity. The maximum and minimum NDVI within each moving window were subsequently assigned to the central pixel as its *Vv* and *Vs*, respectively. Finally, the statistical *Vv* and *Vs* were compared with the MultiVI *Vv* and *Vs* to assess differences in spatial patterns and magnitudes.

3.3.2 Comparison with soil NDVI from the soil spectral library

275 The soil spectra obtained from the soil spectral library were convolved to the red and NIR bands using the spectral response functions of MODIS. The convolved soil spectra were then used to calculate the soil NDVI for comparison with the retrieved *Vs*. The MultiVI *Vs* and statistical *Vs* were averaged for each soil type to facilitate comparison with the soil NDVI. Additionally, the uncertainties in the estimated FVC caused by intra-class variability in *Vs* were assessed for each soil type.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

3.3.3 Assessment with field-measured FVC

The field-measured FVC at the Hebei, Heihe, and Three Gorges Reservoir sites was used to assess the estimated FVC derived from the MultiVI *Vv*/*Vs* and that from the statistical *Vv*/*Vs*. The Landsat 8 NDVI time series were smoothed using a harmonic model (Zhao et al., 2023). The NDVI was generated at half-monthly intervals to coincide with the observation times of the field measurements. Subsequently, the NDVI was converted to FVC using the VI-based mixture model and the retrieved *Vv* and *Vs*. The estimated FVC was averaged across a sliding window of surrounding 3 × 3 pixels for comparison with the field-measured FVC (Weiss et al., 2007).

In the Heihe site, where the plot size (10 m) is smaller than the pixel size of Landsat 8 (30 m), field-measured FVC was upscaled to 90 m for comparison. This upscaling process utilized reflectance data obtained from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). An empirical transfer function was employed to establish the relationship between ASTER NDVI and field-measured FVC (Mu et al., 2014). The FVC at ASTER's resolution (15 m) was estimated from ASTER NDVI using the established transfer relationship. The reference FVC was acquired by upscaling the 15 m resolution FVC to 90 m through arithmetical averaging.

The correlation coefficient (R^2) and the root mean square deviation (RMSD) were used to assess the relationship and differences between the field-measured FVC and the estimated FVC, respectively.

4 Results

4.1 Maps of the MultiVI *Vv* and *Vs*

Figure 5 shows the *Vv* and *Vs* maps generated using the MultiVI algorithm (Figures 5a and 5b) and the statistical method (Figures 5c and 5d). The MultiVI *Vv* and *Vs* maps demonstrate smooth distributions, whereas the statistical *Vv* and *Vs* maps exhibit noticeable stripes along the borders of the Landsat 8 imagery tiles. The spatial patterns of pure NDVI values derived from the MultiVI algorithm and the statistical method show similar trends, predominantly influenced by moisture conditions. Specifically, the humid and sub-humid areas of southeastern China are characterized by elevated pure NDVI values, while the arid and semi-arid regions of northwestern China are associated with lower values.

The statistical *Vv* are generally lower than the MultiVI *Vv* in most areas, particularly in semi-arid and arid regions. In northwestern China, which is primarily covered by grasslands, bare lands, and deserts, the statistical method yields NDVI values of less than 0.3 for pure vegetation pixels (Figure 3b). These values are significantly lower than the *Vv* values reported in most studies (Table 1).

Table 3 presents the MultiVI *Vv* and *Vs* values across various land types. The mean *Vv* values range from 0.82 to 0.91, while the mean *Vs* values range from 0.07 to 0.29. Both MultiVI *Vv* and *Vs* exhibit consistent patterns across different land types, indicating that vegetation with higher *Vv* values tends to also exhibit higher *Vs* values. The shrublands show the lowest values, whereas evergreen needleleaf forests exhibit the highest. The forests generally have higher values than other land types, with

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

evergreen forests surpassing broadleaf forests. In contrast, grasslands and shrublands have lower values compared to other land types. The standard deviation values for *Vs* are greater than those for *Vv*.

The spatial patterns of the MultiVI *Vs* exhibit distinct variations across different soil types and demonstrate a closer alignment with the actual soil distribution in China's mainland (Figure 1) when compared to the statistical *Vs*. The MultiVI *Vs* are generally lower than the statistical *Vs*, especially in the densely vegetated areas of southeastern China. In these humid regions, the statistical *Vs* values exceed 0.4, where evergreen vegetation predominates and shows few bare lands (Figure 3d). These values, which exceed 0.4, are notably higher than the generally accepted soil NDVI values (Table 1).

