

Article

Enhancing Urban Above-Ground Vegetation Carbon Density Mapping: An Integrated Approach Incorporating De-Shadowing, Spectral Unmixing, and Machine Learning

Guangping Qie ^{1,2,3}, Jianneng Ye ⁴, Guangxing Wang ³  and Minzi Wang ^{5,*} ¹ Department of Tourism Management, Moutai Institute, Renhuai 551801, China; qieguangping@mtxy.edu.cn² Department of Geography and Environmental Resources, Southern Illinois University at Carbondale, Carbondale, IL 62901, USA³ School of Earth Systems and Sustainability, Southern Illinois University at Carbondale, Carbondale, IL 62901, USA; gxwang@siu.edu⁴ Department of Student Affairs, Zhejiang Gongshang University, Hangzhou 310018, China; jiannengye@mail.zjgsu.edu.cn⁵ Department of Resource and Environment, Moutai Institute, Renhuai 551801, China

* Correspondence: wangminzi@mtxy.edu.cn; Tel.: +86-1830-0942-108

Abstract: Accurately mapping urban above-ground vegetation carbon density presents challenges due to fragmented landscapes, mixed pixels, and shadows induced by buildings and mountains. To address these issues, a novel methodological framework is introduced, utilizing a linear spectral unmixing analysis (LSUA) for shadow removal and vegetation information extraction from mixed pixels. Parametric and nonparametric models, incorporating LSUA-derived vegetation fraction, are compared, including linear stepwise regression, logistic model-based stepwise regression, k-Nearest Neighbors, Decision Trees, and Random Forests. Applied in Shenzhen, China, the framework integrates Landsat 8, Pleiades 1A & 1B, DEM, and field measurements. Among the key findings, the shadow removal algorithm is effective in mountainous areas, while LSUA-enhanced models improve urban vegetation carbon density mapping, albeit with marginal gains. Integrating kNN and RF with LSUA reduces errors, and Decision Trees, especially when integrated with LSUA, outperform other models. This study underscores the potential of the proposed framework, particularly the integration of Decision Trees with LSUA, for advancing the accuracy of urban vegetation carbon density mapping.

Keywords: urban; vegetation carbon density; mapping; de-shadow; spectral unmixing; machine learning



Citation: Qie, G.; Ye, J.; Wang, G.; Wang, M. Enhancing Urban Above-Ground Vegetation Carbon Density Mapping: An Integrated Approach Incorporating De-Shadowing, Spectral Unmixing, and Machine Learning. *Forests* **2024**, *15*, 480. <https://doi.org/10.3390/f15030480>

Academic Editors: Andrzej Węgiel and Adrian Lukowski

Received: 25 January 2024

Revised: 23 February 2024

Accepted: 2 March 2024

Published: 4 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Urbanization in China, with 65.2% of the population residing in urban areas [1], has led to rapid economic development and increased migration to cities, resulting in reduced vegetation cover and elevated greenhouse gas emissions [2]. Urban areas have become significant contributors to climate change, emphasizing the urgency of understanding and managing urban vegetation carbon density [3]. Recognizing the crucial role of urban vegetation as a carbon sink, accurate estimates of carbon storage and sequestration are essential for informed decision making by governments. Despite increasing awareness of environmental issues, mapping urban vegetation carbon density faces challenges due to complex landscapes, mixed pixels, and mountain- and building-induced shadows [4].

Carbon storage in urban vegetation, predominantly facilitated by urban forests, is a pivotal element in carbon sequestration [5]. Urban forests, encompassing both woody and associated herbaceous plants within and surrounding settlements, demonstrate the capacity to sequester substantial carbon in both above-ground and below-ground biomass [6]. The estimation of carbon levels at various scales involves a diverse array of methods, with local-scale assessments relying on a combination of field and remote sensing data [7–10]. These

assessments take into account factors such as tree species, canopy structure, diameters at breast height (DBH), spatial resolution, spectral bands, and sensor characteristics, etc. Moving to a regional scale, estimates incorporate additional variables related to climate and the environment [11–14]. Commonly, allometric equations, grounded in physiological relationships, are employed for biomass calculations, and their specifics may vary by region [15–20]. While specific allometric equations for urban forests are limited, alternative methods encompass the utilization of volume tables for tree biomass and the transfer of biomass calculations for shrubs and grass in green urban areas. Remote sensing techniques offer a cost-effective strategy for assessing vegetation biomass or carbon density across extensive areas. The extensive adoption of various satellite data sources such as Landsat Thematic Mapper, Landsat 8, Sentinel-2, MODIS, among others, provides multispectral images and seamless integration with Geographic Information Systems (GIS). Various techniques, including spatial interpolation and regression analyses, synergize field data with remote sensing for mapping forest carbon density [21–24]. Remote sensing-based methods, especially utilizing optical sensors like Landsat, MODIS, SPOT, and ALOS, offer promising avenues for mapping vegetation biomass or carbon density [25–27]. Landsat, with its free availability, medium spatial resolution, and historical data, is widely used, yet may underestimate biomass in dense tropical rainforests [7]. Coarse spatial resolution MODIS images suit global estimates but are unsuitable for small areas [28]. Very fine spatial resolution images, e.g., QuickBird and Worldview series, are suitable for urban mapping but hindered by high costs [29]. Active remote sensing techniques, such as LiDAR and TLS, offer advantages in penetrating clouds but pose challenges in cost and sensitivity to certain forest types [30–33]. Urban vegetation mapping is particularly challenging due to fragmented landscapes and building- and mountain-induced shadows, with limited research in this domain.

Fragmented urban landscapes often exhibit mixed pixels, complicating the effective use of remote sensing data for Land Use and Land Cover (LULC) or change detection analysis [34,35]. Spectral unmixing analysis, also known as Spectral Mixture Analysis or Spectral Mixture Modeling, addresses this challenge by extracting sub-pixel information through decomposing mixed pixels into fractional images corresponding to each endmember, representing a pure LULC type [36–40]. Spectral unmixing analysis methods fall into two categories based on the algorithms used: linear mixture models (LMMs) and nonlinear mixture models (NLMMs). LMMs assume that a mixed pixel's DN value is determined by the weights of endmember spectra and corresponding coverage area percentages, with the condition that incident light interacts with only one surface component. While NLMMs face challenges due to their intrinsic complexity, particularly in modeling and obtaining scene parameters, LMMs remain more widely used in spectral unmixing analysis. Notably, there is a lack of reported studies discussing the validation of spectral unmixing analysis results.

Following shadow detection, efforts to remove or minimize shadow impacts involve algorithms targeting topographic, urban building, cloud, and multi-layered shadows. Recent studies emphasize information recovery from shadows before their elimination, acknowledging the weak information recorded by sensors in shadowed areas [41–43]. Terrain shadows in mountainous regions, caused by low sun elevation angles and steep slopes, reduce reflectance, causing spectral heterogeneity in land cover pixels [44]. Neglecting mountain shadows in forest cover mapping leads to underestimation, emphasizing the need for their removal. Methods like NDVI and band ratios, though simple, are influenced by noise and lose spectral. Utilizing DEM combined with NDVI and topographic correction models are employed to mitigate mountain shadows [45,46]. Urban areas face challenges from building-induced shadows, hindering information extraction from high-resolution images. Recent efforts focus on shadow removal for urban regions, employing algorithms assuming a linear relationship between radiance in shadow and non-shadow areas [47–49]. Shadow detection and de-shadowing are crucial preprocessing steps for satellite images, significantly improving land use and land cover (LULC) classification accuracy and facilitating vegetation carbon density mapping.

Despite challenges in mapping urban vegetation, the integration of methods such as shadow removal and spectral unmixing analysis holds promise for modeling carbon density in urban environments. This research addresses this gap by emphasizing the urgency of developing advanced methods to eliminate the effects of mixed pixels and building and mountain shadows, providing a foundation for precise urban vegetation carbon density estimates. The proposed methodology integrates machine learning algorithms, de-shadowing, and spectral unmixing analysis, offering a comprehensive approach to enhance accuracy in carbon density mapping. The objectives include comparing pixel selection methods, developing shadow removal algorithms, and assessing various spatial modeling techniques. By focusing on Shenzhen, a rapidly growing city in China, this study aims to answer critical questions about model accuracy and shadow removal efficacy. This research contributes to the broader context of combating global warming and underscores the importance of precise information for effective urban vegetation carbon management.

2. Study Area and Datasets

2.1. Study Area

The study area is situated in Guangdong Province, southern China. Shenzhen city spans from 113°51' E to 114°21' E and 22°27' N to 22°39' N, covering an area of 1996.85 km² (Figure 1). Bordered by Mirs Bay to the east, the Pearl River estuary to the west, and adjacent to the New Territory of Hong Kong in the south, Shenzhen features a diverse topography. The southeast exhibits rugged terrain, while the northwest is predominantly flat, with mountainous and hilly areas. Wutong Shan, the highest mountain at 943.7 m, overlooks the coastline extending 230 km with six deep-water ports. The region includes 160 rivulets, major rivers like Shenzhen River, Maozhou River, Longgang River, Guanlan River, and Pingshan River, with catchment areas exceeding 100 km² but low surface run-off.

Shenzhen experiences a temperate monsoon climate, with a north/northeast monsoon prevailing from September to mid-March, bringing cool, dry air. The summer monsoon dominates from April to September, resulting in hot, humid weather with a heightened risk of typhoons. The annual mean temperature is 22.4 °C, ranging from 12.1 °C in January to 28.1 °C in July, with extremes recorded at 38.7 °C and 0.2 °C. The frost-free period extends over 355 days, with annual mean precipitation of 1933.3 mm mostly occurring between May and September. The city's geology includes granite, naceous shale, tuff, and metamorphic and sandstone rocks. Two main soil types are prevalent: hill soils and alluvial soils, with the latter confined to riverbanks, river plains, and the seashore.

Historically covered by climatic vegetation, Shenzhen's landscape has evolved due to human disturbances. Existing forests, mainly secondary and man-made, encompass lowland evergreen monsoon forests, montane evergreen broad-leaved forests, ravine rainforests, mangroves, and plantations. Since becoming the first special economic zone in 1979, Shenzhen has undergone rapid development, transitioning from a population of 20,000 in 1979 to over 20,709,400 in 2017. This growth has raised environmental concerns, including urban sprawl, diminishing natural resources, reduced forest cover, and pollution. The city's economic progress, while significant, requires a critical examination of its environmental impact and sustainable resource management.

2.2. Datasets

In this study, Landsat 8 images from the 8th and 15th of November 2014 were obtained from the United States Geological Survey (USGS) website (<http://earthexplorer.usgs.gov/>, accessed on 20 February 2018) and served as the primary data source for tasks such as shadow removal, spectral unmixing analysis, model development for estimating vegetation density. Additionally, high-spatial-resolution images from Pleiades 1A and 1B, dated on the 17th, 19th, and 23rd of November 2014, were acquired for land use and land cover (LULC) classification and validation of spectral unmixing analysis. Table 1 provides details on the spectral and spatial resolution characteristics of Landsat 8 datasets and Pleiades-1A and Pleiades-1B datasets.

