*Article*

# Increasing the Accuracy of Mapping Urban Forest Carbon Density by Combining Spatial Modeling and Spectral Unmixing Analysis

**Hua Sun [1], Guangping Qie [2], Guangxing Wang [1,2,\*], Yifan Tan [3], Jiping Li [1], Yougui Peng [4], Zhonggang Ma [1] and Chaoqin Luo [1]**

[1] Research Center of Forestry Remote Sensing and Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China;
E-Mails: sunhuayiwen@vip.126.com (H.S.); lijiping@vip.163.com (J.L.); mzgbeyond@163.com (Z.M.); chaoqinluo@126.com (C.L.)

[2] Department of Geography, Southern Illinois University at Carbondale, Carbondale, IL 62901, USA; E-Mail: guangpingqie@siu.edu

[3] Xianhu Botanic Garden of Shenzhen, Shenzhen 518004, China; E-Mail: yifantan@tom.com

[4] College of Forestry, South China Agricultural University, Guangzhou 510642, China;
E-Mail: pygui2000@163.com

**\*** Author to whom correspondence should be addressed; E-Mail: gxwang@siu.edu;
Tel.: +1-618-453-6017; Fax: +1-618-453-6465.

**Abstract**: Accurately mapping urban vegetation carbon density is challenging because of complex landscapes and mixed pixels. In this study, a novel methodology was proposed that combines a linear spectral unmixing analysis (LSUA) with a linear stepwise regression (LSR), a logistic model-based stepwise regression (LMSR) and k-Nearest Neighbors (kNN), to map the forest carbon density of Shenzhen City of China, using Landsat 8 imagery and sample plot data collected in 2014. The independent variables that contributed to statistically significantly improving the fit of a model to data and reducing the sum of squared errors were first selected from a total of 284 spectral variables derived from the image bands. The vegetation fraction from LSUA was then added as an independent variable. The results obtained using cross-validation showed that: (1) Compared to the methods without the vegetation information, adding the vegetation fraction increased the accuracy of mapping carbon density by 1%–9.3%; (2) As the observed values increased, the LSR and kNN

residuals showed overestimates and underestimates for the smaller and larger observations, respectively, while LMSR improved the systematical over and underestimations; (3) LSR resulted in illogically negative and unreasonably large estimates, while KNN produced the greatest values of root mean square error (RMSE). The results indicate that combining the spatial modeling method LMSR and the spectral unmixing analysis LUSA, coupled with Landsat imagery, is most promising for increasing the accuracy of urban forest carbon density maps. In addition, this method has considerable potential for accurate, rapid and nondestructive prediction of urban and peri-urban forest carbon stocks with an acceptable level of error and low cost.

# 1. Introduction

During the last 35 years, rapid urbanization of Southern China has led to the development of a cluster of cities, including Shenzhen, Guangzhou, Foshan, and Dongguan. These cities have grown so rapidly that they have merged into an interconnected megalopolis with a population of 42 million. Rapid urban sprawl and increasing human activities have resulted in deforestation and loss of vegetated areas and agricultural lands, increased ratio of anthropogenic to natural carbon stocks, increased carbon emission and aggravated air pollution in this region, and thus contributed to carbon concentration in the atmosphere and to global warming [1]. To mitigate these effects, an effective strategy is to protect forests and re-vegetate the cities, which can thus potentially enhance vegetation carbon sequestration in the region. However, urban vegetation carbon sequestration and its dynamics in this region are still unknown mainly because of the complexity of urban landscapes in which buildings, roads, parking lots, water bodies, forests, trees along city roads, shrubs, grasslands and other greened areas are mixed together. The large number of mixed pixels impedes accurate estimation of urban forest carbon. Currently, most of the methods used to map vegetation carbon are related to general or non-urban forests and are not appropriate for estimating urban vegetation carbon [1,2]. There are also no reports of techniques for introducing the vegetation fraction extracted from the mixed pixels of urban landscapes into spatial interpolation methods to improve mapping urban vegetation carbon. Therefore, there is a strong need to develop a novel method for urban vegetation carbon modeling and dynamic monitoring.

## 1.1. Forest Carbon Modeling

Forest carbon sequestration plays an important role in controlling global climate change [3–7]. During the last 20 years, substantial research on the estimation of forest carbon stock per unit or density has been conducted, and various methods have been developed [8,9]. These methods can be divided into multiple categories: use of biomass or net primary productivity observations from field sample plots [10]; use of biomass and carbon conversion factors or coefficients based on the data from forest inventory [3,11]; eddy correlation and covariance based methods [12–14]; growth and especially

process model-based methods [15] such as BIOME-BGC [16–18]; and remote sensing-based methods [9,19–24]. Compared to other methods, remote sensing-based methods have the following advantages: (1) remotely sensed imagery provides coverage of large areas, reveals the spatial variability of vegetation and facilitates generation of the spatial distribution and pattern of vegetation carbon density for large vegetated areas and complex urban vegetation landscapes; (2) various sensors (optical, Radar, Lidar) and spatial resolution images can be selected to tailor vegetation carbon modeling at various scales; and (3) multi-temporal imagery makes it possible to monitor the dynamics of vegetation carbon density. Thus, remote sensing-based methods are a good alternative for estimating urban vegetation carbon stocks and monitoring their dynamics. The success of the methods will depend on selection of the remotely sensed data, selection of the spectral variables that account for their relationships with vegetation carbon density, selection of sample plots and collection of field observations, and modeling the relationship between vegetation carbon data from sample plots and the spectral variables.

Substantial research has been conducted on estimation of forest carbon density estimation using a combination of sample plot and remotely sensed data. These studies used various remotely sensed data including Moderate Resolution Imaging Spectroradiometer (MODIS), Landsat Thematic Mapper™ and SPOT [8]. Landsat TM images are the most widely used because they are free, downloadable, and have a long time history, large coverage, and medium spatial and temporal resolutions [8,9,20,25,26]. Moreover, the 30 m × 30 m spatial resolution of Landsat TM multi-spectral images often approximates the size of sample plots used for field data collection by the national forest inventories of many countries. For instance, the size of sample plots used by China's national forest inventory is 25.82 m × 25.82 m. This size relationship potentially reduces the uncertainty due to geospatial mismatches between pixels and sample plots. The major disadvantage of Landsat TM imagery is the lack of sensitivity of brightness values to biomass of very dense and multi-layer canopy forests in tropical rain forests, which leads to underestimation of biomass and carbon [8,20]. However, this saturation problem usually does not occur in the estimation of urban forest carbon. SPOT imagery has characteristics similar to those of Landsat TM, but its high cost limits its application. MODIS imagery has been widely used for global, national and regional vegetation carbon modeling, but is generally not applicable for estimating urban forest carbon density due to its coarse spatial resolution. An exception is the study conducted by Villa *et al.* [27] who found that remotely sensed images with spatial resolutions of 500 m × 500 m to 1000 m × 1000 m could be used to effectively estimate carbon stocks for peri-urban areas. This exception can be attributed to the low intensity of urbanization in the peri-urban areas. The fine spatial resolution images such as those from Ikonos and Worldview-3 have the potential to extract vegetation fraction from mixed pixels and, thus, are promising for mapping urban forest carbon density. However, the high cost required is a challenge associated with fine spatial resolution imagery.

Radar imagery overcomes the shortcoming of optical imagery due to clouds [28–32]. However, the images are not very sensitive to the structures of dense and multi-layer canopy forests. Moreover, forest carbon estimation studies during the last decade indicate that both airborne LiDAR and Terrestrial Laser Scanning (TLS) data provide greater potential than other remotely sensed images because of their capacity to estimate tree heights and crown widths [33–35]. In addition, the integration of LiDAR data with optical images can potentially increase the accuracy of forest carbon estimation [36–43]. Like fine spatial resolution optical images, however, the high cost of both Radar and LiDAR data limit their applications.

With remote sensing-based methods, various spatial interpolation techniques, including linear and non-linear regressions, artificial neural networks (ANN), k-Nearest Neighbors (kNN), and geostatistical methods such as cokriging and spatial co-simulation, are used to model the relationships between forest biomass/carbon and spectral variables from images [8,10,44]. The limitations of linear regressions are the linearity and normality assumptions for the data. In addition, linear regressions often lead to not only negative predictions of forest biomass/carbon, but also extremely large and meaningless predictions [45,46]. ANN and kNN do not require the above assumptions and are more flexible for mapping vegetation biomass/carbon [25,47]. However, ANN is still considered a "black box" approach due to unknown weights and difficulty in terminating the learning process [48]. With kNN, the estimate of forest carbon for an unknown location can be calculated as an inverse, distance-weighted mean of observations from k sample plots that are nearest in a multi-dimension spectral space [49]. This method has been widely used with the national forest inventory plot and remotely sensed data [49–54]. In addition, Wang *et al.* [9,22] developed an image-based, spatial co-simulation algorithm based on spatial autocorrelation of variables to map forest carbon density. Fleming *et al.* [45] further compared this algorithm with regressions for mapping forest carbon density and found that the co-simulation was more promising than the regressions. However, the complexity of this co-simulation algorithm impedes its applications.

### 1.2. Urban Forest Carbon Modeling

An urban vegetation landscape is much more complex than general terrestrial vegetation ecosystems. Thus, accurately mapping urban forest carbon density is more difficult because mixed pixels and building-induced shadows often impede extraction of vegetation information. Moreover, urban forests often have less capacity for carbon sequestration compared to other forests [55]. Many authors have studied the estimation of urban forest carbon [56–63] using methods that are similar to those mentioned above for non-urban forest carbon modeling. For example, Kordowski and Kuttler [64] used eddy correlation and covariance methods to estimate the carbon flux of Essen in Germany from 2006 to 2007. Churkina *et al.* [1] applied a simple average per unit for homogeneous polygons to estimate urban carbon stocks of the continental USA. Zhang *et al.* [65] modeled the dynamics of vegetation carbon from 1945 to 2007 for the cities in the southern USA using a process-based model. Strohbach *et al.* [66] proposed a new method in which the sequestration of urban forests due to tree growth was simulated and compared to the corresponding carbon emission. Nowak *et al.* [67] estimated the carbon density for a total of 28 cities across six states in the USA using ground sample plot data.