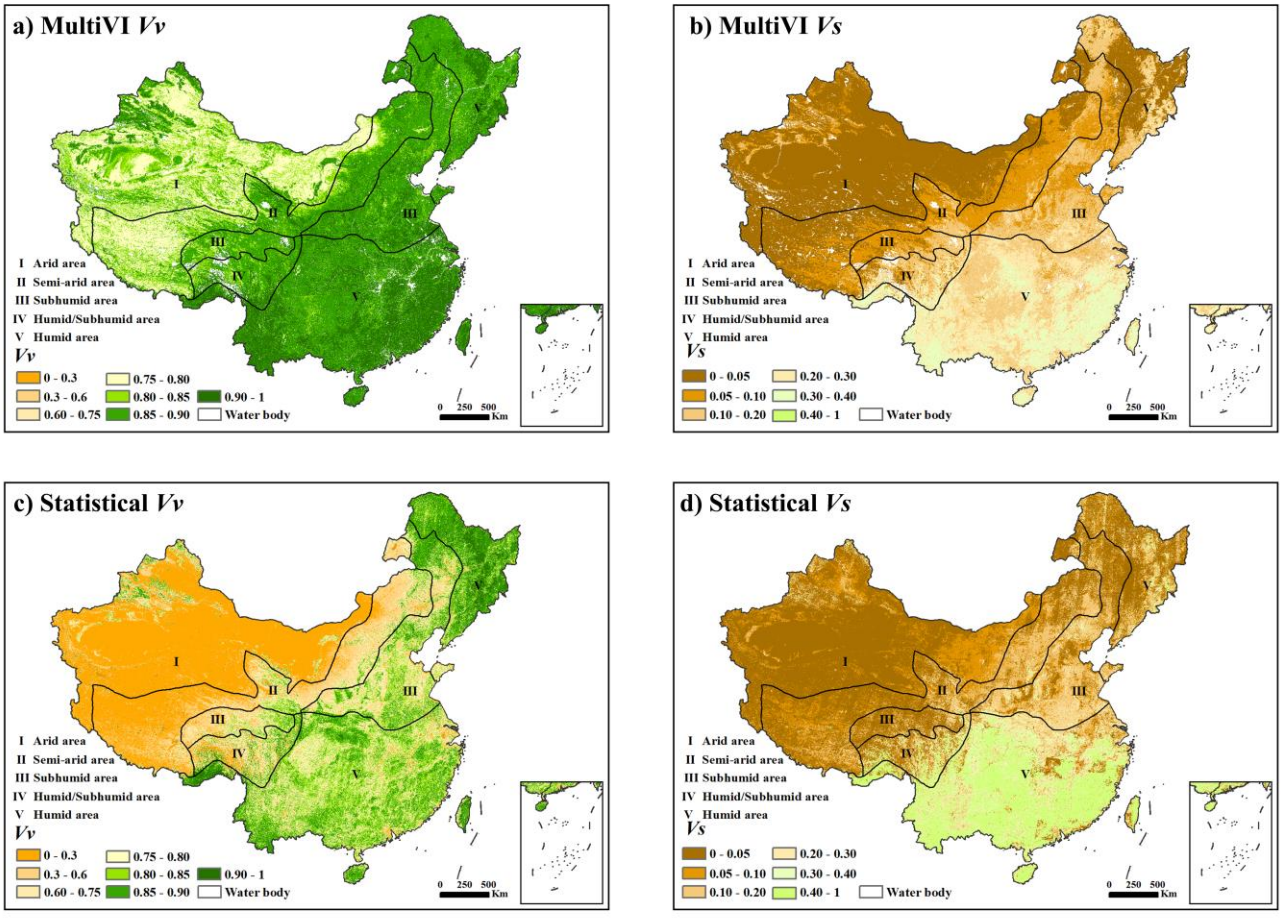


Figure 5: The spatial distributions of *Vv* and *Vs* generated using the MultiVI algorithm and the statistical method, respectively.

Table 3 MultiVI *Vv* and *Vs* for different land types.

Land type	mean *Vv*	standard deviation of *Vv*	mean *Vs*	standard deviation of *Vs*
Grassland	0.8569	0.0455	0.0740	0.0564

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Shrublands	0.8168	0.0469	0.0683	0.0439
Broadleaf croplands	0.8872	0.0310	0.1232	0.0789
Savananas	0.8927	0.0321	0.2194	0.0836
Evergreen broadleaf forests	0.9070	0.0286	0.3136	0.0596
Deciduous broadleaf forests	0.8942	0.0273	0.1737	0.0858
Evergreen needleleaf forests	0.9088	0.0316	0.2870	0.0725
Deciduous needleleaf forests	0.8752	0.0215	0.1564	0.0697

Figure 6 shows the differences between the MultiVI *Vv*/*Vs* and the statistical *Vv*/*Vs*. The largest discrepancies are observed in regions lacking pure pixels, *i.e.*, the arid areas where pure vegetation pixels are absent for *Vv* estimation, and the humid areas where bare soil pixels are lacking for *Vs* estimation.

In the sparse grasslands of northwestern China, the MultiVI *Vv* are significantly higher than the statistical *Vv*, with a bias exceeding 0.3 (Figure 6a). In most humid or subhumid areas of southeastern China, the difference between the two sets of *Vv* values is generally within ±0.1. For relatively sparse vegetation, such as grasslands and croplands, the MultiVI *Vv* are slightly higher than the statistical *Vv*. However, in forested areas, the MultiVI *Vv* are slightly lower than the statistical *Vv*.

In the densely vegetated forests of southeastern China, the MultiVI *Vs* are markedly lower than the statistical *Vs*, with a bias of less than -0.3 (Figure 6b). In arid regions, the MultiVI *Vs* values are slightly lower than the statistical *Vs* values in sparse grasslands and bare lands, but higher in oases, with a bias of approximately ±0.1.

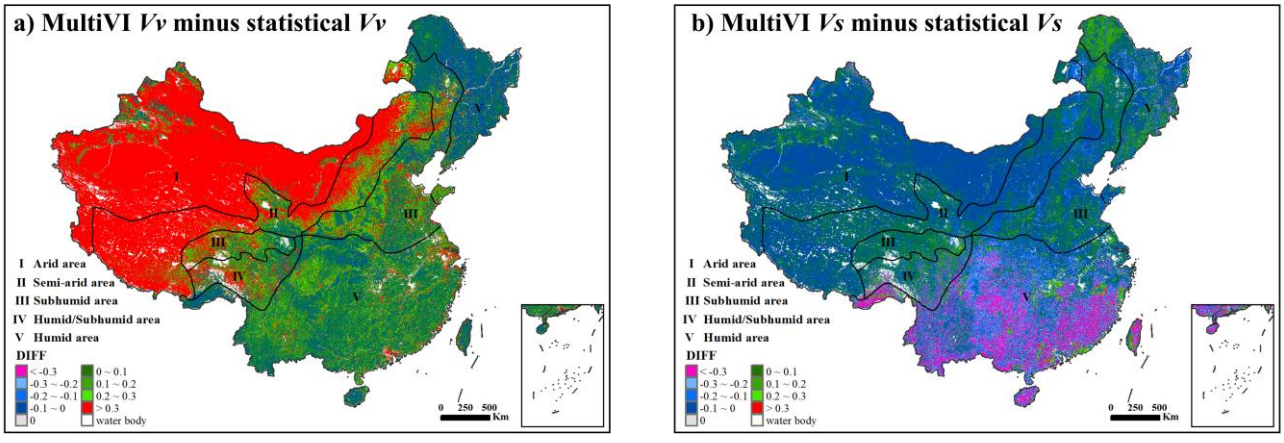


Figure 6: The difference maps between the MultiVI NDVI and statistical NDVI for pure vegetation and bare soil.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

4.2 Evaluation with soil NDVI from the soil spectral library

Figure 7 shows the MultiVI *Vs* and statistical *Vs* in comparison to the soil NDVI derived from the ICRAF soil library. For most soil types, the mean of the MultiVI *Vs* closely aligns with the median NDVI of the corresponding soil samples. In contrast, the mean statistical *Vs* tend to overestimate the median NDVI across most soil types. Both the mean values of the MultiVI and the statistical *Vs* are lower than the median NDVI for Dark Semi-hydromorphic soils, with a bias of approximately 0.1.