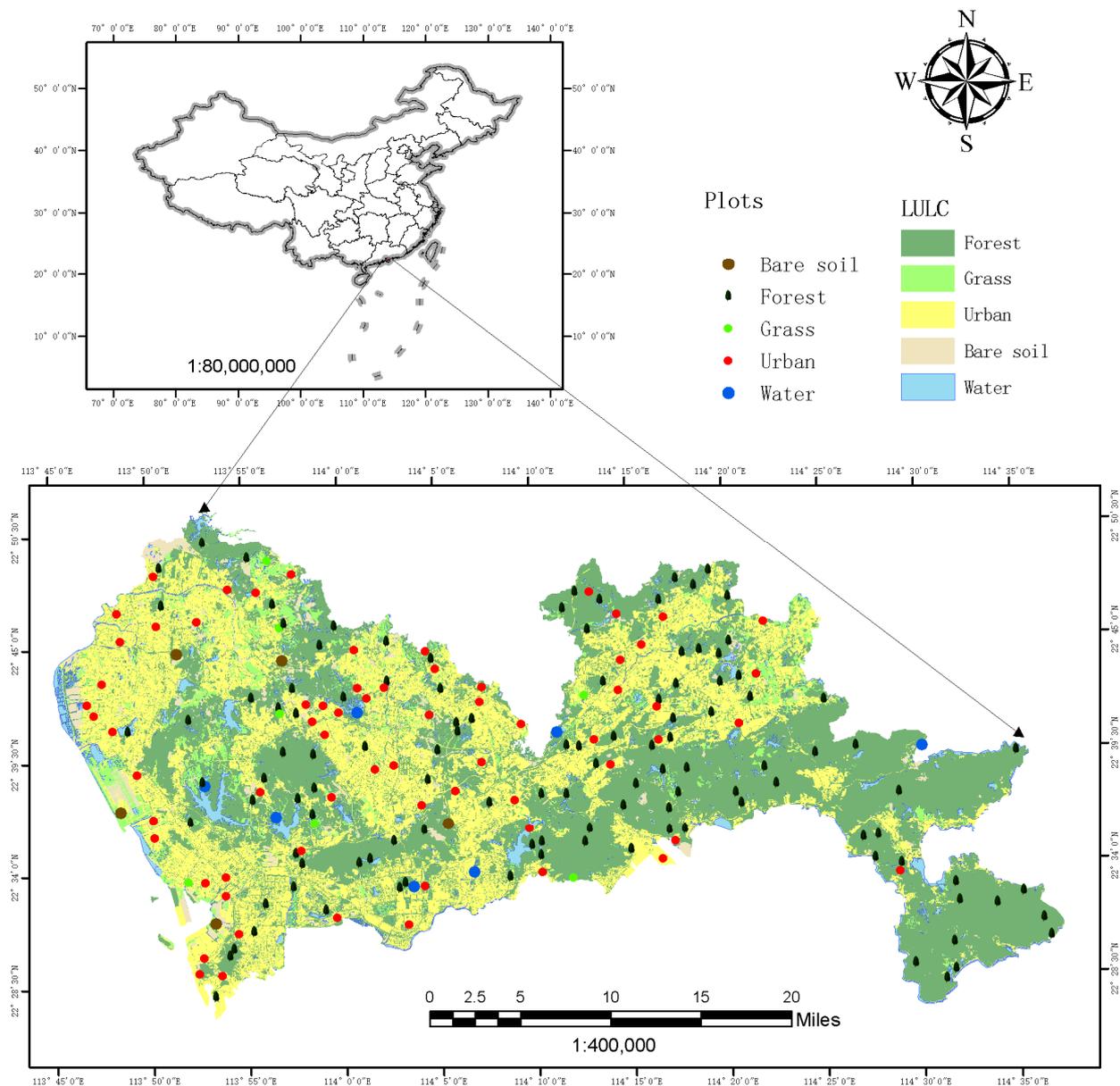
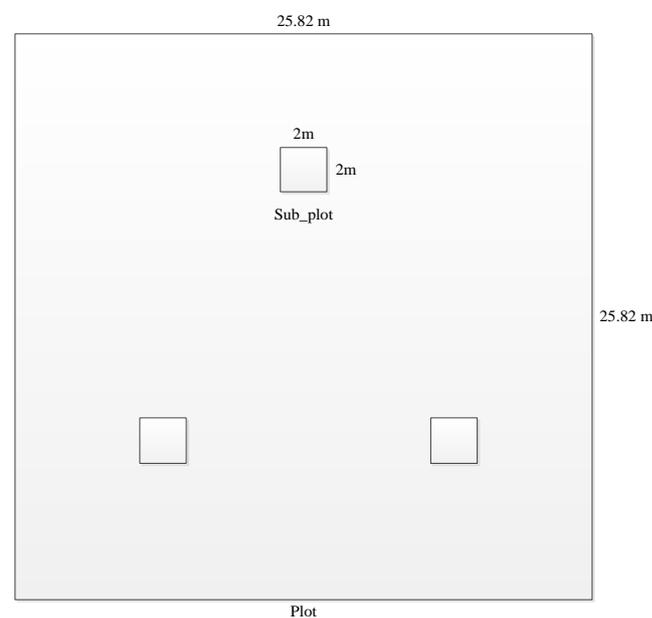


Figure 1. The study area, Shenzhen city, and the spatial arrangement of sample plots within land use and land cover (LULC) categories were determined through the implementation of stratified random sampling to establish the plot sets.

The field survey data utilized in this study were gathered between August 2014 and December 2015, aligning with the period of Landsat 8 image acquisition. A total of 188 plots were measured during this timeframe. The sample plot locations were determined using a global positioning system (GPS) device with a positional error of ± 5 m. Following the sample plot design specified for the National Forest Inventory in China (Chinese Ministry of Forestry, 1996), variables related to grass, shrub, tree, and stand characteristics were measured. Within each plot (Figure 2), tree attributes, including diameter at breast height (DBH), height, and species, were recorded. For subplots with dimensions of $2\text{ m} \times 2\text{ m}$, which contained shrubs and grass, measurements included shrub coverage percentage, ground diameter, height, stock number, and species, as well as grass coverage percentage, species, and height.

Table 1. The band information for Landsat 8, Pleiades-1A, and Pleiades-1B.

Sensor	Band	Range (μm)	Region	Resolution
Landsat 8	Band1	0.433–0.453	Coastal / Aerosol	30 m
	Band2	0.450–0.515	Blue	30 m
	Band3	0.525–0.600	Green	30 m
	Band4	0.630–0.680	Red	30 m
	Band5	0.845–0.885	Near Infrared	30 m
	Band6	1.560–1.660	Short Wavelength Infrared	30 m
	Band7	2.100–2.300	Long Wavelength Infrared	30 m
	Band8	0.500–0.680	Panchromatic	15 m
	Band9	1.360–1.390	Long Cirrus	30 m
	Band10	10.30–11.30	Long Wavelength Infrared	100 m
	Band11	11.50–12.50	Very Long Wavelength Infrared	100 m
Pleiades-1A & 1B	Band0	0.430–0.550	Blue	2 m
	Band1	0.490–0.610	Green	2 m
	Band2	0.600–0.720	Red	2 m
	Band3	0.750–0.950	Near Infrared	2 m
	Band4	0.480–0.830	Panchromatic	0.5 m

**Figure 2.** Schematic representation of the field plot data collection process for trees (represented by the large square) and shrubs and grass (depicted by three smaller squares).

3. Methods

3.1. Above-Ground Vegetation Carbon Density Calculation Based on Survey Data

This study employed Pleiades 1A and 1B images for a visual interpretation-based classification of the study area into five LULC types: forests, grasslands, built-up areas, bare lands, and water bodies. The resulting classification guided a stratified random sampling design across the entire study area, ensuring sample sizes proportional to each class's area and random plot locations within each type. For each plot, the biomass values of

trees, shrubs, and grass were individually calculated and subsequently converted into carbon. The total carbon for each plot was determined by summing the biomass values of trees, shrubs, and grass. Subsequently, the plot carbon density was derived based on the respective plot area. The estimation of tree biomass utilized the Tree Volume Calculation Equation, considering both DBH and height measurements for various tree species in Guangdong province (see Appendix A). The obtained tree volume was then converted into biomass and carbon stock pools using conversion coefficients and empirical equations specified for each tree species (NCFSEM, 2010) as per Equation (1) and Equation (2). Finally, the resulting forest carbon stocks were divided by the plot area to obtain forest carbon density (Mg/ha).

$$\text{BEF} = a + b/M \quad (1)$$

$$B = \text{BEF} * M \quad (2)$$

where BEF represents the biomass expansion factor, which varies across tree species, growing locations, and tree ages. Coefficients 'a' and 'b' are derived from the Biomass and Volume Relationship Parameter Values Table (refer to Appendix B). 'M' represents the volume stock per hectare, and 'B' denotes biomass. The process of converting biomass to carbon is grounded in the Carbon ratio table for different tree species in China (refer to Appendix C).

Regarding shrub and grass carbon estimation, there is a limited number of models or equations available. Fan (2011) proposed two equations, employing a remotely sensed estimation model for estimating shrub and grass biomass. These include the shrub equation Equation (3) and the grass equation Equation (4):

$$\text{Shrub biomass} = 0.0398 \times h_1 - 0.3326 \quad (3)$$

$$\text{Grass biomass} = 0.0175 \times h_2 - 0.2888 \quad (4)$$

3.2. Image Pre-Processing and De-Shadow

Image pre-processing, involving radiometric and geometric correction, is essential for preparing satellite images for tasks such as LULC classification, spectral unmixing analysis, and model development. Radiometric correction involves standardizing the pixel values of remotely sensed imagery, while geometric correction aligns the imagery to a precise spatial location. For radiometric correction, the Landsat 8 data, initially obtained, underwent calibration to convert digital numbers to at-sensor radiance, ensuring consistency and accuracy in the radiometric values. Additionally, atmospheric correction procedures were applied to mitigate atmospheric interference, enhancing the reliability of spectral information. Geometric correction was conducted using ground control points and Digital Elevation Model (DEM) data to rectify spatial distortions caused by terrain variations. This process ensured accurate alignment of the imagery with the Earth's surface, minimizing geometric errors. These preprocessing steps were vital in mitigating distortions, standardizing radiometric values, and enhancing the overall reliability of the dataset for subsequent analysis. For Landsat 8, Level 2 data was acquired from the USGS service. Regarding Pleiades 1A and 1B images in this study, the process involved initial radiometric calibration using ATCOR 9.5, and the conversion of pixel digital number (DN) values to spectral reflectance was carried out using the Model Maker of ERDAS IMAGINE 2023. Following this, geometric calibration was performed using topographic maps, ensuring a root mean square error (RMSE) less than one pixel, to minimize location errors for Pleiades 1A and 1B images.

The accurate mapping of urban vegetation carbon density is hindered by shadows cast by mountains and tall buildings in urban areas. Therefore, prior to utilizing Landsat 8 images for model development, a shadow removal process was undertaken. Most existing shadow removal approaches operate on the assumption that variations in illumination from different materials can be discerned based on their spectra. In these methods, the

spectral value of a pixel within a shadow, denoted as “y”, is treated as a linear combination of endmember spectra from fully illuminated surfaces, e_i :

$$y = \sum_i w_{iy} e_i \quad (5)$$

where w_{iy} is the i th endmember weight for the pixel, then assumed that a pure shadow pixel has the reflectance value of 0, the shadow fraction of a mixed pixel can be calculated using equation Equation (6):

$$f_y = \sum_i w_{ia} - \sum_i w_{iy} \quad (6)$$

where w_{iy} is the i th endmember weight for pixel y , the weight can be extracted from each spectrum by taking its dot product with a filter vector that is orthogonal to all endmembers within the pixel, a is a fully illuminated pixel with weights w_{ia} summing to 1, f_y is the shadow fraction of pixel y . This developed a filter used as a mask which could be applied to an image to estimate its shadow fraction (f_y) for each pixel.

In constrained linear spectral unmixing analysis (LSUA), the i th endmember weight could be extracted from each band with a filter vector v_i that is orthogonal to all endmembers; based on this, Equation (7) could be rewritten as:

$$f_y = -\sum_i v_i^T (y - a) = g^T (y - a) \quad (7)$$

where g is a vector used as shadow filter. If the shadow on a scene is rare, the mean of scene spectrum can be used for a fully illuminated pixel (a). And the matched filter could be defined as Equation (8):

$$q = C^{-1}(t - a) / [(t - a)^T C^{-1}(t - a)] \quad (8)$$

where q is matched filter, C is covariance matrix, t is the target spectrum. When $t = 0$, the shadow matched filter could be expressed as Equation (9):

$$q_{\text{shadow}} = -C^{-1}a / (a^T C^{-1}a) \quad (9)$$

Application of the matched filter to the Landsat 8 image yields an estimate of pixel-level shadow fraction image (f). Then, the result was rebalanced to simulate illumination by a spectrally uniform source using Equation (10):

$$F(\lambda) = \frac{f(d(\lambda) + s(\lambda))}{fd(\lambda) + s(\lambda)} = f(1 + \frac{s(\lambda)}{d(\lambda)}) / (f + \frac{s(\lambda)}{d(\lambda)}) \quad (10)$$

where $F(\lambda)$ is rebalancing result, $d(\lambda)$ is spectrum of direct sun illumination, $s(\lambda)$ is spectral of sky illumination. After rebalancing, the de-shadow spectrum could be calculated by Equation (11):

$$I = F(\lambda) / (1 - f) \quad (11)$$

In this study, the pure shadow pixels were identified on the Landsat 8 image when using the mosaicked Pleiades 1A and 1B image as a reference.

3.3. Spectral Unmixing Analysis

Before linear spectral unmixing analysis (LSUA) was conducted, the Landsat 8 image underwent Minimum Noise Fraction (MNF) transformation to extract noise. Subsequently, a Pixel Purity Index (PPI) was computed using MNF, projecting each pixel onto random vectors in the reflectance space. Pixels scored based on their positions in the projection plot, with the highest scores representing pure pixels. These PPIs were associated with the original image to identify LULC types. N-Dimensional visualization validated endmember purity, eliminating non-corner endmembers for a refined selection.

For spectral unmixing analysis, there is a linear tool in ENVI 4.6, but it is half constrained with the results of unreasonable negative pixel values. In this study, a fully constrained model which solved the negative pixels was developed in Equations (12) and (13) and it guaranteed that the fraction coefficients were positive and their summation was equal to one.

$$\forall i : a_i \geq 0 \tag{12}$$

$$\sum_{i=1}^m a_i = 1 \tag{13}$$

where a_i represents the fraction of each endmember in pixel x .