In addition, remote sensing-based methods have been widely used to estimate and monitor the dynamics of urban forest carbon density [68]. McPherson *et al.* [69] developed spatial distributions of forest carbon density by integrating forest growth models, ground sample plot and remotely sensed data for Los Angeles and Sacramento. Zheng *et al.* [70] combined forest growth models, a land use and land cover (LULC) classification map and a tree canopy cover map from satellite images to estimate urban forest carbon density for county-level cities in northern New England, USA. Strohbach and Haase [66] combined a tree canopy cover map based on hyperspectral color infrared imagery and ground observations to estimate the urban forest carbon stock of Leipzig. Hutyra *et al.* [71] estimated the decrease in forested areas and the loss of forest carbon stocks due to urban sprawl for Seattle from 1986

to 2007 by combining a time series of remotely sensed images and ground sample plot data. Shrestha and Wynne [72] estimated the height and crown width of each tree using airborne LiDAR data and further estimated tree biomass for central Oklahoma, USA. Moskal and Zheng [73] explored methods used to estimate tree heights, diameters, and basal areas for an urban landscape using TLS. Davies *et al.* [74] and Pataki *et al.* [75] reviewed methods for urban vegetation carbon estimation and indicated that the complexity of urban landscapes and great variation in vegetation carbon density may lead to large estimation uncertainties.

In China, urban forest carbon modeling research has mainly been conducted for large cities such as Beijing [76], Shanghai [77], and Guangzhou [78–81]. For example, He *et al.* [76] estimated Beijing's forest biomass by combining TLS data and SPOT imagery. Ying *et al.* [82] estimated the spatial distribution of Harbin's forest carbon density by combining data from field sample plots and a remote sensing-based LULC classification. After reviewing the widely used methods for modeling urban vegetation carbon, Zhou *et al.* [81] suggested that the integration of sample plot data, models and remote sensing technologies would have considerable potential increasing the accuracy of urban vegetation carbon modeling by decomposing the mixed pixels of complex urban vegetation landscapes and introducing vegetation information from spectral unmixing analyses into the models.

### 1.3. Spectral Mixture Analysis

Mixed pixels in complex urban vegetation landscapes can greatly reduce the estimation accuracy of vegetation carbon density. The key for development of methods for mapping urban vegetation carbon density is a method for decomposing mixed pixels and using the derived vegetation information. However, accurately extracting the vegetation information of mixed pixels in images is often a great challenge [83,84]. As the spatial resolutions of images become coarser, the mixed pixels become more problematic. Various methods have been developed for decomposition of land cover components or endmembers in mixed pixels. These methods are divided into linear and nonlinear spectral unmixing algorithms [85–87]. Lu *et al.* [88] used a linear spectral unmixing method and obtained a classification accuracy of 78.2% for Brazilian Amazon vegetation types based on Landsat TM images. Chen *et al.* [89] developed a method based on linear spectral unmixing to estimate leaf area index. Dennison and Roberts [83] discussed the selection of endmembers. Hu and Weng [90] provided a summary of spectral unmixing analysis methods and then compared the fraction image of impervious surface from ANN with that from a linear spectral unmixing analysis using a Hyperion image. Although there have been many reports, general rules on how to select endmembers and spectral unmixing methods are still lacking. In addition, there have been no reported studies on how to integrate spectral unmixing analyses and spatial interpolation methods to map urban vegetation carbon density.

This study aimed at developing an improved method for increasing the accuracy of mapping forest carbon density for complex urban landscapes by combining a linear spectral unmixing analysis (LSUA), Landsat 8 imagery, sample plot data and three spatial interpolation methods: a linear stepwise regression (LSR), a logistic model-based regression (LMSR) and kNN. The combinations were compared by introducing the vegetation fraction from LSUA into the spatial interpolation methods. Methods that use the vegetation fraction are expected to increase estimation accuracy for urban forest carbon density relative to methods that do not use the vegetation fraction.
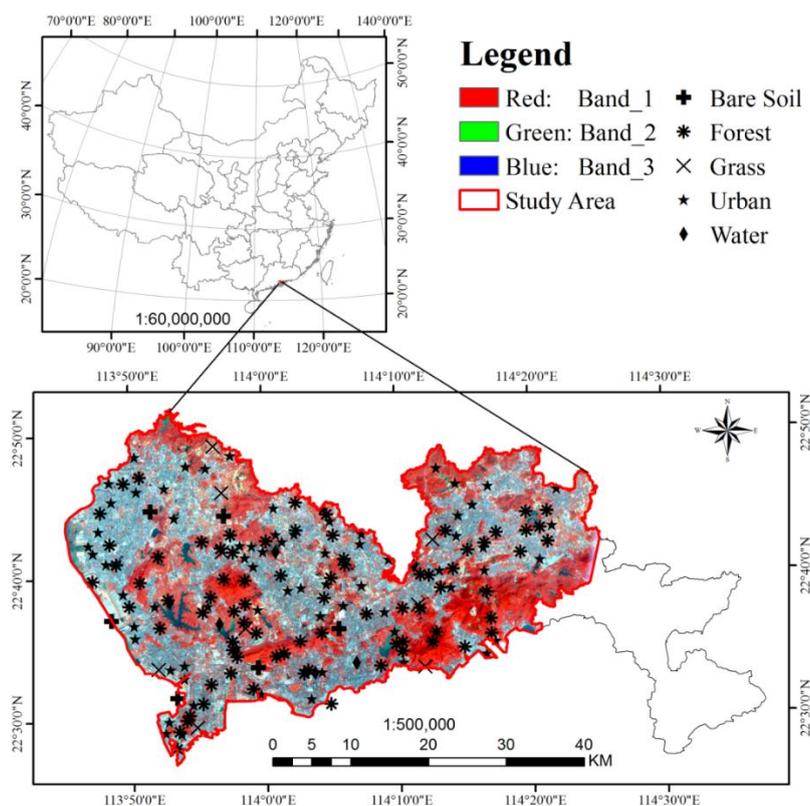
## 2. Study Area and Datasets

This study was conducted in Shenzhen city, one of the four rapidly sprawling cities in Guangdong province of Southern China (Figure 1). During the past 30 years, Shenzhen has rapidly spread from a small village to become a big city with a population of 11 million. The rapid-city expansion and increasing human activities have converted large amounts of the vegetated areas and agricultural lands into built up areas and, thus, led to the great changes in LULC types characterized by deforestation. Currently, Shenzhen has an area of 81.4 km from east to west and 45 km from north to south. It is characterized by a sub-tropical monsoon climate with the annual average temperature and precipitation of 22.5 ℃ and 1924.3 mm. The vegetated area of the city consists of forested lands, shrub lands, grasslands and greened roofs (planted grass and flowers on roofs). Forested lands have an area of 79,700 ha with about 4,880,000 trees planted along roads and streets. Deciduous forests, deciduous and coniferous mixed forests, and shrub lands dominate the forested lands. Deciduous forests mainly consist of *Eucalyptus urophylla* S.T.Blake, *Schefflera actinophylla* (Endl.) Harms, *Ficus benjamina* Linn., *Acacia confusa* Merr., *Evodia glabrifolia* (Champ. ex Benth.) C.C. Huang, *Elaeocarpus chinensis* (Gardn. et Chanp.) Hook. f. ex Benth., *Cratoxylum cochinchinense* (Lour.) Bl., *Schima superba* Gardn. et Champ., *Cinnamomum camphora* (Linn.) Presl, *etc.* *Pinus massoniana* is the only coniferous tree species. In addition, in recent years, Shenzhen has been re-vegetated by planting trees along its roads and streets, expanding grass lands around the buildings and developing greened roofs. However, the city's vegetation and soil carbon stocks, the ratio of anthropogenic over natural carbon stocks and their dynamics are unknown. This study was limited to mapping the urban forest carbon density.

A Landsat 8 image dated 15 October 2014 was acquired. The image consists of eight bands, including coastal aerosol, blue, green, red, near infrared, two shortwave infrared, and cirrus band at a spatial resolution of 30 m × 30 m (Figure 1). The reasons for using the Landsat 8 image included free downloadable data and its moderate spatial resolution of 30 m × 30 m that approximated the size of sample plots used in this study. Moreover, although using fine spatial resolution images such as those from Worldview-2 can enhance the decomposition of mixed pixels in this urban landscape, the extremely high cost of acquiring the images limited their applications in this study. This is also true for mapping forest carbon density for large cities in developing countries. In addition, Landsat 8 is a recent satellite and use of a Landsat image in this study provided an opportunity to evaluate its utility for urban forest carbon modeling. Radiometric correction of the image was conducted using ATCOR2013, and pixel values were converted to reflectance values at the surface of the Earth. The image was also geo-referenced to the Universal Transverse Mercator (UTM) projection and coordinate system with root mean square error (RMSE) of less than one pixel.

The Landsat image was used to classify the entire city landscape into LULC types, including forests, grasslands, built-up areas, bare lands and water bodies. Based on the classification, a random sampling design was conducted to select sample plots for each LULC type. The number of sample plots for each LULC type was proportional to its area and the plots were then randomly located within the area of each type. The plot size was 25.82 m × 25.82 m, which was similar to the spatial resolution of the Landsat 8 image and national forest inventory sample plots. Most of the field work was carried out from 1 August to 20 November 2014, and a total of 161 plots were measured. Because about 23 sample plots located in the eastern part of the city were not measured, the eastern part of the city was not involved in this study.

The sample plots were located using a global positioning system (GPS) receiver that had a positional error of about 5 m. Tree and stand variables were measured based on national forest inventory protocols [91]. For each plot, observations of tree diameters at breast height (1.3 m above ground) (DBH) and height (H) were first collected, and tree volume was predicted using allometric models based on DBH and H by tree species for Guangdong province [91]. The uncertainties associated with the tree volume model predictions were assumed to be negligible. The volume values were then converted into biomass and carbon stock pools using biomass and carbon conversion coefficients by tree species [92]. The resulting forest carbon stocks were further divided by the plot area, leading to plot forest carbon density (Mg/ha).



**Figure 1.** Location of the study area—Shenzhen city and spatial distribution of sample plots within land use and land cover categories.

## 3. Methods

In addition to the eight original image bands, various image transformations were considered including their inversions, ratios between two or three bands, principal component analysis, grey-level co-occurrence matrix-based texture measures (mean, angular second moment, contrast, correlation, dissimilarity, entropy, homogeneity and variance), vegetation indices, and difference vegetation indices. A total of 284 spectral variables were derived (Table 1). Pearson product moment correlation coefficients between plot forest carbon density and the spectral variables were calculated. The statistical significance of the coefficients (whether they significantly differed from zero) was tested using $r_\alpha = \sqrt{t_\alpha^2/(n-2+t_\alpha^2)}$ ($n$ is number of the plots) based on the student's distribution at the significance level of 0.05.