Soil samples of Anthrosols and Ferralisols are primarily distributed in the humid, densely vegetated regions of southeastern China, which results in relatively high NDVI values. For these soil types, the MultiVI *Vs* show an overestimation when compared to the median NDVI values, with biases of approximately 0.15 for Anthrosols and 0.2 for Ferralisols. This overestimation is more pronounced for the statistically derived *Vs*, with biases exceeding 0.2 for Anthrosols and 0.4 for Ferralisols.

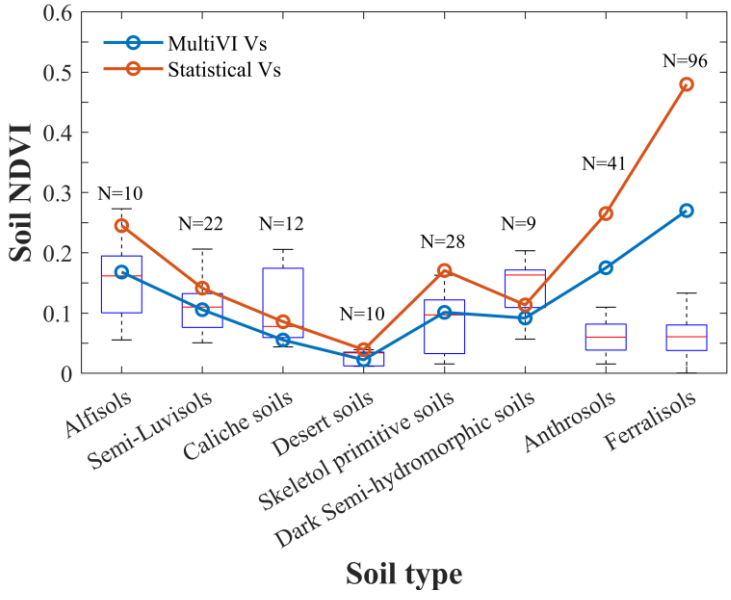
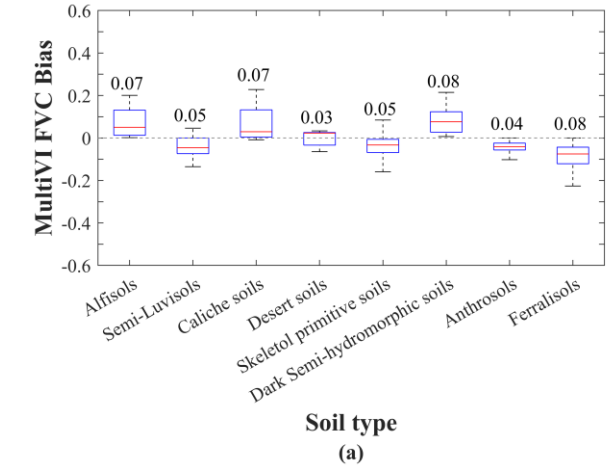


Figure 7: The boxplot of the soil NDVI from the ICRAF soil library for each soil type. Each boxplot features a central red line representing the median. The N above the box indicates the number of sampling plots for each soil type. The lower and upper edges of the box denote the 25th and 75th percentiles, respectively. The whiskers are extended to the most extreme data points excluding outliers. The blue and red lines denote the means of the MultiVI *Vs* and statistical *Vs*, respectively.

Figure 8 illustrates the bias between the MultiVI FVC and the statistical FVC compared to the reference FVC estimated using soil NDVI from the ICRAF soil library. The MultiVI FVC and statistical FVC were calculated using the MultiVI *Vv/Vs* (Figure 8a) and statistical *Vv/Vs* (Figure 8b), respectively. The reference FVC was estimated using the ICRAF soil NDVI in combination with either MultiVI *Vv* (Figure 8a) or statistical *Vv* (Figure 8b).

For the MultiVI FVC, the median bias is within ±0.05, and the mean absolute values consistently remain below 0.1 across all soil types. The overestimation of the MultiVI *Vs* yields a slight underestimation of FVC for Anthrosols and Ferralisols. The overestimations of 0.15 and 0.2 in MultiVI *Vs* lead to underestimations of approximately 0.04 and 0.08 in FVC for Anthrosols

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024



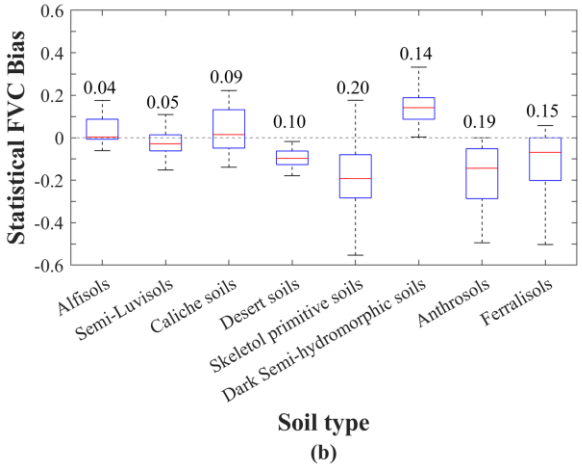


Figure 8: The bias in FVC estimation across different soil types using different *Vs* values. The FVC bias represents the difference between the estimated FVC using (a) MultiVI *Vs* or (b) statistical *Vs* and the reference FVC derived from soil NDVI in the ICRAF soil library. The number above each box indicates the mean absolute bias of FVC for each soil type. The lower and upper edges of the box denote the 25th and 75th percentiles, respectively. The whiskers are extended to the most extreme data points excluding outliers.

4.3 Accuracy of FVC estimation

Figure 9 depicts scatterplots that compare the field-measured FVC with the FVC estimated using the MultiVI *Vv/Vs* and the statistical *Vv/Vs* across three distinct sites: the Hebei site (Figures 9a and 9b), the Heihe site (Figures 9c and 9d), and the Three Gorges Reservoir site (Figures 9e and 9f).

