

PAPER • OPEN ACCESS

Estimating aboveground biomass of urban trees by high resolution remote sensing image: a case study in Hengqin, Zhuhai, China

To cite this article: Linze Bai *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **569** 012053

View the [article online](#) for updates and enhancements.

You may also like

- [Improved quantification of cover crop biomass and ecosystem services through remote sensing-based model-data fusion](#)
Lexuan Ye, Kaiyu Guan, Ziqi Qin et al.
- [Spatial patterns of arctic tundra vegetation properties on different soils along the Eurasia Arctic Transect, and insights for a changing Arctic](#)
Howard E Epstein, Donald A Walker, Gerald V Frost et al.
- [Mixed-species allometric equations and estimation of aboveground biomass and carbon stocks in restoring degraded landscape in northern Ethiopia](#)
Mulugeta Mokria, Wolde Mekuria, Aster Gebrekirstos et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Estimating aboveground biomass of urban trees by high resolution remote sensing image: a case study in Hengqin, Zhuhai, China

Linze Bai^{1,3}, Yuxuan Shu² and Yongxuan Guo²

¹State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China;

²School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

³Email: 2015302590104@whu.edu.cn

Abstract. Aboveground biomass of trees is the weight of organic matter per unit area of trees. The monitoring and evaluation of aboveground biomass of urban trees is helpful to the construction of garden city. With the development of urbanization and the Continuous urban ecological research, the aboveground biomass of urban trees is getting more important for the study of urban carbon cycle and the construction of ecological city. High resolution remote sensing image, which overcomes the time-consuming and labor-consuming drawbacks of traditional methods, is a data source for estimating the aboveground biomass. In this paper, taking Hengqin New Area, Zhuhai City, Guangdong Province, China as an example, we use the method of object-oriented classification of Worldview-03 data to extract Hengqin tree area, and analyze 15 feature parameters in remote sensing images including vegetation indexes and texture features. Combined with field measurement and aboveground biomass computation, we set up Geographical Weighted Regression (GWR) model and Random Forest (RF) model for regression. Finally we estimate and discuss the aboveground biomass of trees in Hengqin.

1. Introduction

The need for new approaches to analysis the urban ecosystem has become increasingly important and pressing as urbanization leads to growing coverage of impervious surface and the proportion of carbon emissions is also on the rise [1,2]. Urban trees are hailed as the lung of the city for they not only play an important role in beautifying the environment, but also are valuable urban carbon storage [3]. Aboveground biomass of urban trees refers to the total amount of organic matter per unit area of trees in urban area at a certain time. It is an important indicator to measure the carbon sequestration level of urban trees, a measure of urban livability, and may also shed light on the law of carbon cycle [4]. In addition, biomass data can be used to alleviate the aggravation of heat island effect and control urban pollution such as soil, water and air pollution. This research is also of great value to urban planning, garden city establishing, environmental protection, fine management and other aspects of social management [5].

The traditional method of calculating aboveground biomass is obtained by a sampling survey in the field directly or through a given model [6]. However, it costs much labor, material resources, money and time, which makes it difficult to apply to large-scale biomass survey in cities. Due to the wide coverage of remote sensing data and its high computational efficiency, the application of remote



sensing methods in biomass inversion has been developed vigorously [7,8]. Based on prior knowledge, the relationship between remote sensing spectral band information and vegetation biophysical parameters is established, and the inversion of tree biomass is realized by using appropriate modeling methods. High resolution satellite images become the data source of biomass assessment. This study estimated the aboveground biomass of trees in Hengqin New Area, Zhuhai, China using Worldview-03 satellite data.

2. Study area and methods

2.1. Study area

The study area is Hengqin New Area, Zhuhai, Guangdong Province, China. As shown in Figure 1, Hengqin New Area is composed of big and small Hengqin islands, from $113^{\circ} 26' 40''$ to $113^{\circ} 33' 59''$ E, $21^{\circ} 5' 10''$ to $22^{\circ} 10' 26''$ N, covering an area of 106.46 square kilometers, and connected with a bridge to Macao in the East.