3.4. Modeling

To model vegetation carbon density, 648 spectral variables were derived from the Landsat 8 image (Table 2), encompassing original bands, diverse transformations, band ratios, PCA-generated bands, vegetation indices, difference vegetation indices, and texture measures. Pearson product-moment correlation coefficients were calculated between these variables and vegetation carbon density to assess their significance, utilizing a significance level of 0.05 based on the student’s distribution.

Table 2. Spectral variables (SVs) obtained from the ten bands of Landsat 8.

Spectral Variables	Definitions of Spectral Variables	# of SV
Original band	Band 1 (Coastal Aerosol), Band 2 (Blue), Band 3 (Green—GRN), Band 4 (Red), Band 5 (Near Infrared—NIR), Band 6 (Shortwave Infrared 1—SWIR1), Band 7 (Shortwave Infrared 2—SWIR2), Band 8 (Cirrus), Band 9 (Long Wavelength), and Band 10 (Long Wavelength)	10
Inversions of bands	$IB_i = \frac{1}{band_i}, i = 1, \dots, 10$	10
Simple two-band ratios	$SR_{i,j} = \frac{Band_i}{Band_j}, i, j = 1, \dots, 10; i \neq j$	90
Three-band ratios	$TR_{i,j,k} = \frac{Band_i}{Band_j + Band_k}, i, j, k = 1, \dots, 8; j \neq k$	359
Difference vegetation indices	$DVI_{i,j} = Band_i - Band_j, i, j = 1, \dots, 10; i \neq j$	45
Shortwave infrared-visible band ratio	$SVR = SWIR1 / \left[\frac{RED + GRN}{2} \right]$	1
Normalized difference vegetation index	$NDVI = \frac{NIR - RED}{NIR + RED}$	1
Modified normalized difference vegetation index	$MNDVI = \frac{NIR - RED}{NIR + RED} \left(1 - \frac{SWIR1 - SWIR1_{min}}{SWIR1_{max} - SWIR1_{min}} \right)$	1
Red-green vegetation index	$GRVI = (RED - GRN) / (RED + GRN)$	1
Reduced simple ratio	$RSR = \frac{NIR}{RED} \left(1 - \left(\frac{SWIR1 - SWIR1_{min}}{SWIR1_{max} - SWIR1_{min}} \right) \right)$	1
Soil adjusted vegetation index	$SAVI_l = \frac{(NIR - RED)(1+l)}{NIR + RED + l}, l = 0.1, 0.25, 0.3, 0.5$	4
Atmospherically resistant vegetation index	$ARVI = [NIR - (2 \times RED - BLUE)] / [NIR + (2 \times RED - BLUE)]$	1
Enhanced vegetation index	$EVI = 2.5(NIR - RED) / (NIR + 6RED - 7BLUE + 1)$	1
Principal component analysis	The first 3 PCs from Principal component analysis (PCA)	3
Texture measures	Texture measures derived from the Grey-Level Co-occurrence Matrix, encompassing mean, angular second moment, contrast, correlation, dissimilarity, entropy, homogeneity, and variance.	80

Spatial autocorrelation analysis revealed significant clustering of plot carbon density values across the study area, indicating varying relationships between vegetation carbon values and original Landsat 8 spectral bands in different locations. This variability and nonlinearity are evident in scatter plots shown in Figure 3.

Considering the intricate urban landscape and the nonlinear relationship between vegetation carbon density and Landsat 8 images’ original bands (Figure 3), this study employed two global parametric spatial interpolation models—Linear Stepwise Regression (LSR) and Logistical Model based Stepwise Regression (LMSR) [50–54]—along with three local non-parametric models—k-Nearest Neighbors (kNN) [55], Decision Trees (DT) [56], and Random Forests (RF) [57]. These models were utilized to map vegetation carbon density, incorporating spectral variables and vegetation fraction derived from LSUA, in comparison to the complex urban context.

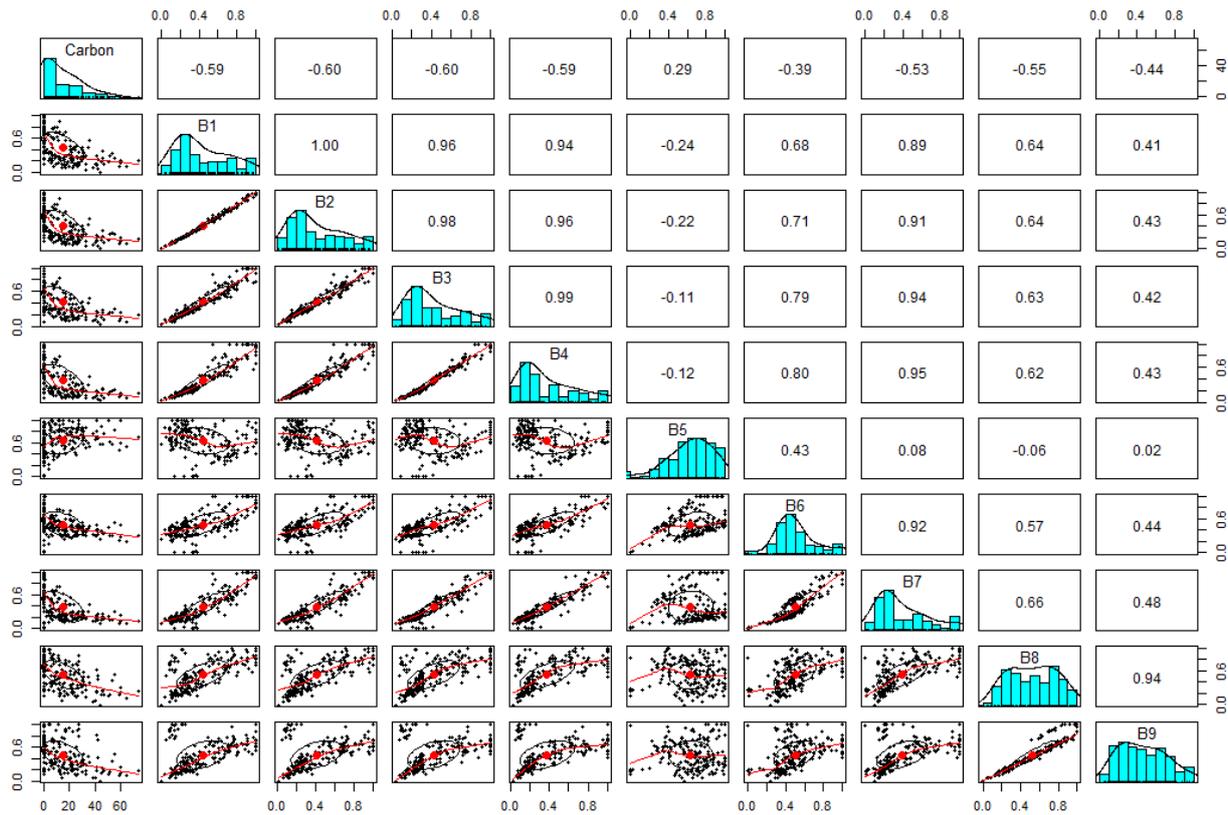


Figure 3. Correlation coefficient analysis examining the relationship between vegetation carbon density and the Landsat 8 images' original bands. The symbols in the figure denoted as Carbon, B1, B2, ..., B9 correspond to vegetation carbon density, Landsat Band 1, Band 2, ..., Band 9, respectively.

3.4.1. Linear Stepwise Regression Model

The LSR model, commonly applied in mapping forest biomass and carbon density, was employed in this study to identify statistically significant spectral variables that enhance model fit and reduce sum of squared errors. Additionally, it assessed collinearity among spectral variables from Landsat 8 imagery, and the chosen significant independent variables were subsequently utilized in the backward-based linear stepwise regression, Equation (14)

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \dots + \beta_n X_n \quad (14)$$

where y is the vegetation carbon density, β_i is the i th coefficient and X_i is the i th spectral variable derived from the Landsat 8 image.

3.4.2. Logistical Model Based Stepwise Regression Model

The LMSR, a probabilistic statistical prediction model handling binary dependent variables, was employed in this study with a dependent variable ranging from 0 to 1. Standardizing the plot vegetation carbon density values to this range, LMSR, coupled with stepwise regression, identified significant spectral variables. Multicollinearity analysis, utilizing the variance inflation factor (VIF), flagged highly correlated variables (coefficients ranging from 0.03 to 0.99). Variables with VIF exceeding 10 were considered indicative of severe multicollinearity. The stepwise logistic regression, conducted in R statistical software, version 3.5.3, employed the LMSR model, denoted by Equation (15).

$$P = \frac{e^{b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n}}{1 + e^{b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n}} \quad (15)$$

where P represents standardized plot vegetation carbon density values, e is the natural logarithm base, b_0 is the interception at y -axis and b_i is the coefficient of the i th independent variable x_i , x_i is the i th significant spectral variable derived from the Landsat 8 image.

3.4.3. k Nearest Neighbors

The kNN is a simple, intuitive, and nonparametric method in statistical discrimination, used for classification or regression. It predicts unknown attributes based on the observed learning set and relies on distance measures. Formally, the model can be described as follows:

Let $L = \{(y_i, x_i), i = 1, 2, 3, \dots, n_L\}$ be observed data and used as training dataset, y_i denotes class membership, and x_i represents the predictor variables. For a new observation (y, x) , the nearest neighbor $(y_{(1)}, x_{(1)})$ is determined by an arbitrary distance function as Equation (16):

$$d(x, x_{(1)}) = \min_i(d(x, x_i)) \quad (16)$$

And $\hat{y} = y_{(1)}$, the nearest neighbor is selected as prediction for y . Classically, distance functions are Euclidean distance or absolute distance.

The idea of using multiple closest observations within the learning set was extended, leading to the k-nearest neighbors method (kNN). Users can set the number of nearest neighbors, k , and consider distance weight in the model. The kNN model assumes that closer neighbors have higher influence on the decision. However, in this study, all k nearest neighbors were assumed to have equal influence. Before searching for the nearest neighbors, similarity measures needed to be standardized for use as weights.

The kNN model is widely used for predicting forest attributes, biomass, and carbon density. It estimates the values of an interest variable based on the similarity of predictor variables with k nearest neighbors or selected plots. In this research, the urban vegetation carbon density at each location was estimated by weighting the carbon density values of the nearest plots using the inverses of Euclidean distances. The variable distance was weighted by triangular, rectangular, Epanechnikov, Gaussian, rank, and optimal kernel functions.

Rectangular k

$$\text{ernel} : \frac{1}{2} * I(|d| \leq 1)$$

$$\text{Triangular kernel} : |(1 - (d)) * I(|d| \leq 1)$$

$$\text{Epanechnikov kernel} : \frac{3}{4}(1 - d^2) * I(|d| \leq 1)$$

$$\text{Gaussian kernel} : \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{d^2}{2}\right)$$

$$\text{Rank kernel} : 1/d$$

$$\text{Optimal kernel} : \frac{35}{32}(1 - d^2)^3 * I(|d| \leq 1)$$

where d is distance, $I(\cdot)$ is indicator function: if defined condition in brackets is true, $I(\cdot) = 1$ and otherwise, $I(\cdot) = 0$. The window width of kernel function was determined by a certain distance from maximum value. The Euclidean distances were standardized by dividing itself using the closest neighbor ($k + 1$) that was not used for predication.

The used significant spectral variables were derived from both LSR and LMSR. As a first step, the spectral variables were standardized by dividing themselves using their standard deviation.

3.4.4. Decision Trees

The DT, introduced by Breiman (1984), is employed for classification or regression predictive analysis. Also known as “decision trees” or CART in some contexts, this algorithm forms the basis for important algorithms like Random Forests [58]. Constructing a DT involves selecting input variables and split points using a greedy algorithm to minimize a cost function. A predefined stopping criterion is essential to avoid an infinite model run.

In regression predictive modeling, the cost function minimization determines split points using the sum squared error calculated from training samples, Equation (17)

$$SQE = \text{sum}(y - \hat{y})^2 \quad (17)$$

where SQE is sum squared error, y is output of the training sample, and \hat{y} is predicted output.

The stopping criterion is crucial in DT, commonly set as a minimum count based on the number of training instances for each leaf node. When the count falls below this minimum, the split stops, and the node becomes a final leaf node. This criterion significantly impacts DT performance. The complexity of DT is linked to the number of splits; simpler trees are preferred to prevent overfitting. Pruning the tree to minimize cross-validation error is necessary for avoiding overfitting issues.