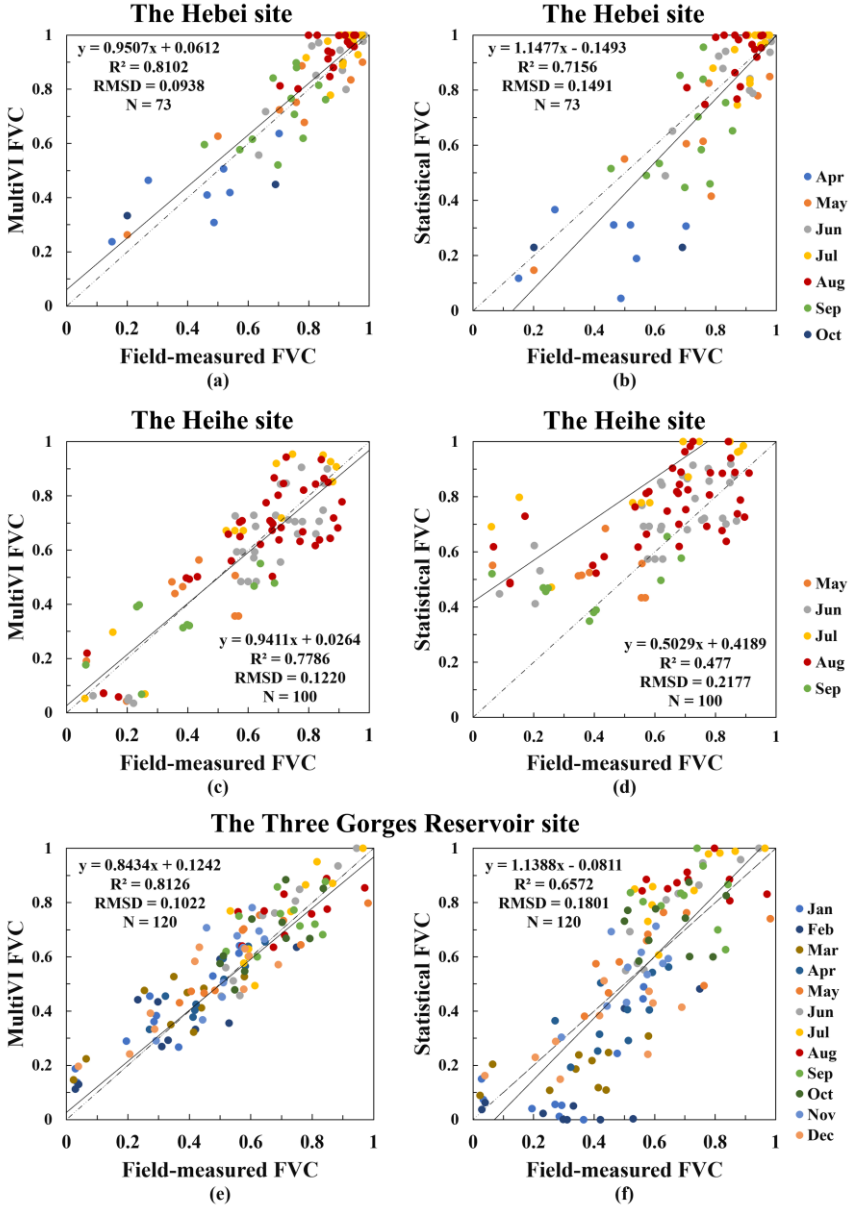
The MultiVI FVC demonstrates superior accuracy relative to the statistical FVC at the Hebei site, with a lower RMSD and a higher R^2 (0.8121 for the MultiVI FVC compared to 0.7156 for the statistical FVC). Both the MultiVI FVC and the statistical FVC show saturation effects for high FVC values at the Hebei site during the summer months (Figures 9a and 9b). However, the MultiVI FVC aligns more closely with the 1:1 line, particularly during the non-growing seasons, specifically in April, September, and October.

The MultiVI FVC outperforms the statistical FVC at the Heihe site, with an RMSD of 0.122 and an R^2 of 0.7786. In contrast, the statistical FVC presents an RMSD exceeding 0.2 and an R^2 below 0.5. The majority of the data points depicted in Figure

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

9d are situated above the 1:1 line, indicating that the statistical method tends to underestimate *Vv* values in arid regions, resulting in a significant overestimation of FVC.

At the Three Gorges Reservoir site, the MultiVI FVC also demonstrates superior accuracy, as indicated by an RMSD of 0.1022, in contrast to 0.1801 for the statistical FVC. Additionally, the correlation between the MultiVI FVC and field-measured FVC is significantly higher (R^2 = 0.8162) than that of the statistical FVC (R^2 = 0.6572). As illustrated in Figure 9f, a part of the statistical FVC data points is saturated at high values and drops to zero at low values, indicating that the statistical method suffers from an underestimation of *Vv* and an overestimation of *Vs* in humid areas.



https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Figure 9: Scatterplots of the MultiVI FVC and the statistical FVC versus the field-measured FVC. The R^2 and RMSD values are also shown. N is the number of samples.

5 Discussion

The VI-based mixture model is widely used to estimate FVC due to its ease of implementation. The model's simplicity is primarily attributed to the preselection of two critical parameters: the NDVI of bare soil (*Vs*) and that of pure vegetation (*Vv*). These parameters are essential for the model's performance and have a significant impact on its accuracy. The traditional statistical method for obtaining *Vv* and *Vs* assumes the existence of pure pixels, either spatially or temporally. In this method, extreme values from regional or temporal datasets are employed to represent *Vv* and *Vs*. However, this approach has two significant limitations: (1) pure pixels may be absent in certain ecosystems, such as evergreen forests, where bare soil pixels are lacking, or in bare lands, where pure vegetation pixels are absent; (2) the *Vv* and *Vs* can vary from pixel to pixel, yet the traditional statistical method often assigns a single value for an entire region or land type.

The global soil spectral library was used to calibrate *Vs* and to account for its spatial variability (Montandon and Small, 2008). Although this method improves the accuracy of FVC estimation compared to using a single *Vs* value across large regions, it is limited in that it only considers soil variability within the spectral library and fails to address the spectral differences among individual pixels. An alternative approach is to use each pixel's historical lowest NDVI values as *Vs* to ensure pixel-wise variability.

In this study, statistical maps of *Vv* and *Vs* were generated using a common statistical criterion (Sect. 3.3.1). However, two significant drawbacks were observed. First, empirical NDVI for fully vegetated pixels are typically reported to exceed 0.5 in most studies (Gao et al., 2020; Zeng et al., 2000; Montandon and Small, 2008), whereas the statistical *Vv* are observed to be below 0.3 in arid regions of China's mainland (Figure 5a). This discrepancy suggests that identifying appropriate NDVI values for pure pixels from datasets with limited spatial or temporal coverage is challenging, making the use of statistical methods to obtain pixel-wise pure NDVI values impractical. Second, inaccuracies in the statistical *Vv* and *Vs* contribute to significant uncertainty in FVC estimation. The overestimation of *Vs* leads to an underestimation of FVC, particularly in arid areas (Figure 10d) and during the non-growing season when NDVI values fall within a sensitive range (0.2 to 0.4) (Figure 10b). Furthermore, the underestimation of *Vv* results in the saturation of FVC (Figure 10f).