**Table 1.** Spectral variables (SV) derived from a total of eight bands for the Landsat 8 image.

| Spectral Variables | Definitions of Spectral Variables | # of SV |
|---|---|---|
| Original $Band_i$ | Landsat 8: 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) and band$_8$-cirrus | 8 |
| Inversions of bands | $IB_i = 1/band_i \; i = 1,...8$ | 8 |
| Simple two-band ratios | $SR_{i,j} = Band_i / Band_j, \; i, j = 1,...8; i \neq j$ | 42 |
| Three-band ratios | $SR_{i,j,k} = Band_i / (Band_j + Band_k)$, $i, j, k = 1,...8; i \neq j \neq k, \; j < k$ | 106 |
| Difference vegetation indices | $DVI_{i,j} = Band_i - Band_j \; i, j = 1,...8; i \neq j$ | 42 |
| Shortwave infrared-visible band ratio | $SVR = SWIR1/[(RED + GRN)/2]$ | 1 |
| Normalized difference vegetation index | $NDVI = (NIR - RED)/(NIR + RED)$ | 1 |
| Modified normalized difference vegetation index | $MNDVI = \dfrac{NIR - RED}{NIR + RED}(1 - \dfrac{SWIR1 - SWIR1_{min}}{SWIR1_{max} - SWIR1_{min}})$ | 1 |
| Red-green vegetation index | $RGVI = (RED - GRN)/(RED + GRN)$ | 1 |
| Reduced simple ratio | $RSR = \dfrac{NIR}{RED}(1 - \dfrac{SWIR1 - SWIR1_{min}}{SWIR1_{max} - SWIR1_{min}})$ | 1 |
| Soil adjusted vegetation index | $SAVI_l = (NIR - RED)(1 + l)/(NIR + RED + l)$, l=0.1, 0.25, 0.3, 0.5 | 4 |
| Atmospherically resistant vegetation index | $ARVI = \dfrac{NIR - (2 \times RED - BLUE)}{NIR + (2 \times RED - BLUE)}$ | 1 |
| Enhanced vegetation index | $EVI = \dfrac{2.5(NIR - RED)}{NIR + 6RED - 7BLUE + 1}$ | 1 |
| Principal component analysis | The first 3 PCs from Principal component analysis (PCA) | 3 |
| Texture measures | Grey-level co-occurrence matrix-based texture measures including mean, angular second moment, contrast, correlation, dissimilarity, entropy, homogeneity and variance | 64 |

Three spatial interpolation methods, LSR, LMSR [45,46,93] and kNN [47,49–54], were used to model the relationship between plot forest carbon density and the spectral variables. LSR is a widely used method for remote sensing-based spatial interpolation of forest biomass and carbon. In this study, it was used to help select the spectral variables that contributed to statistically significantly improving the fit of a model to data and reducing the sum of squared errors. It was also used to test for nonlinearity in the relationship between urban forest carbon density and the spectral variables derived from the Landsat 8 image. The spectral variables selected as significant independent variables were then used in the backward-based linear stepwise regression.

Yan *et al.* [94] found that compared to LSR, LMSR led to more accurate estimates of carbon density for forests that were dominated by plantations. In this study, LMSR was selected for comparison with LSR and also for testing for nonlinearity in the relationship between forest carbon density for the complex urban landscape and the spectral variables derived from Landsat 8 image. For LMSR, the plot forest carbon density values were standardized before the regression so that they had a range of 0–1, and LMSR was then conducted to determine the significant spectral variables. During the stepwise regression for both LSR and LMSR, a variance inflation factor (VIF) that assesses the degree to which the variance of a regression coefficient estimate is inflated by the correlation among the independent variables selected in the model was used to analyze the multicollinearity between each pair of the selected independent variables and to eliminate spectral variables with VIF values greater than 10 [95]. In this study, the correlation coefficients among the independent variables varied from 0.03 to 0.99, which indicated high correlations among at least some pairs of variables. This led to large variances and instability for the regression coefficient estimates and VIF values, and thus reduced the effectiveness of stepwise techniques. In this study, testing of VIF showed that a VIF value exceeding 10 was a good indicator of serious multicollinearity [95]. The regressions were carried out using the R statistical software package.

The significant spectral variables from both LSR and LMSR were also used with kNN to measure the spectral distances of an unknown location from all the sample plots based on Euclidean distance and a neighbor weighting scheme. Results were compared for k = 3, 5, 7 and 10 nearest neighbors. kNN-based estimation was carried out using a computer program developed by the authors in IDL. Because kNN directly measures spectral distances and selects the given number of the nearest neighbors in a spectral space for each location whose value is unknown, it is a more intuitive method than LSR and LMSR and has been widely used in the Nordic countries to map forest types and estimate tree volumes and biomass/carbon [47,49–54]. In this study, we tested if kNN was appropriate for mapping urban forest carbon density.

LSUA, with the constraint that the sum of fractions for all endmembers must equal one, was conducted for the study area using the fully constrained least square linear spectral unmixing tool. Four endmembers were determined: vegetation, urbanized area, water and bare soil. First, the minimum noise fraction (MNF) transformation of Landsat 8 image bands was conducted [86–88], and the endmembers were then selected based on the scatter plots of MNF.

LSUA was combined with the aforementioned LSR, LMSR and kNN by introducing the vegetation fraction into the spatial interpolation methods. The significant spectral variables selected above together with the vegetation fraction were used as the independent variables in the models. The accuracies of the predicted forest carbon density values from the combinations were compared with those from the models that did not include the vegetation fraction. The combinations of the spatial interpolation methods LSR, LMSR and kNN with LSUA and the comparisons with the methods in which the vegetation fraction is not involved are expected to lead to improved estimation. Theoretically, this approach is appropriate for mapping the forest carbon density of complex urban vegetation landscapes having large numbers of mixed pixels, but no reports are known to have been published.

Accuracy assessments of the resulting forest carbon density estimates and maps were conducted using cross-validation. With this method, a sample plot was first randomly selected and removed from the dataset, the remaining plots were used to develop the models, the estimate of forest carbon density for the

removed plot was calculated, and the deviation between the predicted and observed values was computed. The removed plot was then replaced and another plot was randomly selected and removed from the dataset. The corresponding modeling and estimation were conducted for this plot. The process was repeated until all the sample plots yielded their estimates. Using the observations and predictions for the sample plots, the coefficient of determination $R^2$ and RMSE were then computed. In addition, this study led to maps that served mainly as intermediate products that supported increasing the accuracy of estimates. However, the maps can be also used as stand-alone products that depict the spatial distribution of Shenzhen urban forest carbon density. Thus, the evaluation with respect to the final applications of the maps was conducted by calculating the mean density of estimated forest carbon and its variance for each of the maps using model-assisted regression estimators [51].

## 4. Results

A statistical summary of the sample plot data used for mapping urban forest carbon density of Shenzhen is shown in Table 2. The plot-level sample mean and standard deviation were estimated using simple random sampling estimators. The resulting urban forest carbon density estimate was characterized by a large coefficient of variation with a confidence interval of 14.92 Mg/ha–22.15 Mg/ha determined at the significance level of 0.05.
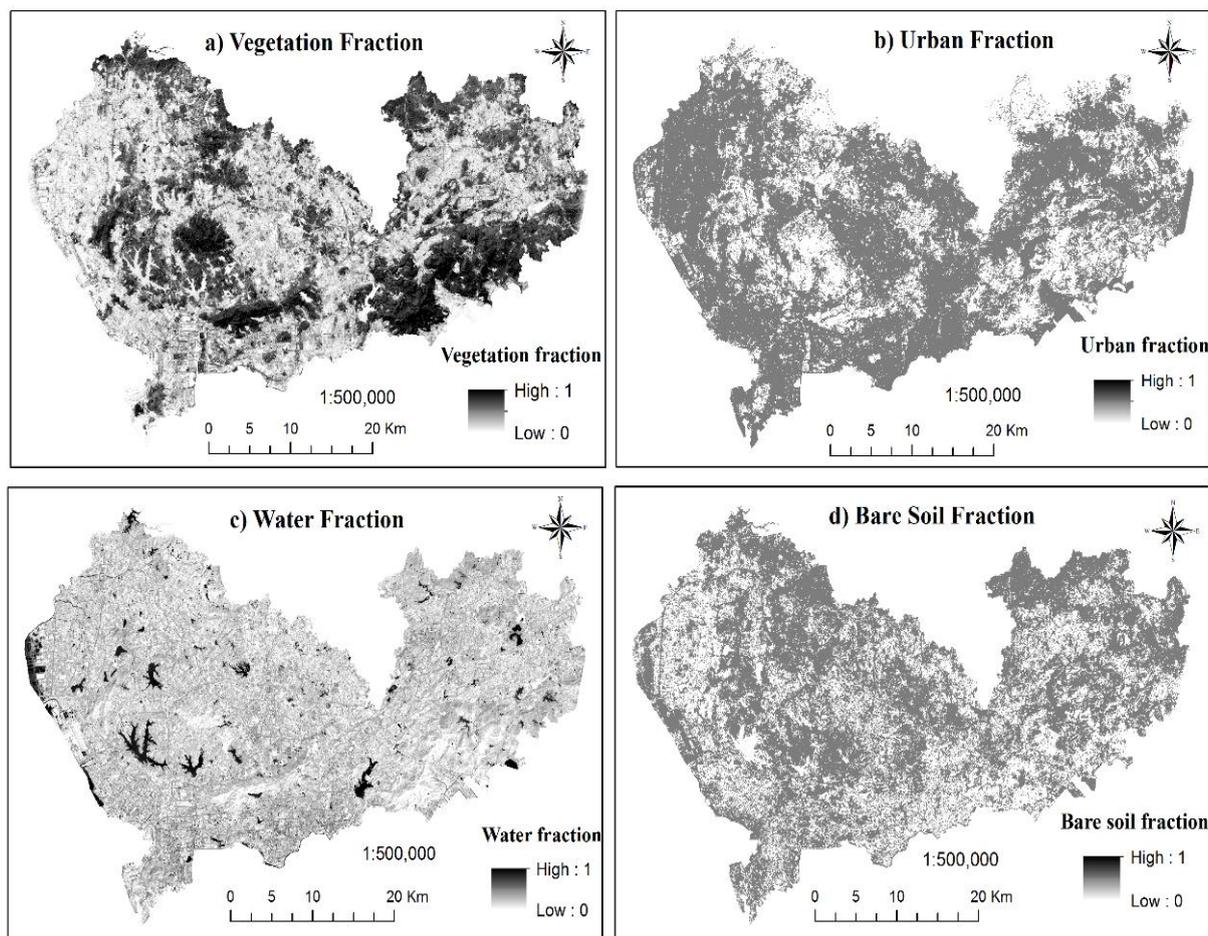
**Table 2.** The statistical summary of sample plot data used for mapping urban forest carbon density of Shenzhen.

| Number of Plots | Minimum (Mg/ha) | Maximum (Mg/ha) | Sample Mean (Mg/ha) | Standard Deviation (Mg/ha) | Coefficient of Variation (%) |
|---|---|---|---|---|---|
| 161 | 0 | 100.65 | 18.54 | 23.33 | 125.88 |

The Pearson product moment correlation coefficients between plot forest carbon density and the 284 spectral variables were calculated. Among them, 241 spectral variables had absolute values of the coefficients varying from 0.156 to 0.655 and significantly differed from zero at the significance level of 0.05. The coefficients of the 60 most correlated spectral variables ranged from 0.557 to 0.655. Atmospherically resistant vegetation index (ARVI) had the greatest correlation with forest carbon density. The widely used vegetation indices, including normalized difference vegetation index (NDVI), modified NDVI (MNDVI), reduced simple ratio (RSR), soil adjusted vegetation indices (SAVI$_i$, i = 0.1, 0.25, 0.3 and 0.5), enhanced vegetation index (EVI), had correlation coefficients of 0.586–0.647 and were included among the 60 most correlated spectral variables. All other spectral variables came from simple two- and three-band ratios (SR$_{i,j}$ and RS$_{i,j,k}$). Compared to the band ratios and vegetation indices, the complicated image transformations PCA and grey-level co-occurrence matrix-based texture measures resulted in smaller coefficients of correlation.