3.4.5. Random Forests

RF was developed by Breiman (2001), enhances categorical variable-based classification and continuous variable-based regression trees (CART) by amalgamating multiple sets of decision trees [59]. As an effective ensemble machine learning model, RF is adept at classification or regression predictive analytics, employing an additive model to combine decisions from a sequence of base models, Equation (18)

$$g(x) = f_0(x) + f_1(x) + f_2(x) + f_3(x) + \dots \quad (18)$$

where $g(x)$ is the sum of simple based models f_i , in this study, each base classifier is a simple decision tree used for regression prediction. For Random Forest regression, all the base models are trained independently using different subsamples of observations. Each tree node splitting is based on a deterministic algorithm by randomly selecting a sub-set of variables and a sample from the training data [59]. In this study, RF begins with randomly drawing many bootstrap sub-samples with replacement from the field plot observations. A regression tree was constructed for each sub-sample. For nodes of each tree, a small part of input variables was selected from the total inputs used for binary partition. For the tree splitting criterion, it was based on the lowest Gini Index (Equation (19)) of the chosen input variable.

$$I_g = 1 - \sum_{j=1}^m f(t_{X(x_i)}, j)^2 \quad (19)$$

where, $f(t_{X(x_i)}, j)$ is the section of samples which the value x_i belongs to the leave j as node t . The predicted carbon density at a location without observance was calculated by averaging the bootstrap selected sub-samples constructing trees.

In the construction of the RF, two parameters must be optimized: the number of trees (ntree) and the optimal minimal size of the terminal nodes of the trees. The optimal numbers of trees and nodes for predicting vegetation carbon density were determined based on the root mean square error (RMSE) of calibration.

3.5. Accuracy Assessment

In comparison, all models employed were combined with LSUA to validate our hypothesis: whether the addition of the vegetation fraction image from LSUA improves the models' performance in predicting carbon density. The accuracy assessment of vegetation carbon density estimates utilized a cross-validation method. In this procedure, a random sample of plot vegetation carbon density was first removed from the field plots, with the remaining plots utilized to train the models. The estimate of vegetation carbon density for the removed sample was calculated, and the deviation between the estimated and observed values was obtained. Another sample was then randomly removed, and the previously removed one was reintroduced into the dataset. The corresponding modeling and estimation were conducted for this sample. This process was repeated until all samples were estimated. Based on the predictions and their comparison with field measurements,

the coefficient of determination R^2 and root mean square error (RMSE) were derived to assess the goodness-of-fit and prediction performance of the models.

4. Results

4.1. Statistics of Field Data

Table 3 presents a statistical summary of the observed plot data employed in mapping urban vegetation carbon density in Shenzhen. The sample mean and standard deviation at the plot level were estimated using simple random sampling estimators. The plot vegetation carbon density values exhibited a substantial coefficient of variation, with a confidence interval ranging from 12.66 Mg/ha to 17.32 Mg/ha at a significance level of 0.05.

Table 3. Statistical summary of sample plot data utilized for urban vegetation carbon density mapping.

Number of Plots	Minimum (Mg/ha)	Maximum (Mg/ha)	Sample Mean (Mg/ha)	Standard Deviation (Mg/ha)	Coefficient of Variation (%)
188	0	73.550	14.99	16.3	108.87

4.2. Correlation of Vegetation Carbon Density with Spectral Variables

The Pearson product–moment correlation coefficients between observed plot vegetation carbon density and 648 spectral variables were calculated. Prior to shadow removal, 523 spectral variables exhibited coefficients ranging from 0.142 to 0.667, significantly different from zero at a 0.05 significance level. After shadow removal, the number of significant variables increased to 534, with coefficients ranging from 0.146 to 0.688. Shadow removal positively impacted the correlation coefficient between field observations and spectral variables. The band-ratio TR536 showed the highest correlation with vegetation carbon density both before and after shadow removal of Landsat 8 imagery. Following shadow removal, Table 4 lists the original Landsat 8 image bands and 45 other spectral variables with the highest correlations to plot carbon density observations. In comparison to vegetation indices and band ratios, PCAs and matrix-based texture variables exhibited smaller correlation coefficients.

Table 4. Pearson correlation coefficients (r) were calculated between field-measured carbon density and spectral variables (top 45 highest correlated variables and 9 original Landsat bands) after shadow removal ($n = 188$; veg-fraction: vegetation fraction obtained using LSUA).

Spectral Variables	Correlation		Spectral Variables	Correlation	
	r	P		r	P
B1	−0.593	0	TR415	−0.661	0
B2	−0.596	0	TR416	−0.608	0
B3	−0.597	0	TR425	−0.660	0
B4	−0.586	0	TR426	−0.596	0
B5	0.293	4.44×10^{-5}	TR435	−0.650	0
B6	−0.394	2.23×10^{-8}	TR436	−0.574	0
B7	−0.529	5.77×10^{-15}	TR458	−0.580	0
B9	−0.554	0	TR459	−0.570	0
B10	−0.435	4.42×10^{-10}	TR516	0.658	0
DVI56	0.642	0	TR517	0.612	0
DVI57	0.617	0	TR526	0.669	0
ARVI	0.626	0	TR527	0.631	0
MNDVI	0.630	0	TR534	0.581	0
SAVI0.1	0.631	0	TR536	0.688	0
SAVI0.25	0.629	0	TR537	0.660	0
SAVI0.5	0.627	0	TR546	0.685	0

Table 4. Cont.

Spectral Variables	Correlation		Spectral Variables	Correlation	
	r	P		r	P
SR57	0.686	0	TR547	0.654	0
SR67	0.639	0	TR567	0.684	0
TR125	−0.584	0	TR637	0.583	0
TR135	−0.543	8.88×10^{-16}	TR647	0.592	0
TR215	−0.621	0	TR715	−0.531	4.88×10^{-15}
TR235	−0.569	0	TR725	−0.531	4.44×10^{-15}
TR258	−0.511	6.33×10^{-14}	TR735	−0.527	8.44×10^{-15}
TR315	−0.650	0	TR745	−0.526	8.88×10^{-15}
TR325	−0.641	0	TR758	−0.510	7.82×10^{-14}
TR345	−0.561	0	TR759	−0.524	1.24×10^{-14}
TR358	−0.516	3.46×10^{-14}	Veg_fraction	0.595	0

4.3. Spectral Unmixing Analysis

For spectral unmixing analysis, two endmember selection methods were compared: spectral characteristics-based automatic selection and operator's knowledge-based manual selection. Additionally, three sets of endmembers (2, 3, and 4) were evaluated: vegetation, urban; vegetation, urban, water; and vegetation, urban, water, bare soil. Results demonstrated that, after shadow removal, the 4-endmember configuration, whether automatically or manually selected, produced the highest correlation coefficient of 0.595 between the vegetation fractional images and field plot carbon density (Table 5). The correlation coefficient increased with the number of endmembers, reaching a peak at 4 endmembers, after which it began to decrease.

Table 5. Summary statistics of correlation coefficients between vegetation fractional images and field-measured carbon density for different endmember configurations (2 endmembers: vegetation, urban; 3 endmembers: vegetation, urban, water; and 4 endmembers: vegetation, urban, water, bare soil) both before and after shadow removal.

Method	2-Endmember	3-Endmember	4-Endmember
Automatic selection (Before)	0.491	0.554	0.589
Manual selection (Before)	0.492	0.555	0.59
Automatic selection (After)	0.495	0.563	0.595
Manual selection (After)	0.498	0.564	0.595

Given the superior performance of the automatic endmember selection method over manual selection, the study employed the results from decomposing mixed pixels using 4 endmembers with automatic pure pixel selection for model development (Figure 4).

Figure 4 illustrates the predominant distribution of vegetation in Shenzhen city's southeast, southwest central, northeast, and northwest regions, where urban fraction estimates were relatively smaller. The linear model effectively identified water bodies, discerning their vegetation and urban fractions approaching zero. This model exhibited a notable correlation coefficient of 0.595 with plot vegetation carbon density.

Validation of the vegetation fraction image was conducted using a visual interpretation map of Land Use and Land Cover (LULC) derived from Pleiades 1A and 1B images with a spatial resolution of 0.5 m. When compared with the vegetation fraction image (Figure 4a), spectral unmixing analysis successfully extracted detailed vegetation cover information. The fully constrained linear spectral unmixing analysis accurately estimated both the spatial pattern and specific coverage rate of urban vegetation. The resulting vegetation cover percentage for the entire study area was 44.2% based on Pleiades image visual interpretation and 41.7% for the fully constrained linear spectral unmixing analysis using Landsat 8 imagery. The pixel-based Root Mean Square Error (RMSE) was 0.16.

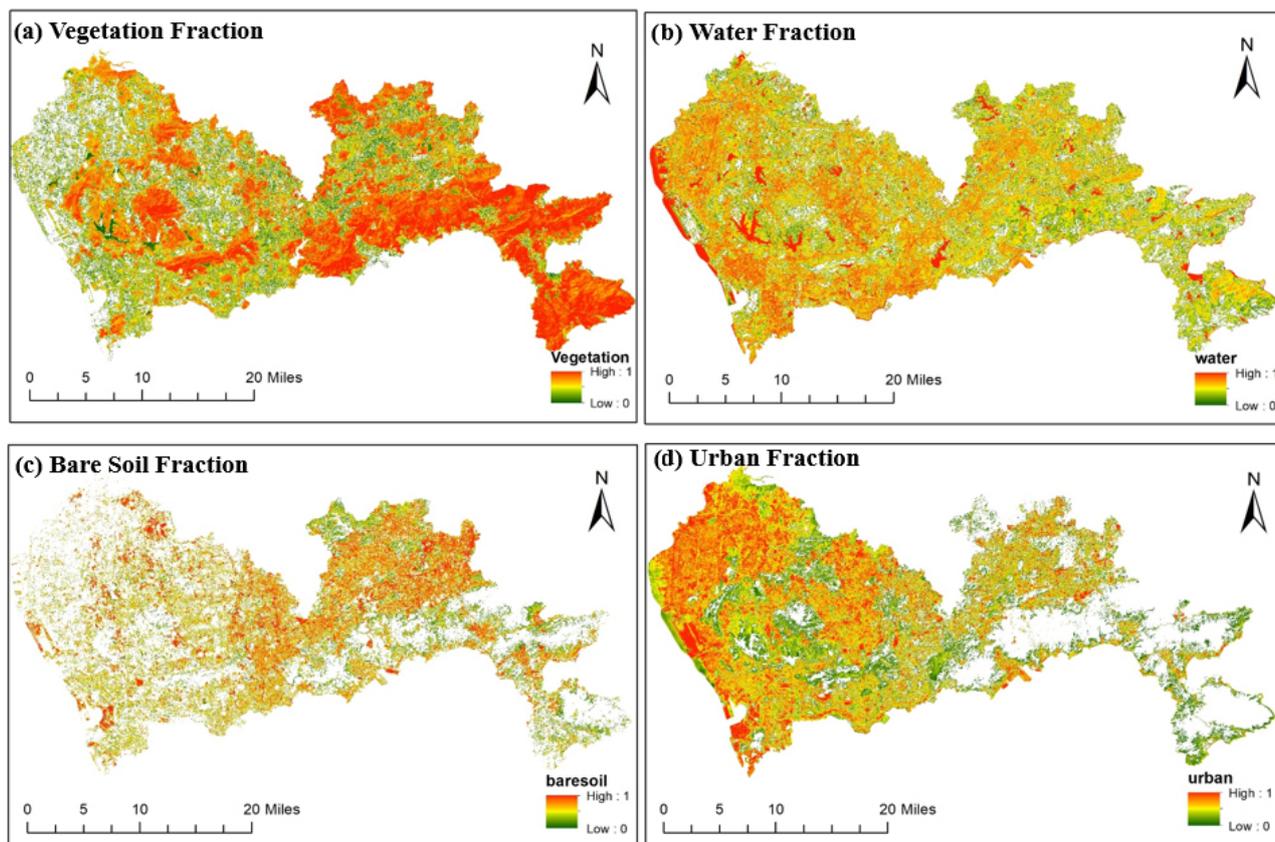


Figure 4. The fractional images were acquired through a mathematical selection method for decomposing mixed pixels utilizing 4 endmembers.

4.4. De-Shadow Results of Landsat 8 Image

Shadow removal, facilitated by LSUA, involved selecting pure endmembers for shadow, vegetation, urban, water, and bare soil in the spectral unmixing analysis. A shadow fraction image served as a mask layer for shadow removal based on each pixel's shadow contribution. Figure 5 displays the image before shadow removal, revealing identifiable mountain shadows. Given the Landsat 8 image's $30\text{ m} \times 30\text{ m}$ spatial resolution, building shadows were challenging to discern, making it difficult to assess the shadow removal effect. The accuracy assessment utilized 300 random sample points, categorizing them as poor, average, or good based on their visual representation of original land cover types compared with the pre-shadow removal Landsat 8 image, with assistance from Pleiades 1A and 1B images as a reference.