The MultiVI algorithm, which uses multi-angle data to retrieve *Vv* and *Vs*, has demonstrated effectiveness in estimating FVC and has been applied to generate FVC time series on a national scale (Mu et al., 2018; Song et al., 2022a). This algorithm retrieves angle-invariant NDVI values for endmembers using observations from two large VZAs. In this study, the retrieval strategy of the MultiVI algorithm was optimized through the construction of well-posed equations based on NDVI time series. Additionally, historical statistical values of *Vv* and *Vs* were incorporated as boundary constraints during the equation-solving process. The downscaling procedure integrates a 30 m land cover product, which introduces finer spatial detail to accurately describe the sub-pixel heterogeneity within MODIS pixels.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

The MultiVI *Vs* show the largest discrepancies for the soil types of Anthrosols and Ferralisols when compared to the soil NDVI from the ICRAF soil spectral library. This discrepancy is attributed to the relatively high NDVI associated with these two soil types, which are predominantly distributed in the humid areas of southern China. However, the overestimation of *Vs* in these areas has a limited impact on the FVC estimation (Figure 8a). Previous studies have also indicated that the influence of *Vs* on FVC estimation is more pronounced in areas with low NDVI values, such as grasslands and croplands / natural vegetation areas, compared to regions with high NDVI values (Ding et al., 2016). Despite the median statistical *Vs* are slightly closer to the soil NDVI from the ICRAF soil library for Caliche soils and Desert soils than MultiVI *Vs* (differences between the two *Vs* smaller than 0.03 in Figure 7), the bias of the statistical FVC is much larger than that of the MultiVI FVC (Figure 8). This indicates that the MultiVI algorithm derives more accurate *Vv* values than statistical methods especially for arid or semi-arid regions, thereby enhancing the precision of FVC estimation.

As a critical vegetation parameter, FVC is frequently demanded in various studies, serving as an input for models or as fundamental data for ecological analyses. The VI-based mixture is the simplest method for converting remotely sensed images into FVC products, provided that the pure NDVI values are obtained beforehand. However, the two essential parameters in the VI-based model, *Vv* and *Vs*, currently lack standardized and reliable data sources. The newly generated 30 m *Vv* and *Vs* maps address this gap. Users can now calculate accurate FVC values quickly using the MultiVI *Vv* and *Vs* with corresponding NDVI to meet their specific requirements.

6 Conclusion

This study demonstrates the feasibility of generating pixel-wise NDVI for pure vegetation (*Vv*) and bare soil (*Vs*) for the VI-based mixture model. In this study, 30 m resolution maps of pure NDVI values for the year 2014 were produced for China's mainland using multi-angle remotely sensed data. The assessment and validation of the *Vv* and *Vs* maps were conducted from three aspects: 1) comparing the MultiVI *Vv* and *Vs* maps with those generated using the statistical method; 2) comparing the derived *Vs* with reference soil NDVI from the ICRAF soil spectral library; and 3) validating the accuracy of FVC calculated from the pure NDVI values against field-measured FVC. Our findings reveal the urgent need for reliable *Vv* and *Vs* per pixel for large-area FVC production. Traditional statistical methods are impractical to achieve this goal due to their reliance on pure pixels. The MultiVI algorithm has proven to be a viable solution, yielding *Vv* and *Vs* with a more coherent spatial pattern and magnitude than statistical methods. The MultiVI *Vs* closely aligned with soil NDVI across various soil types in China's mainland. Furthermore, the FVC estimated using the MultiVI *Vv* and *Vs* demonstrated improved accuracy in comparison to those derived from the statistical *Vv* and *Vs*, with RMSD values around 0.1 and R^2 values near 0.8 for all validation sites. Moreover, the products generated in this study show broad applicability across a variation of climate zones and soil types.

The MultiVI *Vv* and *Vs* maps provide essential parameters for FVC estimation using the widely adopted VI-based mixture model, which is known for its ease of use and reasonable accuracy. Hence, these derived *Vv* and *Vs* maps are anticipated to facilitate the estimation of fine-resolution, high-frequency FVC with reliable quality at large scales.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

7 Data availability

The 30 m MultiVI *Vv* and *Vs* maps are available at https://zenodo.org/records/14060222 (Zhao et al., 2024). The *Vv* and *Vs* data are stored in GeoTIFF files using the WGS-84 (World Geodetic System 1984) coordinate system with UTM (Universal Transverse Mercator) projection. The data are categorized into tiles based on a latitude size of 5° and a longitude size of 6°. The file names consist of 18 characters following these rules: North-South latitude abbreviation (1 digit) + 6-degree zone number (2 digits) + "_" + starting latitude (2 digits) + "_" + product time (4 digits) + "_ " + resolution (3 digits) + "_ " + data attribute (2 digits). Additionally, each image contains a pure NDVI value band range from 0 to 100, with invalid values or water surfaces labelled as 0.

8 Author contribution

Xihan Mu, Wanjuan Song and Tian Zhao designed the methodology. Wanjuan Song and Tian Zhao programmed the software and generated the data. Tian Zhao wrote the original draft, and Xihan Mu reviewed the manuscript. Yun Xie, Donghui Xie, and Guangjian Yan supervised the project.

9 Competing interests

The authors declare that they have no conflict of interest.

References

Asrar, G., Fuchs, M., Kanemasu, E., and Hatfield, J.: Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat 1, Agronomy journal, 76, 300-306, https://doi.org/10.2134/agronj1984.00021962007600020029x, 1984.

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., and Smets, B.: GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production, Remote Sensing of Environment, 137, 299-309, https://doi.org/10.1016/j.rse.2012.12.027, 2013.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J. L., and Leroy, M.: LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm, Remote Sensing of Environment, 110, 275-286, http://dx.doi.org/10.1016/j.rse.2007.02.018, 2007.