LSUA for the study area was conducted using the spectral variables derived from Landsat 8 image bands (Figure 2). Vegetation was mainly distributed in the southeast, southwest central, northeast and northwest parts of Shenzhen city, in which urban fraction had smaller estimates (Figure 2a,b). Bare soil was almost evenly scattered over the city except for the west and south parts dominated by urbanized area (Figure 2d). LSUA accurately detected the water bodies for which the values of vegetation, urban

and bare soil fractions were close to zero (Figure 2c). The resulting vegetation fraction had a correlation coefficient of 0.600 with the plot urban forest carbon density.



**Figure 2.** Fraction images of Shenzhen urban landscape from the linear spectral unmixing analysis (LSUA): (**a**) vegetation fraction, (**b**) urban fraction, (**c**) water fraction, and (**d**) bare soil fraction.

The linear model LSR led to four spectral band ratios that made significant contributions to the estimation of Shenzhen urban forest carbon density (Equation (1)), including $SR_{215}$, $SR_{315}$, $SR_{24}$, and $SR_{34}$, indicating that the visible bands (near infrared, red, green and blue) and coastal aerosol were the most important. This model had a statistical F-value of 40.56 with a p-value of zero. Although ARVI had the highest correlation with the plot forest carbon density, it was excluded from the LSR model mainly because it was highly correlated with all other selected spectral variables with absolute coefficients of correlation varying from 0.807 to 0.950. The resulting estimates of the urban forest carbon density varied from −53 Mg/ha–157 Mg/ha with an average estimate of 18.56 Mg/ha for the sample plots (Table 3), almost the same as the sample mean of 18.54 Mg/ha. However, the negative carbon density estimates were obviously illogical and there were also some unreasonable estimates that were much larger than the maximum observation of 100.65 Mg/ha. Based on cross-validation, the coefficient of determination and RMSE between the estimated and observed values for the sample plots were 0.482 and 16.67 Mg/ha. After the vegetation fraction from LSUA was introduced into LSR as an independent

variable, the integration (Equation (2)) of LSR with LSUA resulted in urban forest carbon density estimates that ranged from −55 Mg/ha–225 Mg/ha with an average estimate of 18.76 Mg/ha for the sample plots, not significantly different from the sample mean (Table 3). Moreover, the integration increased the coefficient of determination to 0.514 and reduced the RMSE to 16.13 Mg/ha. Compared to the results for LSR without the vegetation fraction, the correlation increase and RMSE decrease were 6.6% and 3.3%, respectively. The estimates of the mean obtained from the carbon density maps based on both LSR and its integration with LSUA were slightly smaller than the sample mean, but were included in the confidence interval at the 0.05 significance level (Table 3). Model (2) had a statistical F-value of 32.90 with a p-value of zero.

$$\hat{Y} = 12.847 - 493.602 SR_{215} + 378.587 SR_{315} + 207.412 SR_{24} - 167.151 SR_{34} \tag{1}$$

$$\hat{Y} = -0.755 - 562.751 SR_{215} + 498.171 SR_{315} + 240.889 SR_{24} - 212.644 SR_{34} + 27.394 VF \tag{2}$$

**Table 3.** The accuracy assessments of Shenzhen urban forest carbon density estimates based on cross-validation and the evaluation of map applications by estimating the mean density for each forest carbon density map and its variance using model-assisted regression estimators from the linear stepwise regression (LSR) and the logistic model-based stepwise regression (LMSR) and their integration with linear spectral unmixing analysis (LSUA), that is, LSR&LSUA, LMSR&LSUA (Note: The mean is the average estimate of the observed locations, $R^2$ and RMSE are the coefficient of determination and root mean square error between the estimated and observed values of the sample plots, $\mu_{map}$ and $Var_{map}$ are the mean density of a forest carbon density map and its variance using model-assisted regression estimators, and the confidence interval at the significance level of 0.05 was 14.92 Mg/ha–22.15 Mg/ha).

| Method | Mean (Mg/ha) | $R^2$ | RMSE (Mg/ha) | $\mu_{map}$ | $Var_{map}$ |
|---|---|---|---|---|---|
| LSR | 18.56 | 0.514 | 16.67 | 17.74 | 1.78 |
| LSR&LSUA | 18.76 | 0.514 | 16.13 | 17.96 | 1.67 |
| LMSR | 17.16 | 0.537 | 17.72 | 15.73 | 2.01 |
| LMSR&LSUA | 17.99 | 0.578 | 17.08 | 16.57 | 1.87 |

The non-linear model LMSR also selected four spectral variables that made significant contributions to improving the fit of a model to the data and reducing the sum of squared errors for estimating Shenzhen urban forest carbon density (Equation (3)), including ARVI, $SR_{25}$, $SR_{32}$, and $SR_{34}$. Most of the variables differed from those selected by LSR. However, three simple band ratios were still selected. Both LMSR and its integration with LSUA resulted in a reasonable range of estimates from 0 Mg/ha–100.65 Mg/ha. For LMSR without the vegetation fraction, the resulting coefficient of determination was 0.537, greater than those from LSR with and without the vegetation fraction (Table 3). The RMSE was 17.72 Mg/ha and increased by 6.3% and 9.9%, respectively, compared to those from LSR with and without the vegetation fraction. For LMSR with the vegetation fraction (Equation (4)), the resulting coefficient of determination and RMSE were 0.578 and 17.08 Mg/ha, and increased by 7.6% and decreased by 3.6%, respectively, compared to those from LMSR without the vegetation

fraction. Compared to those from the integration of LSR with LSUA, the combination of LMSR with LSUA increased both the coefficient of determination and RMSE by 12.5% and 5.9%, respectively. For both LMSR and its integration with LSUA, the average estimates of the forest carbon density for both the sample plots and the resulting maps were smaller than the sample mean 18.53 Mg/ha, but were in the confidence interval (Table 3).

$$\hat{Y} = \frac{100.65\exp(15.685 - 3.310\mathrm{ARVI} - 7.594SR_{25} - 13.599SR_{32} + 1.937SR_{34})}{(1 + \exp(15.685 - 3.310\mathrm{ARVI} - 7.594SR_{25} - 13.599SR_{32} + 1.937SR_{34}))} \tag{3}$$

$$\hat{Y} = \frac{100.65\exp(15.984 - 3.587\mathrm{ARVI} - 7.442SR_{25} - 13.919SR_{32} + 1.844SR_{34} + 0.480\mathrm{VF})}{(1 + \exp(15.984 - 3.587\mathrm{ARVI} - 7.442SR_{25} - 13.919SR_{32} + 1.844SR_{34} + 0.480\mathrm{VF}))} \tag{4}$$

Shenzhen urban forest carbon density was also mapped using kNN with four different numbers of nearest neighbors, k = 3, 5, 7 and 10. When the significant variables from LSR were used to calculate the spectral distance between each of the sample plots and an unobserved location for kNN, the results from cross-validation are listed in Table 4. The individual estimates varied from 4.00 Mg/ha to 100.60 Mg/ha. For kNN with and without the vegetation fraction from LSUA, the average estimates of the forest carbon density for the sample plots and the resulting maps ranged from 18.10 Mg/ha to 19.03 Mg/ha and all were in the confidence interval at the 0.05 significance level (Table 4). For the kNN method without the vegetation fraction, as the number of the nearest neighbors increased, the coefficients of determination increased from 0.298 to 0.388 and the RMSE decreased from 20.51 Mg/ha to 18.37 Mg/ha. Compared to the results for kNN without the vegetation fraction, adding the vegetation fraction from LSUA into the set of the significant spectral variables led to increases in the estimation accuracy. The increase for the coefficient of determination varied from 1.7% to 9.3% and the decrease for the RMSE ranged from 1.0% to 3.3%, depending on the numbers of nearest neighbors used. Using k = 10 nearest neighbors resulted in the largest estimation accuracy for kNN, regardless of whether the vegetation fraction was added. However, compared to the results from LSR and LMSR with and without the vegetation fraction, kNN produced larger values of RMSE (Tables 3 and 4).
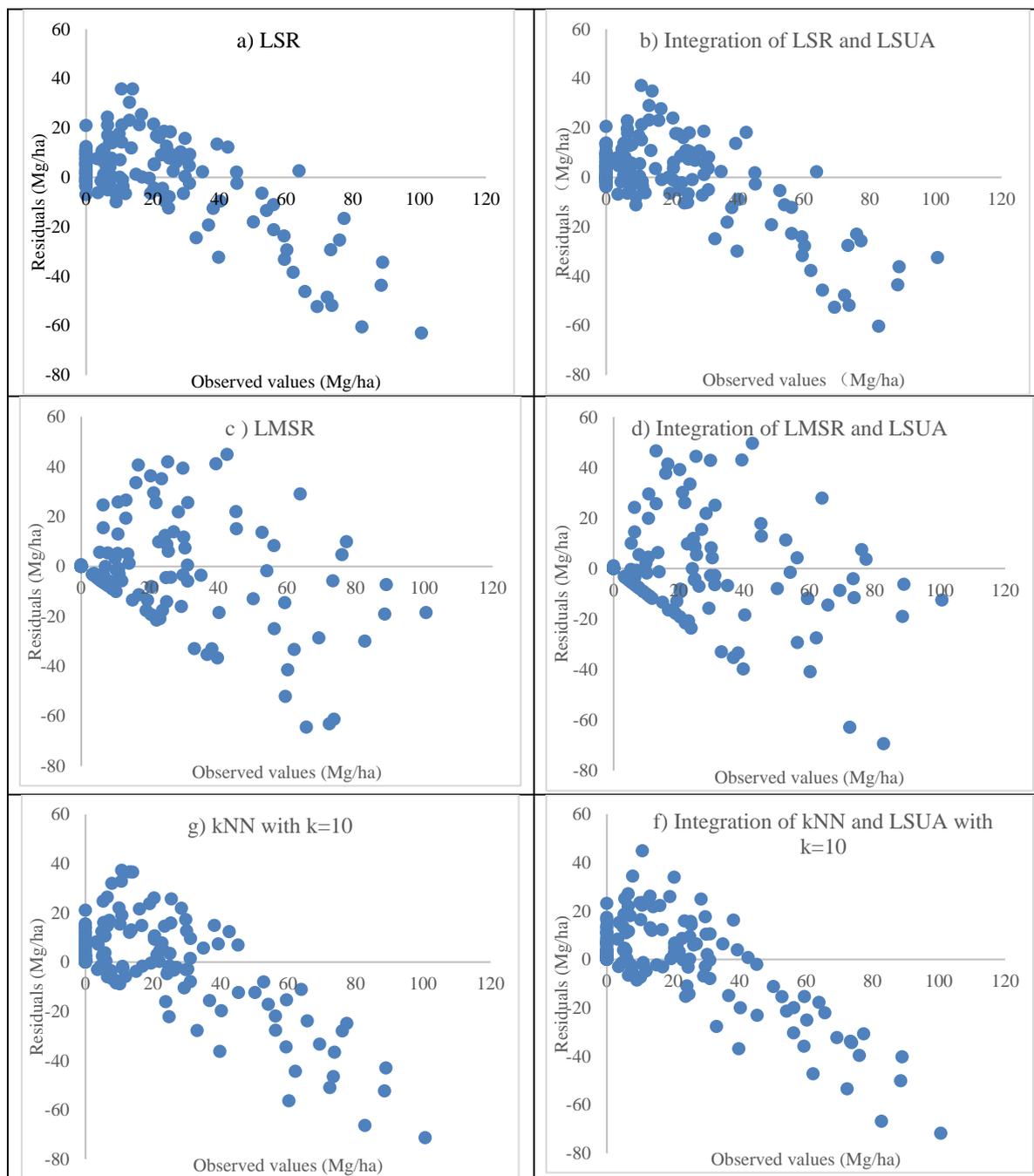
When the significant spectral variables from LMSR were used for kNN, both with and without the vegetation fraction from LSUA, all the average estimates of forest carbon density for the sample plots and the resulting maps were in the confidence interval at the 0.05 significance level (Table 4). The accuracies of the estimates based on cross-validation were slightly greater than those for kNN using the significant variables from LSR. However, they were smaller than those using LMSR, regardless of whether the vegetation fraction was or was not used. With increasing numbers of nearest neighbors, the coefficient of determination also increased and RMSE decreased, indicating that 10 nearest neighbors led to the greatest accuracy. Introducing the vegetation fraction into kNN increased the coefficients of determination by 5.4%–6.6% and decreased the RMSE by 2.1%–3.1%.