Among the 300 pixels, 29 were located in mountain-shadowed areas, where shadows were successfully removed or alleviated, leading to a significant recovery of LULC information. However, in urban areas, only a few pixels could be clearly identified as shadows with the assistance of Pleiades 1A and 1B. The method performed well in shadow removal in mountainous areas but faced challenges in urban areas due to coarse resolution.

Comparing Figure 5 results with Pleiades 1A and 1B as a reference, mountain shadows were effectively removed, while shadows induced by buildings showed a fair improvement in image quality. Table 6 presents a correlation coefficient comparison between plot vegetation carbon density and spectral variables using pre- and post-shadow-removed Landsat 8 images. After shadow removal, correlation coefficients increased by 1.28% to 2.59% across different bands. Given the positive impact of the shadow removal algorithm on all bands, all models were constructed using the post-shadow-removed Landsat 8 image.

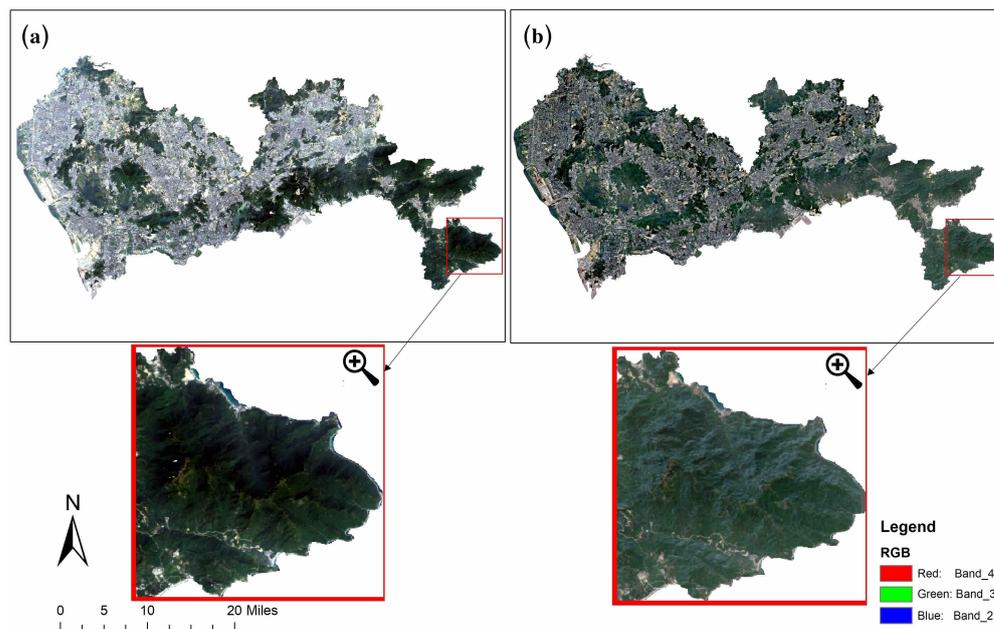


Figure 5. Comparison between pre- and post-shadow removal of Landsat 8: (a) Landsat 8 image before shadow removal displayed in natural color; (b) Landsat 8 image after shadow removal displayed in natural color. The red highlighted area is enlarged to illustrate the differences before and after shadow removal.

Table 6. Correlation coefficients between vegetation carbon density and spectral variables using Landsat 8 images before and after shadow removal.

Landsat 8.	B1	B2	B3	B4	B5	B6	B7	B9	B10
Before	−0.578	−0.581	−0.587	−0.571	0.283	−0.389	−0.518	−0.546	−0.423
After	−0.593	−0.596	−0.597	−0.586	0.294	−0.394	−0.529	−0.554	−0.435

4.5. Vegetation Carbon Density Mapping

The performance of LSR models with and without the vegetation fraction image was compared (Figure 6a,b). Four stepwise-selected spectral variables (excluding the vegetation fraction variable) were used as independent variables to fit the LSR model for estimating vegetation carbon density (Equation (20)). The variable TR536, which exhibited the highest correlation with plot vegetation carbon density, was excluded by stepwise regression due to its high collinearity with other selected variables.

$$\hat{Y} = 24.976 - 14.968 * B9 + 44.513 * TR567 - 25.550 * SR67 - 3.753 * B2_mean \quad (20)$$

Vegetation carbon density estimates using Equation (20) ranged from −77.283 Mg/ha to 454.69 Mg/ha, with a mean estimate of 15.332 Mg/ha, slightly overestimating the observed mean of 14.999 Mg/ha. Leave-one-out cross-validation resulted in an R² of 0.5451 and RMSE of 10.852 Mg/ha (Table 7).

Upon integrating the vegetation fractional image into the LSR model, the selected variables changed, and the model was expressed as Equation (21). This integration led to a slight improvement, with R² increasing to 0.5453, RMSE decreasing to 10.812, and the mean estimation closer to field measurement compared to the LSR model alone (Table 7). Figure 6 illustrates the enhanced estimation of sub-pixel carbon density in the east urban areas along the roads when combining LSR and LSUA, reducing the absolute values of negative estimates from LSR.

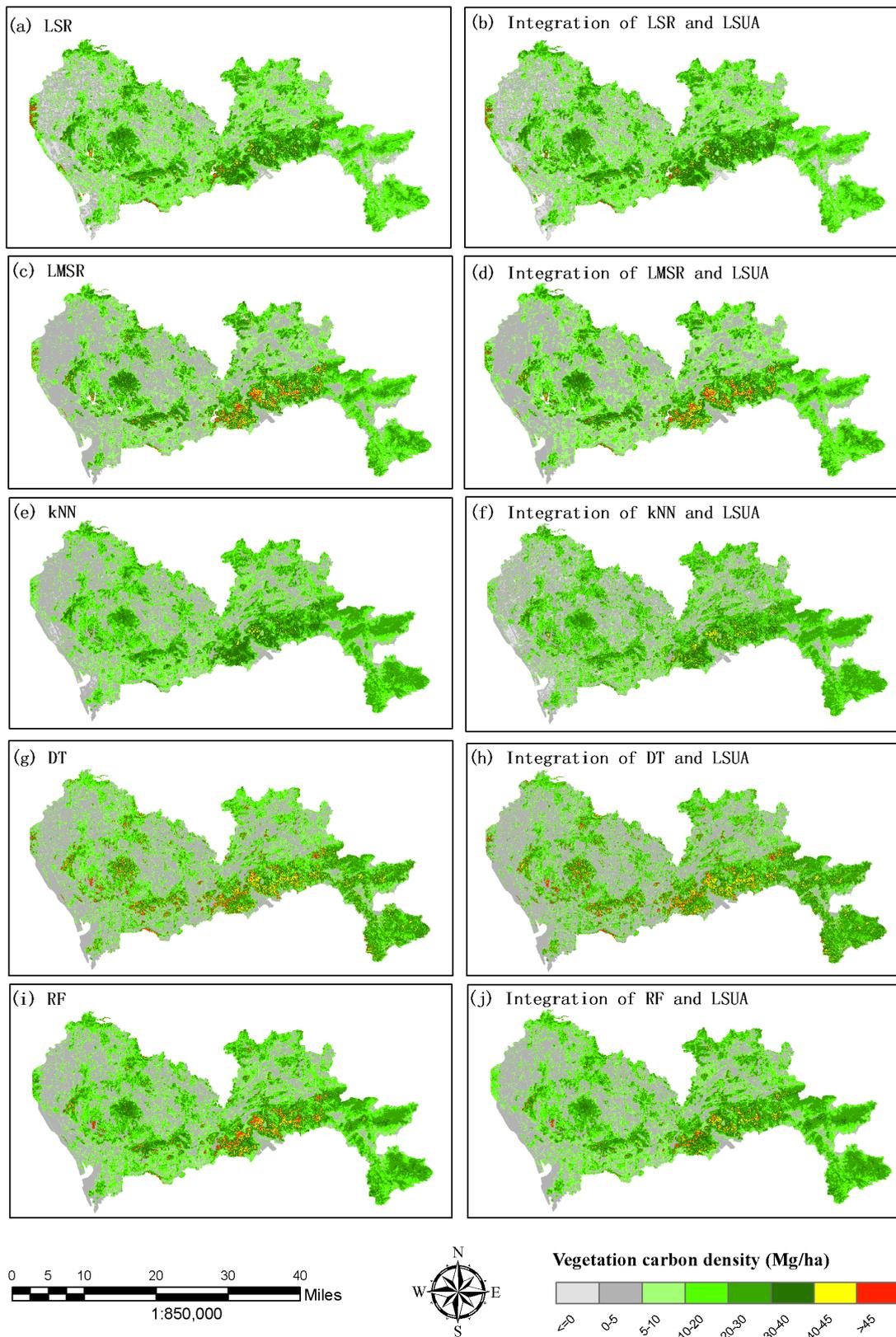


Figure 6. Comparison of the mapping outcomes for vegetation carbon density utilizing LSR, LMSR, kNN, DT, and RF models with and without the incorporation of the vegetation fraction variable derived from LSUA in the modeling process.

Table 7. The accuracy assessment of vegetation carbon density estimates from LSR, LMSR, kNN, DT, RF models and these models integrated with LSUA. In the table, Mean indicates observed and model prediction mean of vegetation carbon density, R^2 is the coefficient of determination and RMSE is mean square error. $\hat{\mu}$ is map mean estimate and Var_{map} is the variance of the map based on model-assisted regression estimators.

Approach	Mean	R^2	RMSE	$\hat{\mu}$ (Mg/ha)	Var_{map}
Observed	14.99	-	-	-	-
LSR	15.07	0.5451	10.852	15.332	1.68
LSR integrated with LSUA	15.05	0.5453	10.812	15.26	1.61
LMSR	14.91	0.5621	9.153	14.091	1.38
LMSR integrated with LSUA	14.94	0.5712	9.046	14.256	1.34
kNN	14.75	0.4620	10.561	14.483	1.89
kNN integrated with LSUA	14.86	0.4641	9.682	14.518	1.77
DT	15.00	0.8171	6.952	14.501	1.26
DT integrated with LSUS	15.00	0.8205	6.888	14.501	1.24
RF	15.33	0.7630	8.741	15.419	1.16
RF integrated with LSUA	15.20	0.7800	8.651	15.136	1.09

$$\hat{Y} = 33.274 - 17.406 * B9 - 14.294 * TR426 + 42.448 * TR567 - 27.295 * SR67 + 2.280 * \text{Veg_fraction} \quad (21)$$

The non-linear LMSR model, utilizing two significant spectral variables (B9 and TR567) selected by stepwise regression (Equation (22)), demonstrated a more reasonable estimate range from 0 Mg/ha to 73.555 Mg/ha compared to the LSR model. The LMSR achieved an R^2 of 0.5621, surpassing both LSR models with and without LSUA integration (Table 7). Additionally, the RMSE was reduced to 9.153 Mg/ha, smaller than those obtained with LSR alone or with LSUA integration.

$$\hat{Y} = \frac{73.555 * \exp(-2.6678 - 1.3674 * B9 + 1.9713 * TR567)}{1 + \exp(-2.6678 - 1.3674 * B9 + 1.9713 * TR567)} \quad (22)$$

When the LMSR model integrated with LSUA (Equation (22)), the coefficient of determination increased to 0.571, surpassing the LMSR model without the inclusion of the vegetation fraction variable. Simultaneously, the RMSE decreased to 9.046 Mg/ha, representing a 1.2% reduction compared to LMSR without the vegetation fraction variable integration (Table 7).

Figure 6c,d illustrates the ability of the LMSR model, with and without the vegetation fraction variable, to capture the spatial patterns of vegetation carbon density across the study area. Areas marked in grey signify low vegetation carbon density, predominantly in developed urban regions. The LMSR model, combined with LSUA, produces a more reasonable sub-pixel vegetation carbon density map compared to the LSR model. In mountainous and urban park areas, carbon density falls within the 20 Mg/ha to 30 Mg/ha range, while values of 30 Mg/ha to 80 Mg/ha are primarily found in low-elevation mountainous areas with favorable soil conditions, lower slopes, and more suitable temperatures than high-elevation counterparts.

For the kNN model, optimal parameters such as the number of nearest neighbors (k), spectral distance parameter, weighting kernel function, and predictor set were determined through iterative kNN imputation and mean square error analysis. The plot vegetation carbon density served as the dependent variable, while the significant variables selected for LSR and LMSR were used as independent variables. For kNN without LSUA, the optimal k was 12, and the best distance weighting kernel function was rectangular, resulting in the smallest mean squared error of 0.0247. With the integration of LSUA, the optimal k remained 10, and the best distance weighting kernel function was also rectangular, resulting in the smallest mean squared error of 0.0244.