Chen, J., Menges, C., and Leblanc, S.: Global mapping of foliage clumping index using multi-angular satellite data, Remote Sensing of Environment, 97, 447-457, https://doi.org/10.1016/j.rse.2005.05.003, 2005.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., and Lu, M.: Global land cover mapping at 30 m resolution: A POK-based operational approach, ISPRS Journal of Photogrammetry and Remote Sensing, 103, 7-27, https://doi.org/10.1016/j.isprsjprs.2014.09.002, 2015.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., Peng, S., Han, G., Zhang, H., and He, C.: Concepts and key techniques for 30 m global land cover mapping, Acta Geodaetica et Cartographica Sinica, 43, 551-557, http://dx.doi.org/10.13485/j.cnki.11-2089.2014.0089, 2014.

Choudhury, B. J., Ahmed, N. U., Idso, S. B., Reginato, R. J., and Daughtry, C.: Relations between evaporation coefficients and vegetation indices studied by model simulations, Remote Sensing of Environment, 50, 1-17, https://doi.org/10.1016/0034-4257(94)90090-6, 1994.

COCHRANE, M. A. and SOUZA Jr, C. M.: Linear mixture model classification of burned forests in the eastern Amazon, International Journal of Remote Sensing, 19, 3433-3440, https://doi.org/10.1080/014311698214109, 1998.

Deardorff, J. W.: Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation, Journal of Geophysical Research: Oceans, 83, https://doi.org/10.1029/JC083iC04p01889, 1978.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Deering, D.: Structure analysis and classification of boreal forests using airborne hyperspectral BRDF data from ASAS, Remote Sensing of Environment, 69, 281-295, https://doi.org/10.1016/S0034-4257(99)00032-2, 1999.
DiMiceli, C., Carroll, M., Sohlberg, R., Huang, C., Hansen, M., and Townshend, J.: Annual global automated MODIS vegetation continuous fields (MOD44B) at 250 m spatial resolution for data years beginning day 65, 2000–2010, collection 5 percent tree cover [dataset], http://doi.org/10.5067/MODIS/MOD44B.006, 2011.
Diner, D. J., Asner, G. P., Davies, R., Knyazikhin, Y., Muller, J.-P., Nolin, A. W., Pinty, B., Schaaf, C. B., and Stroeve, J.: New directions in earth observing: Scientific applications of multiangle remote sensing, Bulletin of the American Meteorological Society, 80, 2209-2228, https://doi.org/10.1175/1520-0477(1999)080<2209:NDIEOS>2.0.CO;2, 1999.
Ding, Y., Zheng, X., Zhao, K., Xin, X., and Liu, H.: Quantifying the impact of NDVIsoil determination methods and NDVIsoil variability on the estimation of fractional vegetation cover in Northeast China, Remote Sensing, 8, 29, https://doi.org/10.3390/rs8010029, 2016.
Donohue, R. J., Hume, I., Roderick, M. L., McVicar, T. R., Beringer, J., Hutley, L. B., Gallant, J. C., Austin, J. M., Van Gorsel, E., and Cleverly, J. R.: Evaluation of the remote-sensing-based DIFFUSE model for estimating photosynthesis of vegetation, Remote Sensing of Environment, 155, 349-365, https://doi.org/10.1016/j.rse.2014.09.007, 2014.
Eriksson, H. M., Eklundh, L., Kuusk, A., and Nilson, T.: Impact of understory vegetation on forest canopy reflectance and remotely sensed LAI estimates, Remote Sensing of Environment, 103, 408-418, https://doi.org/10.1016/j.rse.2006.04.005, 2006.
Escadafal, R. and Huete, A.: Influence of the viewing geometry on the spectral properties (high resolution visible and NIR) of selected soils from Arizona, Proceedings of the 5th International Colloquium, Physical Measurements and Signatures in Remote Sensing, 401-404, https://doi.org/10.1016/0034-4257(89)90013-8, 1991.
Fernández-Guisuraga, J. M., Verrelst, J., Calvo, L., and Suárez-Seoane, S.: Hybrid inversion of radiative transfer models based on high spatial resolution satellite reflectance data improves fractional vegetation cover retrieval in heterogeneous ecological systems after fire, Remote sensing of environment, 255, 112304, https://doi.org/10.1016/j.rse.2021.01.002, 2021.
Gan, M., Deng, J., Zheng, X., Hong, Y., and Wang, K.: Monitoring Urban Greenness Dynamics Using Multiple Endmember Spectral Mixture Analysis, Plos One, 9, https://doi.org/10.1371/journal.pone.0112202, 2014.
Gao, L., Wang, X., Johnson, B. A., Tian, Q., Wang, Y., Verrelst, J., Mu, X., and Gu, X.: Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review, ISPRS Journal of Photogrammetry and Remote Sensing, 159, 364-377, https://doi.org/10.1016/j.isprsjprs.2019.11.018, 2020.
García-Haro, F. J., Sommer, S., and Kemper, T.: A new tool for variable multiple endmember spectral mixture analysis (VMESMA), International Journal of Remote Sensing, 26, 2135-2162, https://doi.org/10.1080/01431160512331337817, 2005.
Garrity, D. and Bindraban, P.: A globally distributed soil spectral library visible near infrared diffuse reflectance spectra, World Agroforestry Centre and ISRIC - World Soil Information [dataset], 2006.
Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D.: Novel algorithms for remote estimation of vegetation fraction, Remote Sensing of Environment, 80, 76-87, https://doi.org/10.1016/S0034-4257(01)00289-9, 2002.
Goel, N. S. and Strebel, D. E.: Simple Beta Distribution Representation of Leaf Orientation in Vegetation Canopies, Agronomy Journal, 76, 800, https://doi.org/10.2134/agronj1984.00021962007600050021x, 1984.
Guan, K., Wood, E. F., and Caylor, K. K.: Multi-sensor derivation of regional vegetation fractional cover in Africa, Remote Sensing of Environment, 124, 653-665, http://dx.doi.org/10.1016/j.rse.2012.06.005, 2012.
Gutman, G. and Ignatov, A.: Satellite-derived green vegetation fraction for the use in numerical weather prediction models, Advances in Space Research, 19, 477-480, https://doi.org/10.1080/014311698215333, 1997.
Gutman, G. and Ignatov, A.: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, International Journal of Remote Sensing, 19, 1533-1543, https://doi.org/10.1080/014311698215333, 1998.
He, L., Chen, J. M., Pisek, J., Schaaf, C., and Strahler, A. H.: Global clumping index map derived from modis BRDF products, Geoscience & Remote Sensing Symposium, https://doi.org/10.1016/J.RSE.2011.12.008,
Hirano, Y., Yasuoka, Y., and Ichinose, T.: Urban climate simulation by incorporating satellite-derived vegetation cover distribution into a mesoscale meteorological model, Theoretical & Applied Climatology, 79, 175-184, https://doi.org/10.1007/s00704-004-0069-0, 2004.
Jensen, J. R.: Remote Sensing of the Environment: An Earth Resource Perspective, Prentice Hall, Upper Saddle River, New Jersey, 333−377 pp.2000.
Jia, K., Li, Y., Liang, S., Wei, X., and Yao, Y.: Combining estimation of green vegetation fraction in an arid region from Landsat 7 ETM+ data, Remote Sensing, 9, 1121, https://doi.org/10.3390/rs9111121, 2017.
Jia, K., Liang, S., Liu, S., Li, Y., Xiao, Z., Yao, Y., Jiang, B., Zhao, X., Wang, X., Xu, S., and Cui, J.: Global Land Surface Fractional Vegetation Cover Estimation Using General Regression Neural Networks From MODIS Surface Reflectance, IEEE Transactions on Geoscience and Remote Sensing, 53, 4787-4796, https://doi.org/10.1109/TGRS.2015.2409563, 2015.
Jiao, Q., Zhang, B., Liu, L., Li, Z., Yue, Y., and Hu, Y.: Assessment of spatio-temporal variations in vegetation recovery after the Wenchuan earthquake using Landsat data, Natural hazards, 70, 1309-1326, https://doi.org/10.1007/s11069-013-0875-8, 2014.
Jiapaer, G., Chen, X., and Bao, A.: A comparison of methods for estimating fractional vegetation cover in arid regions, Agricultural and Forest Meteorology, 151, 1698-1710, https://doi.org/10.1016/j.agrformet.2011.07.004, 2011.
Jun, C., Ban, Y., and Li, S.: China: Open access to Earth land-cover map, Nature, 514, https://doi.org/10.1038/514434c, 2014.