In Figure 3, the residuals for the predicted values for urban forest carbon density were graphed against the observed values for LSR, LMSR and KNN with and without the vegetation fraction from LSUA. For kNN, only the results with the significant variables from LMSR and k = 10 were presented. Obviously, the residuals showed a trend of linear decrease as the observed values increased for both LSR and kNN, which indicates that the smaller forest carbon density values were overestimated, while the greater ones were underestimated (Figure 3). The trend was improved for LMSR with and without the vegetation fraction. Moreover, all the methods led to similar maps of Shenzhen urban forest carbon density
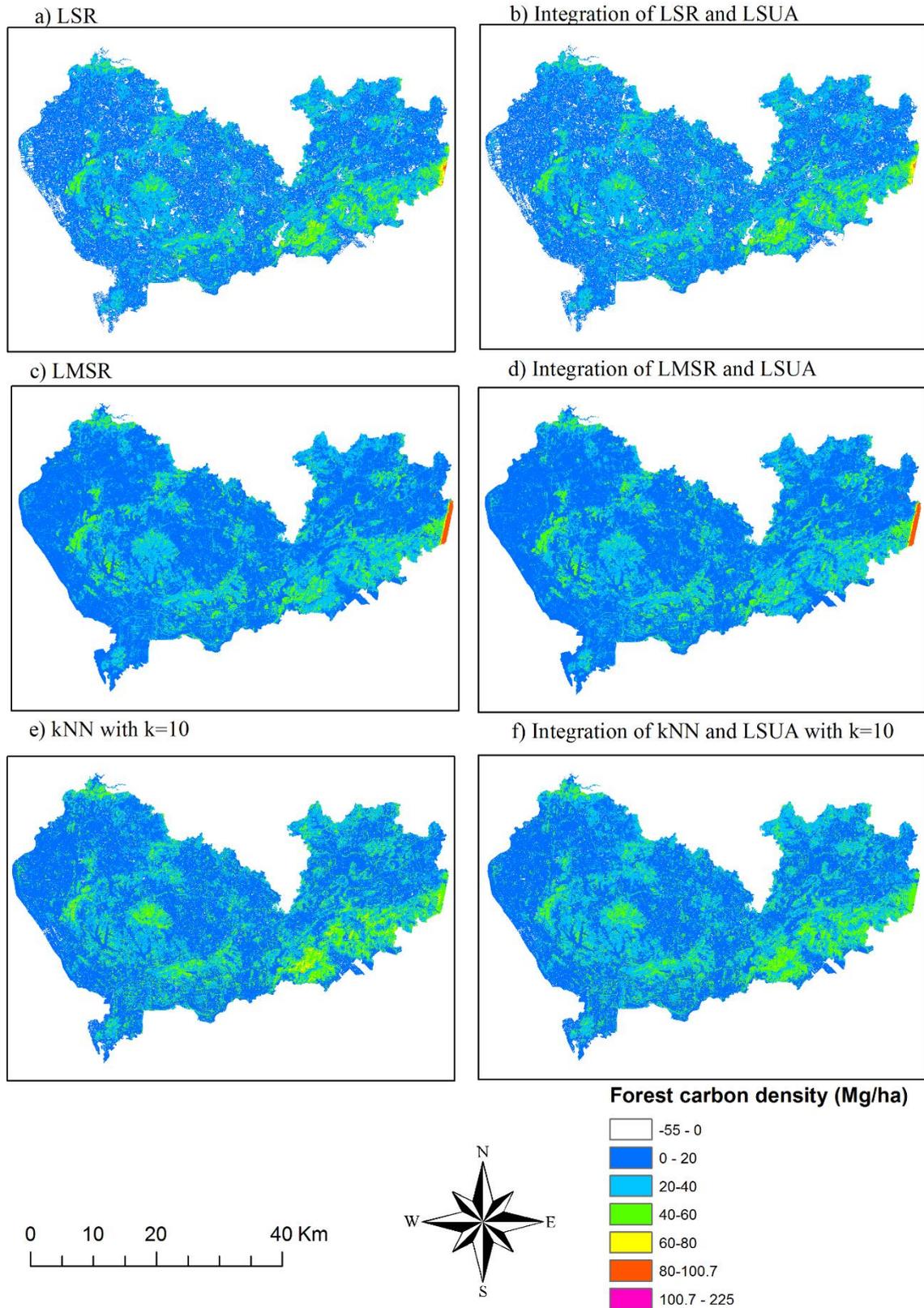
(Figure 4). The carbon density showed greater estimates in the southeast, west central, north and south central parts of the city than the other areas. The spatial distributions were similar to the spatial patterns of the vegetated areas demonstrated by the false color composite image of Landsat 8 in Figure 1. At the eastern border of the study area, very large forest carbon density estimates occurred in a linear pattern. This was mainly caused by noise due to the clouds existing in this portion of the image.

**Table 4.** Accuracy assessments for the Shenzhen urban forest carbon density estimates based on cross-validation and the evaluation of map applications by calculating the mean density of each forest carbon density map and its variance using model-assisted regression estimators from k-Nearest Neighbors (kNN) and the significant spectral variables from both the linear stepwise regression (LSR) and the logistic model-based stepwise regression (LMSR), and their integrations with the linear spectral unmixing analysis (LSUA) (Note: k is the number of nearest neighbors, mean is the average estimate of observed locations, $R^2$ and RMSE are the coefficient of determination and root mean square error between the estimated and observed values of the sample plots, $\hat{\mu}_{map}$ and $Var_{map}$ are the mean density of a forest carbon density map and its variance using model-assisted regression estimators, and the confidence interval of 14.92 Mg/ha–22.15 Mg/ha determined at the significance level of 0.05.

| k | Mean (Mg/ha) | $R^2$ | RMSE (Mg/ha) | $\hat{\mu}_{map}$ | $Var_{map}$ |
|---|---|---|---|---|---|
| kNN based on the significant spectral variables from LSR | | | | | |
| 3 | 18.26 | 0.298 | 20.51 | 19.03 | 2.70 |
| 5 | 18.37 | 0.333 | 19.54 | 18.98 | 2.45 |
| 7 | 18.49 | 0.356 | 19.02 | 18.94 | 2.32 |
| 10 | 18.41 | 0.388 | 18.37 | 18.79 | 2.16 |
| Integration of kNN based on the significant spectral variables from LSR with LSUA | | | | | |
| 3 | 18.11 | 0.303 | 20.32 | 18.66 | 2.65 |
| 5 | 18.35 | 0.363 | 18.94 | 18.83 | 2.31 |
| 7 | 18.28 | 0.389 | 18.40 | 18.62 | 2.17 |
| 10 | 18.10 | 0.397 | 18.19 | 18.33 | 2.12 |
| kNN based on the significant spectral variables from LMSR | | | | | |
| 3 | 18.32 | 0.322 | 20.00 | 18.97 | 2.56 |
| 5 | 18.57 | 0.354 | 19.14 | 19.07 | 2.35 |
| 7 | 18.31 | 0.378 | 18.56 | 18.71 | 2.21 |
| 10 | 18.30 | 0.394 | 18.22 | 18.56 | 2.13 |
| Integration of kNN based on the significant spectral variables from LMSR with LSUA | | | | | |
| 3 | 18.02 | 0.343 | 19.39 | 18.75 | 2.41 |
| 5 | 18.26 | 0.373 | 18.61 | 18.97 | 2.22 |
| 7 | 18.14 | 0.403 | 18.07 | 18.66 | 2.09 |
| 10 | 18.29 | 0.415 | 17.84 | 18.61 | 2.04 |

**Figure 3.** The residuals for the predicted forest carbon density values graphed against the observed values using (**a**) the linear stepwise regression (LSR); (**b**) the integration of LSR and the linear spectral unmixing analysis (LSUA) by adding the vegetation fraction into LSR; (**c**) the logistic model-based stepwise regression (LMSR); (**d**) the integration of LMSR and LSUA; (**e**) the k-Nearest Neighbors (kNN) with k = 10 using the significant variables from LMSR; and (**f**) the integration of kNN and LSUA with k = 10.