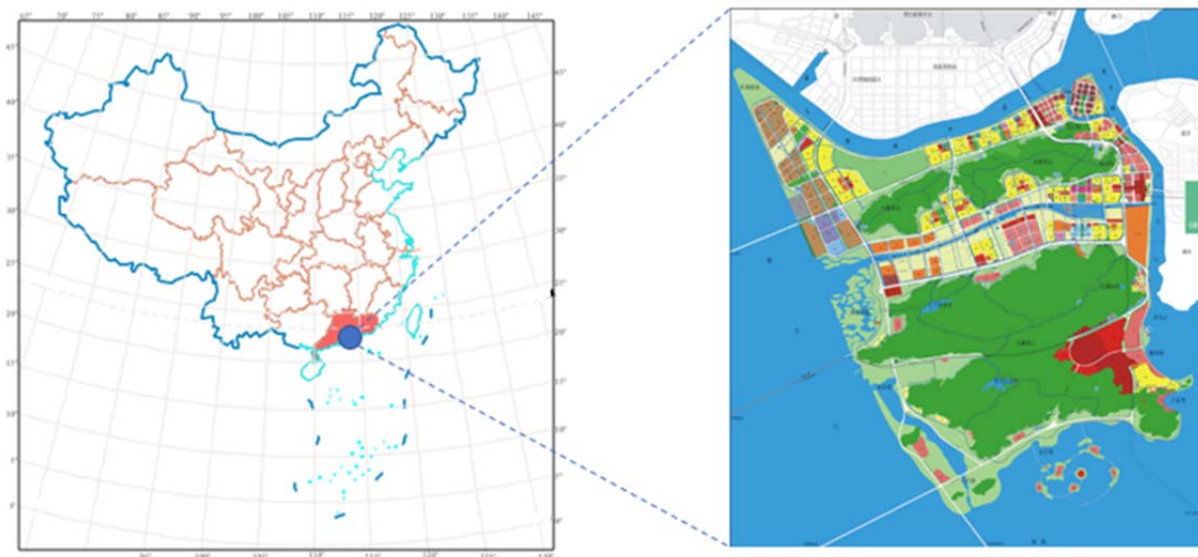


Figure 1. Hengqin area map.

Hengqin is located in the subtropical monsoon climate zone, with an annual average temperature of 22°C . It has a good biodiversity with 896 plant species. According to the statistics of forest resources in Hengqin New Area in 2009, the total area of forest land was 3184.7 hectares, accounting for 28.28% of the total land area, the forest coverage rate of the whole area was 26.7%, and the utilization rate of forest land reached 95.58%. The forest structure of Hengqin Mountain is complex, with tropical evergreen broad-leaved trees dominating the area. The main tree species of Hengqin are aubergine, Xushu, Haiqi, etc. Mangzhou Wetland Park and Binhai Wetland Park are planted with a large area of mangrove. In the built-up area, the original woods were destroyed, and now they are mostly planted forests.

2.2. Field measurement and aboveground biomass computation

Field data was collected in Hengqin from July to August 2018. The data collection steps are as follows: First, the distribution position of the sample plot is determined by subjective estimation method to make it distributed in the whole study area. Second, the H (height) and DBH (the diameter in breast high of trees) of a single tree are measured on site in various sites. Finally, the center coordinates of various sites are measured by GPS (global positioning system), and the tree species in the sample plot are recorded and photographed.

All tree species and land cover types were considered when selecting the sample trees. In the end, we collected 621 sample points. Based on the height of trees, 210 sampling points with precise GPS positioning were selected for subsequent experiments.

The field data are transformed into aboveground biomass through the allometric growth model. This study refers to the migration rate growth equation [9] proposed by he Q s et al. (2012). The specific equation is shown in equation 1 to equation 4.

$$AGB_{branch} = 0.0061 \times ((DBH)^2 \times H)^{0.8905} \quad (1)$$

$$AGB_{leaf} = 0.2650 \times ((DBH)^2 \times H)^{0.4701} \quad (2)$$

$$AGB_{fruit} = 0.0342 \times ((DBH)^2 \times H)^{0.5779} \quad (3)$$

$$AGB = AGB_{branch} + AGB_{leaf} + AGB_{fruit} \quad (4)$$

Where AGB represents aboveground biomass and H represents tree heights.

2.3. Remote sensing data processing

The Worldview-03 satellite data used in this study is consistent with the field measurement time. Worldview-03 was launched in August 2014, with 0.31m resolution panchromatic band image, 8-band multispectral image and 8-band short wave infrared image. Worldview-03 data is used in quantitative analysis and vegetation monitoring of urban areas with strong potential. Pan (Panchromatic band), B (Blue), G (Green), R (Red) and NIR1 (Near Infrared 1) bands are used in this study. The band parameters are shown in Table 1.

Table 1. Band parameters.

Band name	Band range (nm)	Resolution (m)
Pan	450-800	0.31
B	450-510	1.2
G	510-580	1.2
R	630-690	1.2
NIR1	770-895	1.2

We use ENVI5.3 software to preprocess the image. Preprocessing includes ortho correction, image fusion, radiometric calibration, atmospheric correction and image clipping.