Leblanc, S. G., Chen, J. M., Miller, J. R., and Freemantle, J.: Compact Airborne Spectrographic Imager (CASI) Used for Mapping LAI of Cropland, Journal of Geophysical Research: Atmospheres, 104, 27945–27958, https://doi.org/10.1029/1999JD900098, 1999.
Li, G., Jing, Y., Wu, Y., and Zhang, F.: Improvement of two evapotranspiration estimation models using a linear spectral mixture model over a small agricultural watershed, Water, 10, 474, https://doi.org/10.3390/w10040474, 2018a.
Li, L., Mu, X., Macfarlane, C., Song, W., Chen, J., Yan, K., and Yan, G.: A half-Gaussian fitting method for estimating fractional vegetation cover of corn crops using unmanned aerial vehicle images, Agricultural and Forest Meteorology, 262, 379-390, https://doi.org/10.1016/j.dib.2021.106784, 2018b.
Li, X., Zhang, X., Zhang, L., and Wu, B.: Rainfall and Vegetation Coupling Index for soil erosion risk mapping, Journal of Soil and Water Conservation, 69, 213-220, https://doi.org/10.2489/jswc.69.3.213, 2014.
Li, X., Liu, S., Ma, M., Xiao, Q., Liu, Q., Jin, R., Che, T., Wang, W., Qi, Y., Li, H., Zhu, G., Guo, J., Ran, Y., Wen, J., and Wang, S.: HiWATER: An Integrated Remote Sensing Experiment on Hydrological and Ecological Processes in the Heihe River Basin, Advances in Earth Science, 27, 481-498, 10.11867/j.issn.1001-8166.2012.05.0481, 2012.
Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., and Mi, J.: Finer-resolution mapping of global land cover: Recent developments, consistency analysis, and prospects, Journal of Remote Sensing, https://doi.org/10.34133/2021/5289697, 2021.
Lu, D., Moran, E., and Batistella, M.: Linear mixture model applied to Amazonian vegetation classification, Remote sensing of environment, 87, 456-469, https://doi.org/10.1016/j.rse.2002.06.001, 2003.
Matsui, T., Lakshmi, V., and Small, E. E.: The effects of satellite-derived vegetation cover variability on simulated land–atmosphere interactions in the NAMS, Journal of Climate, 18, 21-40, https://doi.org/10.1175/JCLI3254.1, 2005.
Mölders, N. and Olson, M. A.: Impact of urban effects on precipitation in high latitudes, Journal of Hydrometeorology, 5, 409-429, https://doi.org/10.1175/1525-7541(2004)005<0409:IOUEOP>2.0.CO;2, 2004.
Montandon, L. and Small, E.: The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI, Remote Sensing of Environment, 112, 1835-1845, https://doi.org/10.1016/j.rse.2007.09.007, 2008.
Mu, X., Huang, S., and Chen, Y.: HiWATER: Dataset of Fractional Vegetation Cover in the middle reaches of the Heihe River Basin, Heihe Plan Science Data Center [dataset], https://cstr.cn/18406.11.hiwater.043.2013.db, 2013.
Mu, X., Huang, S., Ren, H., Yan, G., Song, W., and Ruan, G.: Validating GEOV1 fractional vegetation cover derived from coarse-resolution remote sensing images over croplands, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8, 439-446, https://doi.org/10.1109/JSTARS.2014.2342257, 2014.
Mu, X., Song, W., Zhan, G., Mcvicar, T. R., Donohue, R. J., and Yan, G.: Fractional vegetation cover estimation by using multi-angle vegetation index, Remote Sensing of Environment, 216, 44-56, https://doi.org/10.1016/j.rse.2018.06.022, 2018.
Mu, X., Hu, M., Song, W., Ruan, G., Ge, Y., Wang, J., Huang, S., and Yan, G.: Evaluation of sampling methods for validation of remotely sensed fractional vegetation cover, Remote Sensing, 7, 16164-16182, https://doi.org/10.3390/rs71215817, 2015.
Mu, X., Zhao, T., Ruan, G., Song, J., Wang, J., Yan, G., Mcvicar, T. R., Yan, K., Gao, Z., and Liu, Y.: High spatial resolution and high temporal frequency (30-m/15-day) fractional vegetation cover estimation over China using multiple remote sensing datasets: Method development and validation, Journal of Meteorological Research, 35, 128-147, https://doi.org/10.1007/s13351-021-0017-2, 2021.
Naqvi, H. R., Mallick, J., Devi, L. M., and Siddiqui, M. A.: Multi-temporal annual soil loss risk mapping employing Revised Universal Soil Loss Equation (RUSLE) model in Nun Nadi Watershed, Uttrakhand (India), Arabian Journal of Geosciences, 6, 4045-4056, https://doi.org/10.1007/s12517-012-0661-z, 2013.
Nilson, T.: A theoretical analysis of the frequency of gaps in plant stands, Agricultural Meteorology, 8, 25-38, https://doi.org/10.1016/0002-1571(71)90092-6, 1971.
Oleson, K., Emery, W., and Maslanik, J.: Evaluating land surface parameters in the Biosphere-Atmosphere Transfer Scheme using remotely sensed data sets, Journal of Geophysical Research: Atmospheres, 105, 7275-7293, https://doi.org/10.1029/1999JD901041, 2000.
Roujean, J.-L., Leroy, M., and Deschamps, P.-Y.: A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data, Journal of Geophysical Research: Atmospheres, 97, 20455–20468, https://doi.org/10.1029/92JD01411, 1992.
Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., and Muller, J.-P.: First operational BRDF, albedo nadir reflectance products from MODIS, Remote sensing of Environment, 83, 135-148, https://doi.org/10.1016/S0034-4257(02)00091-3, 2002.
Song, W., Mu, X., Yan, G., and Huang, S.: Extracting the Green Fractional Vegetation Cover from Digital Images Using a Shadow-Resistant Algorithm (SHAR-LABFVC), Remote Sensing, 7, 10425, https://doi.org/10.3390/rs70810425, 2015.
Song, W., Mu, X., Ruan, G., Gao, Z., Li, L., and Yan, G.: Estimating fractional vegetation cover and the vegetation index of bare soil and highly dense vegetation with a physically based method, International journal of applied earth observation and geoinformation, 58, 168-176, https://doi.org/10.1016/j.jag.2017.01.015, 2017.
Song, W., Mu, X., McVicar, T. R., Knyazikhin, Y., Liu, X., Wang, L., Niu, Z., and Yan, G.: Global quasi-daily fractional vegetation cover estimated from the DSCOVR EPIC directional hotspot dataset, Remote Sensing of Environment, 269, 112835, https://doi.org/10.1016/j.rse.2021.112835, 2022a.