For kNN integrated with LSUA, the coefficient of determination was 0.4641, surpassing that of kNN without the vegetation fraction variable. The RMSE decreased to 9.682 Mg/ha, showing an 8.79% reduction compared to kNN without the vegetation fraction variable (Table 7). The map variance decreased by 6.34%, and the estimated sample mean and map mean increased, aligning more closely with the field measurement mean compared to kNN without the vegetation fraction variable. However, both models exhibited overestimation in areas with low values and underestimation in areas with high vegetation carbon density.

Both kNN and kNN with LSUA tended to underestimate vegetation carbon density in areas with values exceeding 30 Mg/ha and overestimate it in areas with values below 20 Mg/ha, Figure 6e,f. Both models effectively captured the overall patterns of vegetation carbon density in the urban area. The integration of kNN with LSUA performed significantly better in urban areas, particularly in extracting subpixel vegetation carbon information for roadside trees.

The DT is commonly employed for classification and regression tasks. Figure 6g,h depict the DT model's estimation of vegetation carbon density. For both the DT and the DT integrated with LSUA, the flowcharts progress downward from the top, starting with predictor TR567, which exhibits the highest correlation with plot vegetation carbon density among all independent variables in the DT model. Based on the criterion of a TR567 value less than 0.8 or not, the 188 plots are divided into two groups: 90 plots with TR567 less than 0.8 and the remaining plots with TR567 not less than 0.8. This process continues until reaching a final leaf node. The number of splits serves as a crucial indicator for evaluating the effectiveness of the DT model.

With field measurements incorporated into the DT model and a minimum split number of six, both the DT and DT integrated with LSUA achieved the highest coefficient of determination (R^2) between observed and predicted vegetation carbon density values. The DT resulted in an R^2 of 0.8171, with plot and map mean estimates of 15.00 Mg/ha and 14.501 Mg/ha, respectively, and an RMSE of 6.952 Mg/ha. Comparatively, the integration of DT with LSUA showed slight improvement, with an R^2 of 0.8205, plot and map mean of 15.00 Mg/ha and 14.501 Mg/ha, respectively, and an RMSE of 6.888 Mg/ha (Table 7).

Both the DT model and the DT integrated with LSUA demonstrated significant predictive capability for urban vegetation carbon density. Given the potential overfitting concern associated with decision trees, the absence of end nodes with n values of 1 or 2 in this study (Figure 6g,h) suggests reliable model performance.

For the optimal number of trees in RF, RMSE showed a decreasing trend with an increasing number of trees (Figure 7, left). Before integrating RF with LSUA, the error stabilized after the number of trees exceeded 500. Upon integration with LSUA, the error stabilized when the number of trees exceeded 800 (Figure 7, right).

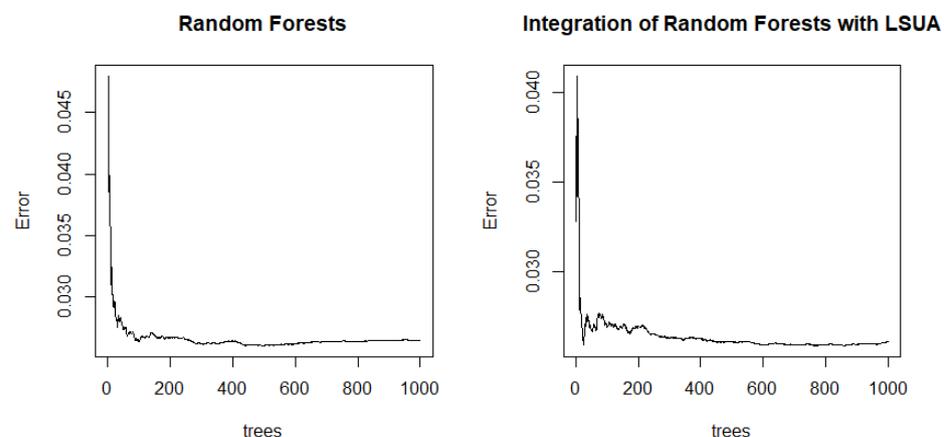


Figure 7. Optimization of the number of trees utilized in Random Forests (left) and the integration of Random Forests with LSUA (right).

Incorporating field measurements of carbon density as the dependent variable and using variables selected by LSR and LMSR as independent variables, RF achieved a coefficient of determination (R^2) of 0.7630, with plot mean and map mean estimates of 15.326 Mg/ha and 15.419 Mg/ha, respectively, and an RMSE of 8.741 Mg/ha. Integration of RF with LSUA improved performance, resulting in an R^2 of 0.7800, plot mean estimate and map mean of 15.207 Mg/ha and 15.136 Mg/ha, respectively, bringing them closer to the field measurement mean.

Both the RF model and the integration of RF with LSUA effectively depicted the spatial distribution of vegetation carbon density across the study area (Figure 6i,j). Both models tended to overestimate areas with 0 Mg/ha and underestimate regions with very high vegetation carbon density. The RF model combined with LSUA demonstrated improved stability and provided more detailed information on vegetation carbon in mixed pixels, requiring a greater number of trees for enhanced performance compared to the RF model without vegetation fraction from LSUA.

5. Discussion

Urban vegetation, comprising forests, shrubs, and grass, plays a vital role in mitigating atmospheric carbon concentration through processes such as photosynthesis. It contributes to air purification, noise reduction, and climate change impact mitigation [60–67]. The accurate mapping of urban vegetation carbon density is critical for governmental planning and residents' well-being. Our study presents a methodological framework for mapping vegetation carbon density in urban areas, utilizing a combination of field plot measurements, Landsat 8 imagery, and Pleiades 1A and 1B imagery data. While the integration of diverse data sources for mapping biomass or carbon density is not new in forested areas [68–70], our research addresses unique challenges in the urban environment. Shadows from buildings and mountains pose difficulties in accurately extracting vegetation information, compounded by the complex and fragmented urban landscape, leading to numerous mixed pixels that complicate the precise mapping of urban vegetation carbon density.

In this study, we employed LSUA to remove shadows induced by buildings and mountains from Landsat 8 images before spatial modeling. The method was evaluated using a random sample of 300 pixels across the study area, including 29 pixels located in mountain-induced shadow areas. While LSUA proved effective in mountainous regions, its performance was somewhat limited in urban areas due to the coarse spatial resolution of Landsat 8. After shadow removal, correlation coefficients between plot vegetation carbon density and shadow-removed Landsat 8 bands increased by 1.28% to 2.59% across different bands compared to the unprocessed Landsat 8 image. Our results align with findings from other studies employing LSUA for shadow removal and vegetation analysis [71,72]. Furthermore, our study validated LSUA by decomposing mixed pixels and extracting vegetation fractional information, achieving the highest correlation coefficient of 0.595 between the vegetation fractional image and plot carbon density when using four endmembers with automatic selection after shadow removal. This surpasses results obtained through manual selection, highlighting the effectiveness of LSUA in improving the accuracy of vegetation information extraction.

Spatial interpolation methods, including parametric (LSR, LMSR) and non-parametric (kNN, DT, RF) models, were utilized for mapping urban vegetation carbon density. The utilization of parametric and non-parametric models allows for a nuanced exploration of the complex relationships governing urban vegetation carbon density. While parametric models provide straightforward insights into linear relationships, non-parametric models, especially Decision Trees and Random Forests, offer a more flexible and adaptive approach, crucial for capturing the heterogeneity inherent in urban landscapes. The integration of diverse models aims to capitalize on their respective strengths, compensating for limitations in individual approaches. It is important to note that the selection of these models is not arbitrary but based on their suitability for addressing the specific challenges outlined in the study, emphasizing the need for a versatile and robust methodology in urban carbon

mapping. The ensuing comparative analysis of these models provides a comprehensive understanding of their performance, contributing valuable insights to the field of remote sensing and urban ecology.

The study employed R^2 and RMSE as performance metrics, revealing that Decision Trees consistently outperformed other models with the highest R^2 and the lowest RMSE, indicative of superior accuracy. The integration of LSUA enhanced prediction accuracies across models, although the improvements were not statistically significant, suggesting LSUA's role in refining models without a substantial boost in accuracy. Relative errors were compared among models, with Decision Trees and Random Forests consistently demonstrating lower errors, underscoring their reliability in capturing the complexities of the urban landscape. In the comparison between Decision Trees and Random Forests, Decision Trees, especially when integrated with LSUA, exhibited superior performance, addressing overfitting concerns and resulting in a slight improvement in correlation. Integration of these models with vegetation fractional images from LSUA improved mapping accuracy by 0.20% to 2.70%, depending on the model used. This enhancement, though relatively modest compared to some studies, signifies the potential for increased accuracy in mapping urban vegetation carbon density, especially in mixed environments with varying land features. The R^2 and RMSE results revealed that the DT model exhibited the best performance, demonstrating the highest R^2 and lowest RMSE among all models, whether integrated with vegetation fraction images from LSUA or not. LSR, regardless of LSUA integration, yielded extremely large and illogically negative estimates. Both LMSR and its integration with LSUA produced consistent estimates within the range of field plot values. kNN and kNN with LSUA tended to underestimate carbon density for large values and overestimate for small values. Relative RMSE values for urban vegetation carbon density estimates were 72.4% and 72.1% for LSR and LSR with LSUA, 61.4% and 60.3% for LMSR and LMSR with LSUA, 70.5% and 64.5% for kNN and kNN with LSUA, 58.3% and 57.7% for RF and RF with LSUA, and 46.35% and 45.87% for DT and DT with LSUA, respectively. The DT and DT with LSUA demonstrated the best performance, followed by RF and RF with LSUA, LMSR and LMSR with LSUA, and kNN and kNN with LSUA. LSUA integration improved prediction accuracies, albeit not significantly.

Compared to Yan et al. (2015) [73], the relative errors in this study were larger due to a higher coefficient of variation (108.87%) in urban vegetation carbon density resulting from a complex landscape, mixed pixels, and building-induced shadows. The dominance of built-up areas in Shenzhen city, as opposed to the forested areas in Yan et al.'s study, contributed to these differences. While kNN has shown effectiveness in mapping forest parameters in other studies [54,74–77], its application to estimate urban vegetation carbon density in this research resulted in lower accuracy compared to other methods. This discrepancy arises from the spectral distance-based approach of kNN, which encounters challenges in urban areas with mixed pixels exhibiting varied spectral reflectance due to diverse structures and compositions. The DT and DT with LSUA demonstrated strong predictive capabilities for urban vegetation carbon density, achieving high correlations of 0.8171 and 0.8205, respectively. The slight improvement in correlation with LSUA integration suggests that the DT model effectively mitigated overfitting issues. RF and RF with LSUA also exhibited high predictive capabilities, with RF integrated with LSUA identified as the more promising model for accurately mapping urban vegetation carbon density. This method holds great potential for rapid and accurate mapping with an acceptable level of error and cost-effectiveness.

In this study, although very high spatial resolution Pleiades 1A and 1B images were acquired, they were solely used for stratified random sampling and validating vegetation fraction images derived from LSUA. Despite Pleiades images providing superior visual interpretation capabilities compared to Landsat 8, their disadvantages include increased internal variability within homogeneous land cover polygons and limited spectral resolution with only four bands. The fine spatial resolution of Pleiades images also demands extensive storage and high-performance computation. Consequently, due to cost constraints,

Pleiades 1A and 1B images were not utilized for mapping vegetation carbon density in this study. The aim of this research was to establish a cost-efficient methodology for mapping urban vegetation carbon density in developing countries. The high cost of Pleiades images, amounting to \$58,608 (USD) for Shenzhen city coverage, renders their use impractical for mapping vegetation carbon density in large Chinese cities or in the regional scale. Utilizing freely available Landsat 8 images emerges as the optimal option, significantly reducing research costs. Future studies should investigate the impact of spatial and spectral resolutions on the accuracy of estimating urban vegetation carbon density. Urban sprawl induces land use and land cover (LULC) conversion, influencing total carbon stock and carbon pools. Previous studies highlighted the proportionate increase in anthropogenic carbon stock and decline in soil and vegetation carbon stock in populated areas [78–81]. While this study concentrated on mapping urban vegetation carbon density, future research should address and assess the effects of urbanization on carbon pool dynamics, considering data availability.

The study establishes a methodological framework to enhance the precision of urban vegetation carbon density mapping through spatial modeling, spectral unmixing analysis, and de-shadowing techniques. The validated efficacy of this framework indicates improved accuracy. To advance the field, future studies should prioritize innovative methods for refining vegetation information extraction from mixed pixels and mitigating shadow impacts. A promising avenue for accuracy enhancement lies in exploring data fusion techniques, especially by incorporating optical imagery with LiDAR and RADAR data. Further research directions involve exploring additional machine learning models, extending the methodology to diverse urban environments, and addressing challenges like spatial resolution limitations. Insights into the effects of urbanization on carbon dynamics and anthropogenic activities significantly contribute to a broader understanding of urban carbon mapping.