**Figure 4.** The spatial distributions of the predicted forest carbon density values using (**a**) the linear stepwise regression (LSR); (**b**) the integration of LSR and the linear spectral unmixing analysis (LSUA); (**c**) the logistic model-based stepwise regression (LMSR); (**d**) the integration of LMSR and LSUA; (**e**) the k-Nearest Neighbors (kNN) with k = 10 using the significant variables from LMSR; and (**f**) the integration of kNN and LSUA with k = 10.

## 5. Discussion

Urban forests play an important role in the mitigation of carbon concentration in the atmosphere, the reduction of human activity induced air pollution and noise and the control of global climate change [56–70,77–81]. Accurately estimating and mapping urban forest carbon density has thus become very important. However, a great challenge for this purpose is selecting or developing methods for extracting vegetation fraction from the mixed pixels of complex urban landscapes and then introducing the information into spatial modeling [81]. The results of this study showed that by comparing three spatial interpolation methods—LSR, LMSR and kNN—without the vegetation fraction, introduction of the vegetation fraction images derived using LSUA into the modeling methods increased the accuracy of mapping Shenzhen urban forest carbon density by 1.0%–9.3%, depending on the used methods. This indicates that extraction and use of the vegetation fraction has considerable potential for increasing the accuracy of mapping the urban forest carbon density. This finding was supported the study of Yan *et al.* [94] in which increases in accuracy of 31%–39% were achieved by adding the vegetation fraction information into a regression model for mapping the carbon density of general or non-urban forests dominated by pure plantations. Comparatively, the increase in estimation accuracy for this study was relatively smaller, mainly because Shenzhen had a more complex landscape and more mixed pixels than the non-urban forests.

For this study, a limited budget precluded acquisition and use of fine spatial resolution images such as from Worldview-2; instead, a new and free downloaded Landsat 8 image was used. This image had a spatial resolution of 30 m $\times$ 30 m, which was relatively too coarse for the decomposition of the mixed pixels in the complex Shenzhen city landscape. The use of fine spatial resolution images may increase the accuracy of unmixing the mixed pixels and, thus, the accuracy of estimating the urban forest carbon density. In this study, moreover, the lack of fine spatial resolution images or land use and land cover maps limited the accuracy assessment of the vegetation fraction from LSUA. In addition, the lack of fine spatial resolution images also made it impossible to separate the vegetated areas of Shenzhen city into forested lands, shrub lands and grasslands, mainly because, in this city, the trees and shrubs were often mixed together, and the grasslands were in small segments around buildings and bare lands. Readers should thus use the conclusions drawn in this study with caution.

On the other hand, mapping urban forest carbon density requires a higher cost compared to the generation of a carbon density map for non-urban forests for the same size with the same estimation accuracy. The cost increase results from the requirement for more field observations because of more complex urban landscapes and higher spatial and temporal variability of urban forest carbon density. For example, the coefficient of variation of 125.9% for the carbon density of the urban forests in Shenzhen was very large. For accuracy expressed as a relative error of 15% or less, a total of 161 sample plots were selected using a random sampling design at a 0.05 significance level. For the carbon density of non-urban forests, the coefficient of variation generally varies from 60% to 100% and a total 35–100 sample plots is required. Thus, the increase of cost and time to collect and process the additional sample plot data is about 60%–360%. Moreover, free Landsat images at a spatial resolution of 30 m $\times$ 30 m are often adequate for mapping the carbon density of non-urban forests. For urban forests, remotely sensed images at spatial resolutions finer than 5 m $\times$ 5 m may be needed to unmix the mixed pixels. The cost to acquire the images varies depending on the spatial resolution of the images. For

instance, the image costs for Shenzhen city are $4688 for a RapidEye basic product at a 5 m ×5 m spatial resolution, $18,864 for SPOT 6 pan-sharpened data at a 1.5 m × 1.5 m spatial resolution, and $47,619 for Pleiades pan-sharpened images and $58,608 for Worldview-2 at a 0.5 m ×0.5 m spatial resolution. The extremely high cost often makes it impossible to use the fine spatial resolution images to map forest carbon density for large cities in developing countries. Thus, using a free Landsat 8 image could greatly reduce the image cost. Instead, the estimation accuracy may decrease, but the amount of decrease is unknown and a further study is necessary.

The results of RMSE from cross-validation showed that LSR with and without the vegetation fraction led to the smallest RMSE values, followed by LMSR with and without the vegetation fraction, and kNN with the significant spectral variables from LMSR and the vegetation fraction together. The largest RMSE came from kNN with the significant variables from LSR and without the vegetation fraction. However, LSR, regardless of whether the vegetation fraction was used, resulted in not only illogically negative estimates, but also unreasonably large estimates. Moreover, in this study, urban forest carbon density and all the selected spectral variables were characterized by skewed instead of normal distributions. This required transformation of the data to normality. Thus, a logarithm transformation, which is most appropriate for skewed distributions [96], was applied to the data before the linear regressions. However, the results showed that LSR with and without the vegetation fraction led to average estimates of the urban forest carbon density that were outside the confidence interval obtained using simple random sampling design at the significance level of 0.05 for both the sample plots and the resulting maps with extremely large estimates. This indicated that, in practice, it is not always necessary or desirable to transform a data set to achieve normality [96]. This finding was consistent with both Fleming *et al.* [45] and Wang *et al.* [46].

In addition, both LSR and kNN and their integration with LSUA led to a linearly decreasing trend of the residuals over increasing observations of urban forest carbon density, *i.e.*, overestimates for the smaller observations and underestimates for larger observations. The over- and underestimates were mainly caused by the non-linear relationships between urban forest carbon density and the selected spectral variables. The linearly decreasing trend of the residuals was not apparent for the non-linear model LMSR with and without the vegetation fraction. This indicated that the non-linear logistic model was a better choice than the linear LSR and kNN techniques for modeling the relationship between urban forest carbon density and the spectral variables and improving estimation of Shenzhen urban forest carbon density.

In this study, the resulting relative errors of estimating the urban forest carbon density were 87.0% and 89.9% for LSR, 92.1% and 95.6 for LMSR, and 96.2%−110.6% for kNN with and without the vegetation fraction, respectively. Compared to those from Yan *et al.* [94], these relative errors were much larger. This was mainly caused by the larger coefficient of variation of 125.9% for the urban forest carbon density in the complex Shenzhen landscape and the greater number of mixed pixels compared to the study of Yan *et al.* [94] in which the study area was dominated by the pure plantations. This was supported by the findings from Davies *et al.* [74] and Pataki *et al.* [75]. This indicated greater potential for further increasing the accuracy of mapping urban forest carbon density by improving the decomposition of mixed pixels using finer spatial resolution images.

Studies in the Nordic countries and the USA have shown that kNN is a powerful tool for mapping forest types and estimating tree volumes and biomass by combining national forest inventory sample

plot data and remotely sensed images [47,49–54]. However, its performance for mapping Shenzhen urban forest carbon density in this study produced a much smaller accuracy than that using the non-linear regression LMSR. The reason was mainly because kNN is based on the spectral distance between pure pixels [88–90], whereas the complex Shenzhen urban vegetation landscape has a large number of mixed pixels with a variety of structures and compositions of endmembers. The spectral reflectance values of the mixed pixels varied greatly depending on their inherent structures and compositions. Thus, the resulting spectral distances between the mixed pixels did not correspond well with those for pure pixels. On the other hand, shadows in the urban landscape may also affect the representativeness of the spectral distance to the similarity among pixels.

Urbanization causes LULC changes and, thus, changes in the total carbon stock and its pools of cities. Based on the study by Scalenghe *et al.* [97], in a populated area the proportion of anthropogenic carbon stock increased and was accompanied by a large increase in the carbon emission per capita per unit area, while the proportions of soil and vegetation carbon stock decreased. Moreover, Villa *et al.* [27] also quantified the ratio of anthropogenic to natural carbon stocks for a peri-urban area in Italy using Landsat multi-temporal data. However, this study concentrated only on the estimation and mapping of the forest carbon density in Shenzhen and did not deal with assessing the impacts of urbanization on the amounts of other carbon pool stocks because of a lack of data. In the future, anthropogenic (buildings, roads, impervious surface, *etc.*), soil and vegetation carbon stocks, carbon emission and their dynamics due to urbanization, sprawl patterns, surface imperviousness and soil sealing, should be studied.

This study demonstrated the potential of combining spatial interpolation methods and spectral unmixing analysis for improving mapping and estimation of urban forest carbon density. Further studies will focus on the use of fine spatial resolution images and spectral unmixing analysis to separate vegetated areas into vegetation types such as forests, individual trees along streets, shrub lands and grasslands. Moreover, selecting or developing methods for removing or reducing the effects of building-induced shadows on the estimation of urban forest carbon density is another great challenge. In addition, airborne LiDAR data or the data from terrestrial 3D laser scanning and their data fusion with optical images may also offer great potential for increasing the accuracy of forest carbon density estimation [72,73].

## 6. Conclusions

In this study, a novel approach was proposed to map the urban forest carbon density of Shenzhen, a rapidly sprawling city in southern China. In this approach, three spatial interpolation methods—LSR, LMSR and kNN—were integrated with LSUA by introducing the derived vegetation fraction image into the modeling methods to increase estimation accuracy. The results led to the following conclusions: (1) Adding the vegetation fraction from LSUA into the linear and non-linear regressions and kNN increased the accuracy of mapping Shenzhen urban forest carbon density by 1.0%−9.3% compared to the methods without the vegetation fraction; (2) All the average estimates of urban forest carbon density from these methods were in the confidence interval estimated using only the sample plot data at the significance level of 0.05. However, LSR resulted in illogically negative and extremely large estimates, and kNN produced greater RMSEs than LSR and LMSR; (3) Integrating LMSR and LSUA improved the systematical over and underestimations caused by LSR and KNN for the smaller and greater urban

forest carbon density values, respectively. The integration also led to reasonable estimates with a RMSE value that was much smaller than those from kNN with and without the vegetation fraction. Thus, combining the spatial modeling method LMSR and the spectral unmixing analysis LUSA coupled with Landsat images is most promising for increasing the accuracy of mapping urban forest carbon density. In addition, this method is characterized by the great potential for accurate, rapid and nondestructive prediction of urban and peri-urban forest carbon stocks with an acceptable level of error and low cost.

## Acknowledgments

## Author Contributions

G.W., H.S., G.Q., Y.T. and J.L. conceived and designed the experiments; Y.P., H.S, G.Q., Y.T. and Z.M. performed the experiments; H.S, G.Q., G.W. J.L. and C.L. analyzed the data; H.S, G.Q. and G.W. wrote the paper.