https://doi.org/10.5194/essd-2024-535
Preprint. Discussion started: 9 December 2024

Song, W., Zhao, T., Mu, X., Zhong, B., Zhao, J., Yan, G., Wang, L., and Niu, Z.: Using a vegetation index-based mixture model to estimate fractional vegetation cover products by jointly using multiple satellite data: Method and feasibility analysis, Forests, 13, 691, https://doi.org/10.3390/f13050691, 2022b.

Verrelst, J., Schaepman, M. E., Koetz, B., and Kneubühler, M.: Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data, Remote Sensing of Environment, 112, 2341-2353, https://doi.org/10.1016/j.rse.2007.11.001, 2008.

Wang, C. and Qi, J.: Biophysical estimation in tropical forests using JERS-1 VNIR imagery. I: Leaf area index, International Journal of Remote Sensing, 29, 6811-6826, https://doi.org/10.1080/01431160802270115, 2008.

Weiss, M., Baret, F., Garrigues, S., and Lacaze, R.: LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products, Remote Sensing of Environment, 110, 317-331, https://doi.org/10.1016/j.rse.2007.03.001, 2007.

Weiss, M., Baret, F., Smith, G., Jonckheere, I., and Coppin, P.: Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling, Agricultural and forest meteorology, 121, 37-53, https://doi.org/10.1016/j.agrformet.2003.08.001, 2004.

Wu, D., Wu, H., Zhao, X., Zhou, T., Tang, B., Zhao, W., and Jia, K.: Evaluation of Spatiotemporal Variations of Global Fractional Vegetation Cover Based on GIMMS NDVI Data from 1982 to 2011, Remote Sensing, 6, 4217, https://doi.org/10.3390/rs6054217, 2014.

Xiao, J. and Moody, A.: A comparison of methods for estimating fractional green vegetation cover within a desert-to-upland transition zone in central New Mexico, USA, Remote Sensing of Environment, 98, 237-250, http://dx.doi.org/10.1016/j.rse.2005.07.011, 2005.

Xiao, Z., Wang, T., Liang, S., and Sun, R.: Estimating the Fractional Vegetation Cover from GLASS Leaf Area Index Product, Remote Sensing, 8, 337, https://doi.org/10.3390/rs8040337, 2016.

Xie, M., Wang, Y., and Fu, M.: An Overview and Perspective about Causative Factors of Surface Urban Heat Island Effects, Progress in Geography, 30, 35-41, https://doi.org/10.11820/dlkxjz.2011.01.004, 2011.

Yang, H. and Yang, Z.: A modified land surface temperature split window retrieval algorithm and its applications over China, Global and Planetary Change, 52, 207-215, https://doi.org/10.1016/j.gloplacha.2006.02.015, 2006.

Zeng, X., Dickinson, R. E., Walker, A., Shaikh, M., DeFries, R. S., and Qi, J.: Derivation and Evaluation of Global 1-km Fractional Vegetation Cover Data for Land Modeling, Journal of Applied Meteorology, 39, 826-839, 10.1175/1520-0450(2000)039<0826:DAEOGK>2.0.CO;2, 2000.

Zhang, Y., Odeh, I., and Ramadan, E.: Assessment of land surface temperature in relation to landscape metrics and fractional vegetation cover in an urban/peri-urban region using Landsat data, International Journal for Remote Sensing, 34, 168-189, https://doi.org/10.1080/01431161.2012.712227, 2013.

Zhao, T., Song, W., Mu, X., Xie, Y., Xie, D., and Yan, G.: Normalized Difference Vegetation Index Maps of Pure Pixels over China for Estimation of Fractional Vegetation Cover (2014) (1.0.0) [dataset], https://doi.org/10.5281/zenodo.14060222, 2024.

Zhao, T., Mu, X., Song, W., Liu, Y., Xie, Y., Zhong, B., Xie, D., Jiang, L., and Yan, G.: Mapping Spatially Seamless Fractional Vegetation Cover over China at a 30-m Resolution and Semimonthly Intervals in 2010–2020 Based on Google Earth Engine, Journal of Remote Sensing, 3, 0101, https://doi.org/10.34133/remotesensing.0101, 2023.