6. Conclusions

The study focuses on accurately mapping urban above-ground vegetation carbon density, addressing challenges posed by complex urban landscapes, mixed pixels, and building-induced shadows. A novel methodological framework is introduced, combining linear spectral unmixing analysis (LSUA) for shadow removal, spatial modeling, and integration of diverse data sources, including Landsat 8, Pleiades 1A and 1B, DEM, and field measurements. The shadow removal algorithm effectively operates in mountainous areas but shows limitations in urban settings due to Landsat 8's coarse spatial resolution. LSUA improves correlation after shadow removal, and integration with spatial models enhances mapping accuracy, with Decision Trees exhibiting superior performance. While relative improvements are modest, the potential for increased accuracy in mapping urban vegetation carbon density is highlighted. Despite challenges, the DT model, especially integrated with LSUA, demonstrates the best performance. However, relative errors are larger compared to a similar study, attributing this to the complex urban landscape. Cost considerations favor the use of freely available Landsat 8 images over higher-cost Pleiades images. Future research should explore the impact of spatial and spectral resolutions, assess urbanization effects on carbon pools, and develop novel methods for vegetation information extraction and shadow mitigation. Overall, the study provides a cost-efficient methodology with potential for accurate urban vegetation carbon density mapping.

Author Contributions: Conceptualization, G.Q., G.W. and M.W.; methodology, G.Q.; software, G.Q.; validation, G.W., M.W. and J.Y.; formal analysis, G.Q.; investigation, G.W.; resources, G.Q.; data curation, G.Q.; writing—original draft preparation, G.Q.; writing—review and editing, G.W. and M.W.; visualization, G.Q. and G.W.; supervision, G.W.; project administration, G.W., G.Q. and M.W.; funding acquisition, G.W., G.Q. and M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the research project “Shenzhen vegetation biomass and carbon modeling” awarded by Shenzhen Xianhu Botanic Garden, grant number #8851; Municipal Science and Technology Cooperation, grant number (2022) 163; Moutai College High-Level Talents Research Initiation Fund Project, grant number (2022) 134; Moutai College High-Level Talents Research Initiation Fund Project, grant number (2022) 049.

Data Availability Statement: All data will be available upon request.

Acknowledgments: Authors appreciate the assistance of Yifan Tan and Hua Sun on the field data collection and funding from the Shenzhen Xianhu Botanic Garden, Shenzhen, China, Department of Tourism of Management, Moutai Institute, China, and we appreciate all the anonymous reviewers of this article who provided valuable revision comments.

Conflicts of Interest: The authors declare no conflicts of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A

Table A1. Tree volume calculation equations.

Tree Species	Volume Calculation Equation
Eucalypts	$V = 8.71419 \times 10^{-5} D^{1.94801} H^{0.74929}$
Pinus elliotii	$V = 7.81515 \times 10^{-5} D^{1.79967} H^{0.98178}$
Acacia rachii	$V = 7.32715 \times 10^{-5} D^{1.65483} H^{1.08069}$
Chinese red pine	$V = 7.98524 \times 10^{-5} D^{1.74220} H^{1.01198}$
Castanopsis fissa	$V = 6.29692 \times 10^{-5} D^{1.81296} H^{1.01545}$
Broad-leaved	$V = 6.74286 \times 10^{-5} D^{1.87657} H^{0.92888}$
Cunningham lanceolata	$V = 6.97483 \times 10^{-5} D^{1.81583} H^{0.99610}$
Hard latissimus	$V = 6.01228 \times 10^{-5} D^{1.87550} H^{0.98496}$

Note: V—tree volume, D—diameter at breast height (1.3 m), H—height of tree.

Appendix B

Table A2. Biomass and volume relationship parameter values table.

Forest Types	a (Mg/m ³)	b (Mg)	N	R ²
Picea asperata Mast/Abies alba	0.5519	48.861	24	0.78
Bethula	1.0687	10.237	9	0.70
Casuarinaequisetifolia	0.7441	3.2377	10	0.95
Cunningham lanceolata	0.4652	19.141	90	0.94
Cedarwood	0.8893	7.3965	19	0.87
Cupressus funebris	1.1453	8.5473	12	0.98
Quercus subg Quercus sect	0.8873	4.5539	20	0.8
Eucalyptus robusta smith	0.6096	33.806	34	0.82
Larix princhipis-rupprechtii	0.9292	6.494	24	0.83
Subtropical evergreen broad-leaved forest	0.8136	18.466	10	0.99
Theropencedrymion	0.9788	5.3764	35	0.93
Broadleaf mixed plantations	0.5856	18.744	9	0.91
Pinus armandi	0.5723	16.489	22	0.93
Pinus massoniana	0.5034	20.547	52	0.87
Sylvestris/Pinus	1.112	2.6951	15	0.85

Table A2. *Cont.*

Forest Types	a (Mg/m ³)	b (Mg)	N	R ²
Pinustabuliformis	0.869	9.1212	112	0.91
Others Conifer	0.5292	25.087	19	0.86
Aspen	0.4969	26.973	13	0.92
Tsugachinensis/Criptomeriafortunei	0.3491	39.816	30	0.79
Tropical forests	0.7975	0.4204	18	0.87

Appendix C

Table A3. Carbon ratio table for different tree species in China.

Trees Species	Ratio	Tree Species	Ratio
Picea asperata Mas	0.4994	Schima	0.5115
Tsuga chinensis	0.5022	Others broad-leaved hard wood	0.4901
Larix gmelinii	0.5137	Aspen	0.4502
Pinus koraiensis Sieb	0.5113	Eucalyptus	0.4748
Pinus thunbergii Parl	0.5146	Acacia rachii	0.4666
Pinus tabulaeformis	0.5184	Others broad-leaved soft wood	0.4502
Pinus armandii Franch	0.5177	Broadleaf mixed trees	0.4796
Pinus massoniana Lamb	0.5271	Economic trees	0.4700
Pinus elliotii	0.5311	Cupressus funebris Endl	0.5088
Others Pinus	0.4963	Coniferous mixed forest	0.5168
Cunninghamia lanceolate	0.5127	* Bush	0.4672
Conifer-broadleaf forest	0.4893	* Herbal	0.3270

Note: * Bush is a joint name of all kinds of different shrub species, * Herbal is a joint name of all kinds of different grass species.

References

- Kan, K.; Chen, J. Rural urbanization in China: Administrative restructuring and the livelihoods of urbanized rural residents. *J. Contemp. China* **2022**, *31*, 626–643. [[CrossRef](#)]
- Wang, G.; Shi, X.; Cui, H.; Jiao, J. Impacts of migration on urban environmental pollutant emissions in China: A comparative perspective. *Chin. Geogr. Sci.* **2020**, *30*, 45–58. [[CrossRef](#)]
- Chen, W.Y. The role of urban green infrastructure in offsetting carbon emissions in 35 major Chinese cities: A nationwide estimate. *Cities* **2015**, *44*, 112–120. [[CrossRef](#)]
- Sun, H.; Qie, G.; Wang, G.; Tan, Y.; Li, J.; Peng, Y.; Luo, C. Increasing the accuracy of mapping urban forest carbon density by combining spatial modeling and spectral unmixing analysis. *Remote Sens.* **2015**, *7*, 15114–15139. [[CrossRef](#)]
- Ren, Z.; Zheng, H.; He, X.; Zhang, D.; Shen, G.; Zhai, C. Changes in spatio-temporal patterns of urban forest and its above-ground carbon storage: Implication for urban CO₂ emissions mitigation under China's rapid urban expansion and greening. *Environ. Int.* **2019**, *129*, 438–450. [[CrossRef](#)]
- Contosta, A.R.; Lerman, S.B.; Xiao, J.; Varner, R.K. Biogeochemical and socioeconomic drivers of above-and below-ground carbon stocks in urban residential yards of a small city. *Landsc. Urban Plan.* **2020**, *196*, 103724. [[CrossRef](#)]
- Abbas, S.; Wong, M.S.; Wu, J.; Shahzad, N.; Muhammad Irteza, S. Approaches of satellite remote sensing for the assessment of above-ground biomass across tropical forests: Pan-tropical to national scales. *Remote Sens.* **2020**, *12*, 3351. [[CrossRef](#)]
- Pasetto, D.; Arenas-Castro, S.; Bustamante, J.; Casagrandi, R.; Chrysoulakis, N.; Cord, A.F.; Ziv, G. Integration of satellite remote sensing data in ecosystem modelling at local scales: Practices and trends. *Methods Ecol. Evol.* **2018**, *9*, 1810–1821. [[CrossRef](#)]
- Campbell, M.J.; Dennison, P.E.; Kerr, K.L.; Brewer, S.C.; Anderegg, W.R. Scaled biomass estimation in woodland ecosystems: Testing the individual and combined capacities of satellite multispectral and lidar data. *Remote Sens. Environ.* **2021**, *262*, 112511. [[CrossRef](#)]
- Wu, H.; Li, Z.L. Scale issues in remote sensing: A review on analysis, processing and modeling. *Sensors* **2009**, *9*, 1768–1793. [[CrossRef](#)] [[PubMed](#)]

11. Keith, H.; Mackey, B.G.; Lindenmayer, D.B. Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11635–11640. [[CrossRef](#)]
12. Clough, B.J.; Curzon, M.T.; Domke, G.M.; Russell, M.B.; Woodall, C.W. Climate-driven trends in stem wood density of tree species in the eastern United States: Ecological impact and implications for national forest carbon assessments. *Glob. Ecol. Biogeogr.* **2017**, *26*, 1153–1164. [[CrossRef](#)]
13. Keith, H.E.A.T.H.E.R.; Mackey, B.; Berry, S.; Lindenmayer, D.; Gibbons, P. Estimating carbon carrying capacity in natural forest ecosystems across heterogeneous landscapes: Addressing sources of error. *Glob. Chang. Biol.* **2010**, *16*, 2971–2989. [[CrossRef](#)]
14. Brown, S. Measuring carbon in forests: Current status and future challenges. *Environ. Pollut.* **2002**, *116*, 363–372. [[CrossRef](#)]
15. Zianis, D.; Mencuccini, M. On simplifying allometric analyses of forest biomass. *For. Ecol. Manag.* **2004**, *187*, 311–332. [[CrossRef](#)]
16. Fortier, J.; Truax, B.; Gagnon, D.; Lambert, F. Allometric equations for estimating compartment biomass and stem volume in mature hybrid poplars: General or site-specific? *Forests* **2017**, *8*, 309. [[CrossRef](#)]
17. Cole, T.G.; Ewel, J.J. Allometric equations for four valuable tropical tree species. *For. Ecol. Manag.* **2006**, *229*, 351–360. [[CrossRef](#)]
18. Vargas-Larreta, B.; López-Sánchez, C.A.; Corral-Rivas, J.J.; López-Martínez, J.O.; Aguirre-Calderón, C.G.; Álvarez-González, J.G. Allometric equations for estimating biomass and carbon stocks in the temperate forests of North-Western Mexico. *Forests* **2017**, *8*, 269. [[CrossRef](#)]
19. Yuen, J.Q.; Fung, T.; Ziegler, A.D. Review of allometric equations for major land covers in SE Asia: Uncertainty and implications for above-and below-ground carbon estimates. *For. Ecol. Manag.* **2016**, *360*, 323–340. [[CrossRef](#)]
20. Xing, D.; Bergeron, J.C.; Solarik, K.A.; Tomm, B.; Macdonald, S.E.; Spence, J.R.; He, F. Challenges in estimating forest biomass: Use of allometric equations for three boreal tree species. *Can. J. For. Res.* **2019**, *49*, 1613–1622. [[CrossRef](#)]
21. Zhu, J.; Huang, Z.; Sun, H.; Wang, G. Mapping forest ecosystem biomass density for Xiangjiang River Basin by combining plot and remote sensing data and comparing spatial extrapolation methods. *Remote Sens.* **2017**, *9*, 241. [[CrossRef](#)]
22. Galeana-Pizaña, J.M.; López-Caloca, A.; López-Quiroz, P.; Silván-Cárdenas, J.L.; Couturier, S. Modeling the spatial distribution of above-ground carbon in Mexican coniferous forests using remote sensing and a geostatistical approach. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *30*, 179–189. [[CrossRef](#)]
23. Mauya, E.W.; Koskinen, J.; Tegel, K.; Hämäläinen, J.; Kauranne, T.; Käyhkö, N. Modelling and predicting the growing stock volume in small-scale plantation forests of Tanzania using multi-sensor image synergy. *Forests* **2019**, *10*, 279. [[CrossRef](#)]
24. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, *2017*, 1353691. [[CrossRef](#)]
25. Ahmad, A.; Gilani, H.; Ahmad, S.R. Forest aboveground biomass estimation and mapping through high-resolution optical satellite imagery—A literature review. *Forests* **2021**, *12*, 914. [[CrossRef](#)]
26. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *Int. J. Dig. Earth.* **2016**, *9*, 63–105. [[CrossRef](#)]
27. Zhang, L.; Shao, Z.; Liu, J.; Cheng, Q. Deep learning based retrieval of forest aboveground biomass from combined LiDAR and landsat 8 data. *Remote Sens.* **2019**, *11*, 1459. [[CrossRef](#)]
28. Tan, B.; Woodcock, C.E.; Hu, J.; Zhang, P.; Ozdogan, M.; Huang, D.; Yang, W.; Knyazikhin, Y.; Myneni, R.B. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. *Remote Sens. Environ.* **2006**, *105*, 98–114. [[CrossRef](#)]
29. Mahabir, R.; Croitoru, A.; Crooks, A.T.; Agouris, P.; Stefanidis, A. A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. *Urban Sci.* **2018**, *2*, 8. [[CrossRef](#)]
30. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote sensing technologies for enhancing forest inventories: A review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [[CrossRef](#)]
31. Surový, P.; Kuželka, K. Acquisition of forest attributes for decision support at the forest enterprise level using remote-sensing techniques—A review. *Forests* **2019**, *10*, 273. [[CrossRef](#)]
32. Liang, X.; Kukko, A.; Balenović, I.; Saarinen, N.; Junntila, S.; Kankare, V.; Holopainen, M.; Makarovs, M.; Surovy, P.; Kaartinen, H.; et al. Close-Range Remote Sensing of Forests: The state of the art, challenges, and opportunities for systems and data acquisitions. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 32–71. [[CrossRef](#)]
33. Mulatu, K.A.; Decuyper, M.; Brede, B.; Kooistra, L.; Reiche, J.; Mora, B.; Herold, M. Linking terrestrial LiDAR scanner and conventional forest structure measurements with multi-modal satellite data. *Forests* **2019**, *10*, 291. [[CrossRef](#)]
34. Reba, M.; Seto, K.C. A systematic review and assessment of algorithms to detect, characterize, and monitor urban land change. *Remote Sens. Environ.* **2020**, *242*, 111739. [[CrossRef](#)]
35. Paudel, S.; Yuan, F. Assessing landscape changes and dynamics using patch analysis and GIS modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *16*, 66–76. [[CrossRef](#)]
36. Somers, B.; Asner, G.P.; Tits, L.; Coppin, P. Endmember variability in spectral mixture analysis: A review. *Remote Sens. Environ.* **2011**, *115*, 1603–1616. [[CrossRef](#)]
37. Zanutta, D.C.; Haertel, V.; Shimabukuro, Y.E.; Renno, C.D. Linear spectral mixing model for identifying potential missing endmembers in spectral mixture analysis. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 3005–3012. [[CrossRef](#)]
38. Nielsen, A.A. Spectral mixture analysis: Linear and semi-parametric full and iterated partial unmixing in multi-and hyperspectral image data. *J. Math. Imaging Vis.* **2001**, *15*, 17–37. [[CrossRef](#)]