## 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.

## References

1.  Churkina, G.; Brown, D.; Keoleian, G. Carbon stored in human settlements: The conterminous United States. *Glob. Change Biol.* **2010**, *16*, 135–143.
2.  Strohbach, M.W.; Arnold, E.; Haase, D. The carbon footprint of urban green space-A life cycle approach. *Landsc. Urban Plan.* **2012**, *104*, 220–229.
3.  Fang, J.; Chen, A.; Peng, C.; Zhao, S.; Ci, L. Changes in forest biomass carbon storage in China between 1949 and 1998. *Science* **2002**, *292*, 2320–2322.
4.  Grace, J.; Lloyd, J.; Mcintyre, J.; Miranda, A.C. Carbon dioxide uptake by an undisturbed tropical rain forest in Southwest Amazonia, 1992 to 1993. *Science* **1995**, *270*, 778.
5.  Schimel, D.S.; House, J.I.; Hibbard, K.A.; Bousquet, P.; Ciais, P.; Peylin, P.; Braswell, B.H.; Apps, M.J.; Baker, D.; Bondeau, A.; *et al.* Recent patterns and mechanism of carbon exchange by terrestrial ecosystems. *Nature* **2001**, *414*, 169–172.

6. Smith, J.E.; Heath, L.S. Carbon stocks and stock changes in U.S. forests and Appendix C. In *U.S. Agriculture and Forestry Greenhouse Gas Inventory: 1990–2005*; Technical Bulletin No. 1921; Global Change Program Office, Office of the Chief Economist: Washington, DC, USA, 2008; Available online: http://www.usda.gov/oce/global_change/AFGGInventory1990_2005.htm/ (accessed on 28 August 2015).

7. US Climate Change Science Program. The North American Carbon Budget and Implications for the Global Carbon Cycle. Available online: http://www.cfr.org/climate-change/north-american-carbon-budget-implications-global-carbon-cycle/p14868/ (accessed on 28 August 2015).

8. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Moran, E. A survey of remote sensing-based aboveground biomass estimation methods. *Int. J. Digit. Earth* **2014**, doi:10.1080/17538947.2014.990526

9. Wang, G.; Oyana, T.; Zhang, M.; Adu-Prah, S.; Zeng, S.; Lin, H.; Se, J. Mapping and spatial uncertainty analysis of forest vegetation carbon by combining national forest inventory data and satellite images. *For. Ecol. Manage.* **2009**, *258*, 1275–1283.

10. Lu, D. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* **2006**, *27*, 1297–1328.

11. Intergovermental Panel on Climate Change (IPCC). *Land Use, Land-Use Change and Forestry*; Cambridge University Press: Cambridge, UK, 2000.

12. Yang, H.; Zhou, Q.; Chen, Z.; Zhang, S. Estimation methods and application analysis of forest carbon stock. *J. Geoinf. Sci.* **2007**, *9*, 5–12. (In Chinese)

13. Zhou, X.; Wang, Z.; Zhu, Q. A Review of global climate change and forest carbon sequestration. *Shanxi For. Sci. Tech.* **2011**, *2*, 47–52.

14. Fu, D.; Chen, B.; Zhang, H.; Wang, J.; Black, A.; Amiro, B.D. Estimating landscape net ecosystem exchange at high spatial-temporal resolution based on Landsat data, an improved upscaling model framework, and eddy covariance flux measurements. *Remote Sens. Environ.* **2014**, *141*, 90–104.

15. Landsberg, J.J.; Waring, R.H. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance, and partitioning. *For. Ecol. Manage.* **1997**, *95*, 209–228.

16. Running, S.W.; Coughlan, J.C. A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecol. Model.* **1988**, *42*, 125–154.

17. Kimball, J.S.; Keyser, A.R.; Running, S.W.; Saatch, S.S. Regional assessment of boreal forest productivity using an ecological process model and remote sensing parameter maps. *Tree Physiol.* **2000**, *20*, 761–775.

18. Running, S.W.; Hunt, E.R., Jr. Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global scale models. In *Scaling Physiological Processes: Leaf to Globe*; Ehleringer, J.R., Field, C.B., Eds.; Academic Press, Inc.: San Diego, CA, USA, 1993; pp. 141–158.

19. Neumann, M.; Zhao, M.; Kindermann, G.; Hasenauer, H. Comparing MODIS net primary production estimates with terrestrial national forest inventory data in Austria. *Remote Sens.* **2015**, *7*, 3878–3906.

20. Lu, D.; Chen, Q.; Wang, G.; Moran, E.; Batistella, M.; Zhang, M.; Laurin, G.V.; Saah, D. Estimation and uncertainty analysis of aboveground forest biomass with Landsat and LiDAR data: Brief overview and case studies. *Int. J. For. Res.* **2012**, *1*, 1–16.

21. Tian, X.; Su, Z.; Chen, E.; Li, Z.; van der Tol, C.; Guo, J.; He, Q. Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. *Int. J. Appl. Earth Observ. Geoinf.* **2012**, *14*, 160–168.

22. Wang, G.; Zhang, M.; Gertner, G.Z.; Oyana, T.; McRoberts, R.E.; Ge, H. Uncertainties of mapping forest carbon due to plot locations using national forest inventory plot and remotely sensed data. *Scand. J. For. Res.* **2011**, *26*, 360–373.

23. Wang, G.; Zhang, M. Upscaling with conditional co-simulation for mapping aboveground forest carbon. In *Scale Issue in Remote Sensing*; Weng, Q., Ed.; John Wiley and Sons: Hobboken, NJ, USA, 2014; pp. 108–125.

24. Zhang, M.; Lin, H.; Zeng, S.; Li, J.; Shi, J.; Wang, G. Impacts of plot location errors on accuracy of mapping and up-scaling aboveground forest carbon using sample plot and Landsat TM data. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1483–1487.

25. Foody, G.M.; Cutler, M.E.; McMorrow, J.; Pelz, D.; Tangki, H.; Boyd, D.S.; Douglas, I. Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Glob. Ecol. Biogeogr.* **2001**, *10*, 379–387.

26. Foody, G.M.; Boyd, D.S.; Cutler, M.E. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sens. Environ.* **2003**, *85*, 463–474.

27. Villa, P; Malucelli, F; Scalenghe, R. Carbon stocks in peri-urban areas: A case study of remote sensing capabilities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4119–4128.

28. Foody, G.M.; Green, RM.; Lucas, R.M.; Curran, P.J.; Honzak, M.; Do Amaral, I. Observations on the relationship between SIR-C radar backscatter and biomass of regenerating tropical forests. *Int. J. Remote Sens.* **1997**, *18*, 687–694.

29. Harrell, P.A.; Kasischke, E.S.; Bourgeau-Chavez, L.L.; Haney, E.M.; Christensen, N.L., Jr. Evaluation of approaches to estimating aboveground biomass in southern pine forests using SIR-C data. *Remote Sens. Environ.* **1997**, *59*, 223–233.

30. Kasischke, E.S.; Goetz, S.; Hansen, M.C.; Ustin, S.L.; Ozdogan, M.; Woodcock, C.E.; Rogan, J. Temperate and boreal forests. In *Remote Sensing for Natural Resource Management and Environmental Monitoring*; Ustin, S.L. Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2004; pp. 147–238.

31. Kuplich, T.M.; Salvatori, V.; Curran, P.J. JERS-1/SAR backscatter and its relationship with biomass of regenerating forests. *Int. J. Remote Sens.* **2000**, *21*, 2513–2518.

32. Askne, J.I.H.; Fransson, J.E.S.; Santoro, M.; Soja, M.J.; Ulander, L.M.H. Model-Based biomass estimation of a hemi-boreal forest from multitemporal TanDEM-X acquisitions. *Remote Sens.* **2013**, *5*, 5574–5597.

33. Lefsky, M.A.; Harding, D.J.; Cohen, W.B.; Parker, G.G.; Shugart, H.H. Surface LiDAR remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sens. Environ.* **1999**, *67*, 83–98.

34. Lefsky, M.A.; Cohen, W.B.; Harding, D.J.; Parker, G.G.; Acker, S.A.; Gower, S.T. Lidar remote sensing of above-ground biomass in three biomes. *Glob. Ecol. Biogeogr.* **2002**, *11*, 393–399.

35. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. Lidar remote system for ecosystem studies. *Am. Inst. Biol. Sci.* **2002**, *52*, 19–30.

36. Asner, G.P. Tropical forest carbon assessment: Integrating satellite and airborne mapping approaches. *Environ. Res. Lett.* **2009**, *4*, 1–11.

37. Asner, G.P.; Hughes, R.F.; Varga, T.A.; Knapp, D.E.; Kennedy-Bowdoin, T. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems* **2009**, *12*, 261–278.

38. Asner, G.P.; Powell, G.V.N.; Mascaro, J.; Knapp, D.E.; Clark, J.K.; Jacobson, J.; Kennedy-Bowdoin, T.; Balaji, A.; Paez-Acosta, G.; Victoria, E.; *et al.* High-Resolution forest carbon stocks and emissions in the Amazon. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16738–16742.

39. Asner, G.P.; Hughes, R.F.; Mascaro, J.; Uowolo, A.; Knapp, D.E; Jacobson, J.; Kennedy-Bowdoin, T.; Clark, J.K.; Balaji, A. High-resolution carbon mapping on the million-hectare Island of Hawaii. *Front. Ecol. Environ.* **2011**, doi:10.1890/100179.

40. Chen, Q. Lidar remote sensing of vegetation biomass. In *Remote Sensing of Natural Resources*; Wang, G., Weng, Q., Eds.; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2013; pp. 399–420.

41. Holopainen, M.; Vastaranta, M.; Liang, X.; Hyyppä, J.; Jaakkola, A.; Kankare, V. Estimation of forest stock and yield using LiDAR data. In *Remote Sensing of Natural Resources*; Wang, G., Weng, Q., Eds.; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2013; pp. 260–290.

42. Mascaro, J.; Asner, G.P.; Muller-Landau, H.C.; van Breugel, M.; Hall, J.; Dahlin, K. Control over aboveground forest carbon density on Barro Colorado Island, Panama. *Biogeosciences* **2011**, *8*, 1615–1629.

43. Mascaro, J.; Detto, M.; Asner, G.P.; Muller-Landau, H.C. Evaluating uncertainty in mapping forest carbon with airborne Lidar. *Remote Sens. Environ.* **2011**, *115*, 3770–3774.

44. Heiskanen, J. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *Int. J. Remote Sens.* **2006**, *27*, 1135–1158.

45. Fleming, A.; Wang, G.; McRoberts, R. Comparison of methods toward multi-scale forest carbon mapping and spatial uncertainty analysis: Combining national forest inventory plot data and Landsat TM images. *Eur. J. For. Res.* **2015**, *134*, 125–137.