Based on remote sensing image, vegetation indexes and texture features are extracted. Vegetation index is characterized by dominant vegetation index under different scenarios, and altogether 7 vegetation indexes are extracted [10]. NIR(Near Infrared) single band image is selected for biomass inversion based on texture features. The calculation window size is set to 3×3 , and the sliding step size is set to 1 to get the best effect[11,12]. The specific equation description is shown in Table 2.

2.4. Tree area extraction

In high-resolution images of urban areas, buildings have complex surfaces, strong spatial heterogeneity, and shadows. We use eCognition software to generate homogeneous objects through multi-scale segmentation. Considering the shadow scene, we established the key features of segmented objects and extracted the tree area. The extraction process is as follows.

2.4.1. Multiscale segmentation. The purpose of segmentation is to obtain homogeneous regions, and to separate and represent ground object at the optimal scale. After experiment and visual interpretation, the optimal segmentation scale is set to 80, the spectral factor is set to 0.5, and the shape factor is set to 0.5.

Table 2. Optical feature parameters.

Variable type	Variable name	Computing method
Vegetation index	NDVI	$\frac{NIR - R}{NIR + R}$
	NDGI	$\frac{NIR - G}{NIR + G}$
	OSAVI	$(1 + 0.16) \frac{NIR - R}{NIR + R + 0.16}$
	MSAVI	$NIR + 0.5 - \sqrt{(NIR + 0.5)^2 - 2 \times (NIR - R)}$
	ARVI	$\frac{NIR - (2R - B)}{NIR + (2R - B)}$
	EVI	$\frac{2.5(NIR - R)}{1 + NIR + 6R - 7.5B}$
	Clgreen	$\frac{NIR}{G} - 1$
Texture feature	Mean	$\sum_{i,j=0}^{N-1} iP_{i,j}$
	Homogeneity	$\sum_{i,j=0}^{N-1} i \frac{P_{i,j}}{1 + (i - j)^2}$
	Contrast	$\sum_{i,j=0}^{N-1} iP_{i,j}(i - j)^2$
	Dissimilarity	$\sum_{i,j=0}^{N-1} iP_{i,j} i - j $
	Entropy	$\sum_{i,j=0}^{N-1} iP_{i,j}(-\ln P_{i,j})$
	Variance	$\sum_{i,j=0}^{N-1} P_{i,j} \left(1 - \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} P_{i,j}\right)$
	Second Moment	$\sum_{i,j=0}^{N-1} iP_{i,j}^2$
	Correlation	$\sum_{i,j=0}^{N-1} i \frac{\sum_{i,j=0}^{N-1} ijP_{i,j} - \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} P_{i,j} \sum_{j=0}^{N-1} j \sum_{i=0}^{N-1} P_{i,j}}{\sigma_i^2 \sigma_j^2}$

Notes: NDVI refers to Normalized Difference Vegetation Index. NDGI refers to Normalized Difference Green Index. OSAVI refers to Optimize Soil Adjusted Vegetation Index. MSAVI refers to Modified Soil Adjusted Vegetation Index. ARVI refers to Atmospherically Resistant Vegetation Index. EVI refers to Enhanced Vegetation Index.

2.4.2. Vegetation information extraction. NDVI and NDGI were used to extract vegetation information. NDVI is the most commonly used best factor to reveal the growth state and spatial

density of plants at present. However the threshold classification of NDVI index easily Misclassify non-vegetated areas as vegetation areas in urban shadow area. NDGI can weaken the non-vegetation features in the shadow area, but the bright roofs in the non-shaded area are easily misclassified as vegetation. In the experiment, the multi parameter threshold method was used to get the vegetation coverage area. The selection of threshold is shown in equation 5 and equation 6.

$$NDVI = (NIR - R) / (NIR + R) \geq 0.1 \quad (5)$$

$$NDGI = (NIR - G) / (NIR + G) \geq 0.1 \quad (6)$$

Where, NIR, R and G represent the pixel values of near infrared band, red band and green band respectively.

2.4.3. Classification of trees and grasslands in non-shadow areas. Objects whose brightness are greater than 10% of the maximum brightness value of the image are referred to as non-shadow areas. The classification of trees and grassland in non-shadow area is realized by Intensity and GLDVA (Grey Difference Level Vector ASM). Compared with trees, grassland has a higher intensity and GLDVA values. By setting a threshold value to 0.02 for the product of Intensity and GLDVA ($\text{Intensity} \times \text{GLDVA}$), we can distinguish the information of trees and grasslands in the non-shadow area.

2.4.4. Classification of trees and grasslands in shadow areas. Objects whose brightness feature is within 10% of the maximum brightness value of the image are shadow areas. The classification of trees and grassland in shadow area is realized by NDVI and brightness features. In the shadow area, the grass color is darker, while the NDVI value of trees in the shadow area is higher. By setting the threshold value of the product of $\text{brightness} \times \text{NDVI}$ to 13.5, we can distinguish the information of trees and grasslands in the shadow area.