39. Chen, F.; Wang, K.; Tang, T.F. Spectral unmixing using a sparse multiple-endmember spectral mixture model. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5846–5861. [[CrossRef](#)]
40. Song, C. Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability? *Remote Sens. Environ.* **2005**, *95*, 248–263. [[CrossRef](#)]
41. Su, N.; Zhang, Y.; Tian, S.; Yan, Y.; Miao, X. Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2568–2582. [[CrossRef](#)]
42. Dare, P.M. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 169–177. [[CrossRef](#)]
43. Shahtahmassebi, A.; Yang, N.; Wang, K.; Moore, N.; Shen, Z. Review of shadow detection and de-shadowing methods in remote sensing. *Chin. Geogr. Sci.* **2013**, *23*, 403–420. [[CrossRef](#)]
44. Wen, J.; Liu, Q.; Xiao, Q.; Liu, Q.; You, D.; Hao, D.; Wu, S.; Lin, X. Characterizing land surface anisotropic reflectance over rugged terrain: A review of concepts and recent developments. *Remote Sens.* **2018**, *10*, 370. [[CrossRef](#)]
45. Yang, X.; Zuo, X.; Xie, W.; Li, Y.; Guo, S.; Zhang, H. A Correction Method of NDVI Topographic Shadow Effect for Rugged Terrain. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8456–8472. [[CrossRef](#)]
46. Jiang, H.; Chen, A.; Wu, Y.; Zhang, C.; Chi, Z.; Li, M.; Wang, X. Vegetation Monitoring for Mountainous Regions Using a New Integrated Topographic Correction (ITC) of the SCS+ C Correction and the Shadow-Eliminated Vegetation Index. *Remote Sens.* **2022**, *14*, 3073. [[CrossRef](#)]
47. Luo, S.; Shen, H.; Li, H.; Chen, Y. Shadow removal based on separated illumination correction for urban aerial remote sensing images. *Signal Process.* **2019**, *165*, 197–208. [[CrossRef](#)]
48. Shi, L.; Zhao, Y.F. Urban feature shadow extraction based on high-resolution satellite remote sensing images. *Alex. Eng. J.* **2023**, *77*, 443–460. [[CrossRef](#)]
49. Azevedo, S.; Silva, E.; Colnago, M.; Negri, R.; Casaca, W. Shadow detection using object area-based and morphological filtering for very high-resolution satellite imagery of urban areas. *J. Appl. Remote Sens.* **2019**, *13*, 036506. [[CrossRef](#)]
50. Jalkanen, A.; Mattila, U. Logistic regression models for wind and snow damage in northern Finland based on the National Forest Inventory data. *For. Ecol. Manag.* **2000**, *135*, 315–330. [[CrossRef](#)]
51. Guo, F.; Zhang, L.; Jin, S.; Tigabu, M.; Su, Z.; Wang, W. Modeling anthropogenic fire occurrence in the boreal forest of China using logistic regression and random forests. *Forests* **2016**, *7*, 250. [[CrossRef](#)]
52. Shi, Y.; Feng, C.; Yang, S. Predictive Modeling of Forest Fires in Yunnan Province: An Integration of ARIMA and Stepwise Regression Analysis. *Appl. Sci.* **2023**, *14*, 256. [[CrossRef](#)]
53. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [[CrossRef](#)]
54. Gjertsen, A.K. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sens. Environ.* **2007**, *110*, 420–430. [[CrossRef](#)]
55. McNerney, D.O.; Nieuwenhuis, M. A comparative analysis of k NN and decision tree methods for the Irish National Forest Inventory. *Int. J. Remote Sens.* **2009**, *30*, 4937–4955. [[CrossRef](#)]
56. Gleason, C.J.; Im, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [[CrossRef](#)]
57. Esteban, J.; McRoberts, R.E.; Fernández-Landa, A.; Tomé, J.L.; Næsset, E. Estimating forest volume and biomass and their changes using random forests and remotely sensed data. *Remote Sens.* **2019**, *11*, 1944. [[CrossRef](#)]
58. Breslow, L.A.; Aha, D.W. Simplifying decision trees: A survey. *Knowl. Eng. Rev.* **1997**, *12*, 1–40. [[CrossRef](#)]
59. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
60. Vieira, J.; Matos, P.; Mexia, T.; Silva, P.; Lopes, N.; Freitas, C.; Correia, O.; Santos-Reis, M.; Branquinho, C.; Pinho, P. Green spaces are not all the same for the provision of air purification and climate regulation services: The case of urban parks. *Environ. Res.* **2018**, *160*, 306–313. [[CrossRef](#)]
61. Gong, C.; Xian, C.; Ouyang, Z. Assessment of NO₂ Purification by Urban Forests Based on the i-Tree Eco Model: Case Study in Beijing, China. *Forests* **2022**, *13*, 369. [[CrossRef](#)]
62. Van Renterghem, T. Towards explaining the positive effect of vegetation on the perception of environmental noise. *Urban For. Urban Green.* **2019**, *40*, 133–144. [[CrossRef](#)]
63. Canadell, J.G.; Raupach, M.R. Managing forests for climate change mitigation. *Science* **2008**, *320*, 1456–1457. [[CrossRef](#)]
64. Alemu, B. The role of forest and soil carbon sequestrations on climate change mitigation. *Res. J. Agr. Environ. Manag.* **2014**, *3*, 492–505.
65. Mader, S. Plant trees for the planet: The potential of forests for climate change mitigation and the major drivers of national forest area. *Mitig. Adapt. Strateg. Glob. Chang.* **2020**, *25*, 519–536. [[CrossRef](#)]
66. Lundmark, T.; Bergh, J.; Hofer, P.; Lundström, A.; Nordin, A.; Poudel, B.C.; Sathre, R.; Taverna, R.; Werner, F. Potential roles of Swedish forestry in the context of climate change mitigation. *Forests* **2014**, *5*, 557–578. [[CrossRef](#)]
67. Kauppi, P.E.; Stål, G.; Arnesson-Ceder, L.; Sramek, I.H.; Hoen, H.F.; Svensson, A.; Wernick, I.K.; Högberg, P.; Lundmark, T.; Nordin, A. Managing existing forests can mitigate climate change. *For. Ecol. Manag.* **2022**, *513*, 120186. [[CrossRef](#)]
68. Avitabile, V.; Herold, M.; Henry, M.; Schmillius, C. Mapping biomass with remote sensing: A comparison of methods for the case study of Uganda. *Carbon Bal. Manag.* **2011**, *6*, 7. [[CrossRef](#)]

69. Kumar, L.; Mutanga, O. Remote sensing of above-ground biomass. *Remote Sens.* **2017**, *9*, 935. [[CrossRef](#)]
70. Zheng, G.; Chen, J.M.; Tian, Q.J.; Ju, W.M.; Xia, X.Q. Combining remote sensing imagery and forest age inventory for biomass mapping. *J. Environ. Manag.* **2007**, *85*, 616–623. [[CrossRef](#)]
71. Yang, J.; He, Y.; Caspersen, J. Fully constrained linear spectral unmixing based global shadow compensation for high resolution satellite imagery of urban areas. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 88–98. [[CrossRef](#)]
72. Yang, J.; Li, P. Impervious surface extraction in urban areas from high spatial resolution imagery using linear spectral unmixing. *Remote Sens. Appl. Soc. Environ.* **2015**, *1*, 61–71. [[CrossRef](#)]
73. Yan, E.; Lin, H.; Wang, G.; Sun, H. Improvement of forest carbon estimation by integration of regression modeling and spectral unmixing of Landsat data. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2003–2007.
74. Koukal, T.; Suppan, F.; Schneider, W. The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes. *Remote Sens. Environ.* **2007**, *110*, 431–437. [[CrossRef](#)]
75. Baffetta, F.; Corona, P.; Fattorini, L. A matching procedure to improve k-NN estimation of forest attribute maps. *For. Ecol. Manag.* **2012**, *272*, 35–50. [[CrossRef](#)]
76. Vega Isuhuaylas, L.A.; Hirata, Y.; Ventura Santos, L.C.; Serrudo Torobeo, N. Natural forest mapping in the Andes (Peru): A comparison of the performance of machine-learning algorithms. *Remote Sens.* **2018**, *10*, 782. [[CrossRef](#)]
77. Beaudoin, A.; Bernier, P.Y.; Guindon, L.; Villemaire, P.; Guo, X.J.; Stinson, G.; Bergeron, T.; Magnussen, S.; Hall, R.J. Mapping attributes of Canada's forests at moderate resolution through k NN and MODIS imagery. *Can. J. For. Res.* **2014**, *44*, 521–532. [[CrossRef](#)]
78. Certini, G.; Scalenghe, R. Anthropogenic soils are the golden spikes for the Anthropocene. *Holocene* **2011**, *21*, 1269–1274. [[CrossRef](#)]
79. Villa, P.; Malucelli, F.; Scalenghe, R. Multitemporal mapping of peri-urban carbon stocks and soil sealing from satellite data. *Sci. Total Environ.* **2018**, *612*, 590–604. [[CrossRef](#)]
80. Sarzhanov, D.A.; Vasenev, V.I.; Vasenev, I.I.; Sotnikova, Y.L.; Ryzhkov, O.V.; Morin, T. Carbon stocks and CO₂ emissions of urban and natural soils in Central Chernozemic region of Russia. *Catena* **2017**, *158*, 131–140. [[CrossRef](#)]
81. Tao, Y.; Li, F.; Liu, X.; Zhao, D.; Sun, X.; Xu, L. Variation in ecosystem services across an urbanization gradient: A study of terrestrial carbon stocks from Changzhou, China. *Ecol. Model.* **2015**, *318*, 210–216. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.