46. Wang, G.; Wente, S.; Gertner, G.Z.; Anderson, A.B. Improvement in mapping vegetation cover factor for universal soil loss equation by geo-statistical methods with Landsat TM images. *Int. J. Remote Sens.* **2002**, *23*, 3649–3667.

47. Muukkonen, P.; Heiskanen, J. Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sens. Environ.* **2005**, *99*, 434–447.

48. Stuemer, W.; Kenter, B.; Koehl, M. Spatial interpolation of *in situ* data by self-organizing map algorithms (neural networks) for the assessment of carbon stocks in European forests. *For. Ecol. Manage.* **2010**, *260*, 287–293.

49. Tomppo, E. Multi-source National Forest Inventory of Finland. In Proceedings of the Subject Group S4.02–00 "Forest resource inventory and monitoring" and Subject Group S4.12–00 "Remote Sensing Technology", European Forest Institute, Joensuu, Finland, 6–12 August 1995; pp. 27–41.

50. McRoberts, R.; Tomppo, E.O.; Finley, A.O.; Heikkinen, J. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. *Remote Sens. Environ.* **2007**, *111*, 466–480.

51. McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbors techniques. *For. Ecol. Manage.* **2012**, *272*, 3–12.

52. McRoberts, R.E.; Næsset, E.; Gobakken, T. Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. *Remote Sens. Environ.* **2015**, *163*, 13–22.

53. Tomppo, E.; Olsson, H.; Ståhl, G.; Nilsson, M.; Hagner, O.; Katila, M. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens. Environ.* **2008**, *112*, 1982–1999.

54. Tomppo, E.; Halme, M. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: A genetic algorithm approach. *Remote Sens. Environ.* **2004**, *92*, 1–20.

55. Guan, D.; Chen, Y. Status of urban vegetation in Guangzhou City. *J. For. Res. (Harbin)* **2003**, *14*, 249–252. (In Chinese).

56. Escobedo, F.; Varela, S.; Zhao, M.; Wagner, J.E.; Zipperer, W. Analyzing the efficacy of subtropical urban forests in offsetting carbon emissions from cities. *Environ. Sci. Policy* **2010**, *13*, 362–372.

57. Golubiewski, N.E. Urbanization increases grassland carbon pools: Effects of landscaping in Colorado's Front Range. *Ecol. Appl.* **2006**, *16*, 555–571.

58. Liu, C.; Li, X. Carbon storage and sequestration by urban forests in Shenyang, China. *Urban For. Urban Green.* **2012**, *11*, 121–128.

59. McPherson, E.G. Atmospheric carbon dioxide reduction by Sacramento's Urbana forests. *J. Arboric.* **1998**, *24*, 215–223.

60. Nowak, D.J. Atmospheric carbon reduction by urban trees. *J. Environ. Manage.* **1993**, *37*, 207–217.

61. Nowak, D.J.; Crane, D.E. Carbon storage and sequestration by urban trees in the USA. *Environ. Pollut.* **2002**, *116*, 381–389.

62. Nowak, D.J.; Grane, D.E.; Dwyer, J.F. Compensatory value of urban trees in the United States. *J. Arboric.* **2002**, *28*, 194–199.

63. Schreyer, J.; Tigges, J.; Lakes, T.; Churkina, G. Using airborne LiDAR and QuickBird data for modelling urban tree carbon storage and its distribution—A case study of Berlin. *Remote Sens.* **2014**, *6*, 10636–10655.

64. Kordowski, K.; Kuttler, W. Carbon dioxide fluxes over an urban park area. *Atmos. Environ.* **2010**, *44*, 2722–2730.

65. Zhang, C.; Tian, H.; Chen, G.; Chappelka, A.; Xu, X.; Ren, W.; Hui, D.; Liu, M.; Lu, C.; Pan, S.; *et al.* Impacts of urbanization on carbon balance in terrestrial ecosystems of the southern United States. *Environ. Pollut.* **2012**, *164*, 89–101.

66. Strohbach, M.; Haase, D. The above-ground carbon stock of a central European city: Patterns of carbon storage in trees in Leipzig, Germany. *Landsc. Urban Plan.* **2012**, *104*, 95–104.

67. Nowak, D.J.; Greenfield, E.J.; Hoehn, R.E.; Lapoint, E. Carbon storage and sequestration by trees in urban and community areas of the United States. *Environ. Pollut.* **2013**, *178*, 229–236.

68. Myeong, S.; Nowak, D.J.; Duggin, M.J. A temporal analysis of urban forest carbon storage using remote sensing. *Remote Sens. Environ.* **2006**, *101*, 277–282.

69. McPherson, E.G.; Xiao, Q.; Aguaron, E. A new approach to quantify and map carbon stored, sequestered and emissions avoided by urban forests. *Landsc. Urban Plan.* **2013**, *120*, 70–84.

70. Zheng, D.; Duceya, M.J.; Heath, L.S. Assessing net carbon sequestration on urban and community forests of northern New England, USA. *Urban For. Urban Green.* **2013**, *12*, 61–68.

71. Hutyraa, L.R.; Yoon, B.; Hepinstall-Cymermanc, J.; Alberti, M. Carbon consequences of land cover change and expansion of urban lands: A case study in the Seattle metropolitan region. *Landsc. Urban Plan.* **2011**, *103*, 83–93.

72. Shrestha, R.; Wynne, R.H. Estimating biophysical parameters of individual trees in an urban environment using small footprint discrete-return imaging LiDAR. *Remote Sens.* **2012**, *4*, 484–508.

73. Moskal, L.M.; Zheng, G. Retrieving forest inventory variables with Terrestrial Laser Scanning (TLS) in urban heterogeneous forest. *Remote Sens.* **2012**, *4*, 1–20.

74. Davies, Z.G.; Dallimer, M.; Edmondson, J.L.; Leake, J.R.; Gaston, K.J. Identifying potential sources of variability between vegetation carbon storage estimates for urban areas. *Environ. Poll.* **2013**, *183*, 133–142.

75. Pataki, P.E.; Alig, R.J.; Fung, A.S.; Golubiewski, N.E.; Kennedy, C.A.; McPherson, E.G.; Nowak, D.J.; Pouyat, R.V.; Romerolankao, P. Urban ecosystems and the North American carbon cycle. *Glob. Change Biol.* **2006**, *12*, 1–11.

76. He, C.; Convertino, M.; Feng, Z.; Zhang, S. Using LiDAR data to measure the 3D green biomass of Beijing urban forest in China. *PLoS ONE* **2013**, *8*, e75920.

77. Wang, Z.; Cui, X.; Yin, S.; Shen, G; Han, Y.; Liu, C. Characteristics of carbon storage in Shanghai's urban forest . *Chin. Sci. Bull.* **2013**, *58*, 1130–1138. (In Chinese).

78. Li, X.; Kang, W. Function of carbon sink of forest ecosystem in Guangzhou. *J. Central South Univ. For. Tech.* **2008**, *28*, 8–13. (In Chinese).

79. Wang, Z.H.; Liu, H.M.; Guan, Q.W.; Wang, X.; Hao, J.; Ling, N.; Shi, C. Carbon storage and density of urban forest ecosystems in Nanjing. *J. Nanjing For. Univ.: Natl. Sci. Ed.* **2011**, *35*, 18–22. (In Chinese).

80. Zhou, G. Measures to increase carbon sink in Guangzhou based on carbon storage dynamics in recent years. *J. Chin. Urban For.* **2007**, *5*, 24–27. (In Chinese).

81. Zhou, J.; Xiao, R.; Zhuan, C.; Deng, Y. A review of city forest carbon sequestration methods. *Chin. J. Ecol.* **2013**, *32*, 3368–3378. (In Chinese).

82. Ying, T.; Li, M.; Fang, W. Estimation of Harbin forest carbon stock. *J. Northeast For. Uni.* **2009**, *37*, 33–35. (In Chinese).

83. Dennison, P.E.; Roberts, D.A. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. *Remote Sens. Environ.* **2003**, *87*, 123–135.

84. Lennington, R.K.; Sorensen, C.T.; Heydorn, R.P. A mixture model approach for estimating crop areas from Landsat data. *Remote Sens. Environ.* **1984**, *14*, 197–206.

85. Adams, J.B.; Sabol, D.E.; Kapos, V.; Almerida Filho, R.; Roberts, D.A.; Smith, M.O.; Gillespie, A.R. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. *Remote Sens. Environ.* **1995**, *52*, 137–154.

86. Roberts, D.A.; Gardner, M.; Church, R.; Ustin, S; Scheer, G.; Green, R.O. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sens. Environ.* **1998**, *65*, 267–279.

87. Small, C. A global analysis of urban reflectance. *Int. J. Remote Sens.* **2005**, *26*, 661–681.

88. Lu, D.; Morana, E.; Batistell, M. Linear mixture model applied to Amazonian vegetation classification. *Remote Sens. Environ.* **2003**, *87*, 456–469.

89. Chen, L.; Zhang, X.; Jiao, Z. Reversion of leaf area index in forest based on linear mixture model. *Trans. Chin. Soc. Agri. Eng.* **2013**, *29*, 124–129.

90. Hu, X.; Weng, Q. Extraction of impervious surfaces from hyperspectral imagery: Linear *vs.* nonlinear methods. In *Remote Sensing of Natural Resources*; Wang, G., Weng, Q., Eds.; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2013; pp. 141–154.

91. Chinese Ministry of Forestry-Department of Forest Resource and Management (CMF-DFRM). *Forest Resources of China 1949 to 93*; Chinese Forestry Press: Beijing, China, 1996.

92. National Center of Forestry Carbon Sequestration Estimation and Monitoring (NCFCSEM). *Guidelines for National Forest Carbon Sequestration Estimation and Monitoring Technology*; National Institute of Forest Inventory and Planning, State Forestry Administration of China: Beijing, China, 2010; p. 39. (In Chinese).

93. Howard, H.; Wang, G.; Singer, S.; Anderson, A.B. Modeling and Prediction of Land Condition for Fort Riley Military Installation. *Trans. ASABE* **2013**, *56*, 643–652.

94. 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.

95. Harrell, F.E. Prespecification of predictor complexity without later simplification. In *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*; Springer-Verlag: New York, NY, USA, 2001; pp. 64–65.

96. McDonald, J.H. *Handbook of Biological Statistics*; Sparky House Publishing: Baltimore, MD, USA, 2008.

97. Scalenghe, R; Malucelli, F; Ungaro, F; Perazzone, L; Filippi, N; Edwards, A.C. Influence of 150 years of land use on anthropogenic and natural carbon stocks in Emilia-Romagna Region (Italy). *Environ. Sci. Tech.* **2011**, *45*, 5112–5117.