Figure 2 is a schematic diagram of the classification results of a building intensive area in Hengqin. On the left, it can be seen that the surface type of this area is complex and there are building shadows, including trees and grasslands. The right image is the result of classification. Trees and grassland have been extracted respectively, and the misclassification of shadow area is effectively reduced.

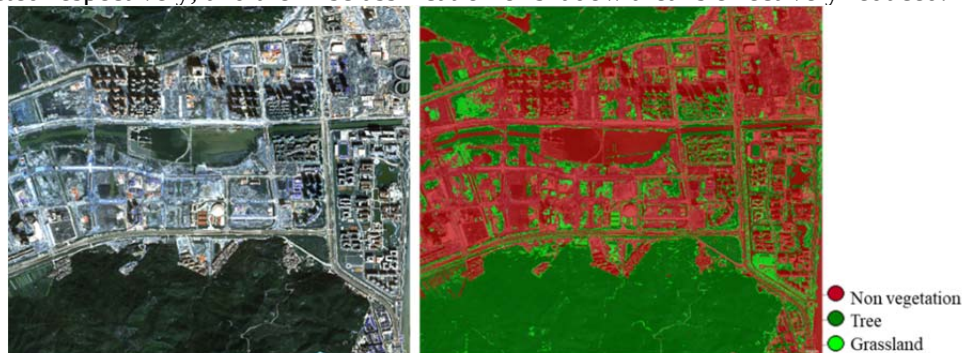


Figure 2. Sketch map of tree extraction results.

The extracted tree area in the whole Hengqin was used for subsequent experiments.

2.5. Regression algorithms

In order to establish the regression relationship between the 15 feature parameters and the aboveground biomass of trees, GWR (geographical weighted regression) and RF (random forest model) are used for comparison.

GWR: Spatial non-stationarity is the change of the relationship between variables or structure caused by the change of geographical location, which is a manifestation of spatial heterogeneity [13]. Standard regression analysis assigns the same weight to each sample without considering the spatial

non-stationarity, so it is difficult to get a reasonable regression result. GWR model is an extension of general linear regression model. GWR embeds the spatial position of the data into the regression equation. GWR considers the local effect of the spatial object by establishing the local regression equation at each point in the spatial range.

$$y_i = \alpha(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \dots + \beta_p(u_i, v_i)x_{ip} + \varepsilon_i \quad (7)$$

Where (u_i, v_i) refers to the coordination of location, $\{\hat{\alpha}(u_i, v_i), \hat{\beta}_1(u_i, v_i), \dots, \hat{\beta}_p(u_i, v_i)\}$ refers to the model parameters in the location.

GWR assigns a weight to each sample, the weight is given by the spatial weight matrix, and the coefficient is calculated according to the least squares rule. W is the weight matrix.

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) \bar{y} \quad (8)$$

$$W(u_i, v_i) = \begin{pmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{pmatrix} \quad (9)$$

By selecting different weight functions, different spatial weight matrices can be obtained. The spatial weight function selected in this paper is Gauss function, where d_{ij} is the distance between i and j , b is the bandwidth parameter.

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{2b^2}\right) \quad (10)$$

RF: Random forest is a Bagging method that uses CART (Classification and Regression Trees) decision tree as a weak learner. When each tree is generated, the selected features are a small number of randomly selected features, ensuring the randomness of features. Compared with the general Bagging algorithm, the sample number N collected by RF is the same as that of the training set.

RF model is characterized by fast prediction speed and strong anti-noise capability. Due to its randomness, RF generally does not need additional pruning to obtain better generalization ability and anti-overfitting ability. However, RF model only has the best performance in large sample data, and the regression prediction effect on small sample data will decrease to a certain extent.

3. Results and discussion

3.1. Regression results

The scatter diagram of regression results based on 210 selected field measurement points is shown in Figure 3 and Figure 4. Figure 3 shows the regression result of GWR, R^2 is 0.603, RMSE is 35.44 Mg/ha. Figure 4 is the regression result of RF, R^2 is 0.774, RMSE is 26.72 Mg/ha.

The figure above shows that the estimation accuracy of RF model is higher than that of GWR. As a machine learning model, RF model can better extract information from feature parameters. Therefore, we choose RF model to estimate the aboveground biomass of trees in Hengqin.

3.2. Estimating aboveground biomass of trees

We extract 15 feature parameters according to the method described in 2.3 and the tree area according to the method described in 2.4. We use the result of RF model to estimate the aboveground biomass of trees in Hengqin. Finally, the thematic map of aboveground biomass of trees in Hengqin is shown in Figure 5.

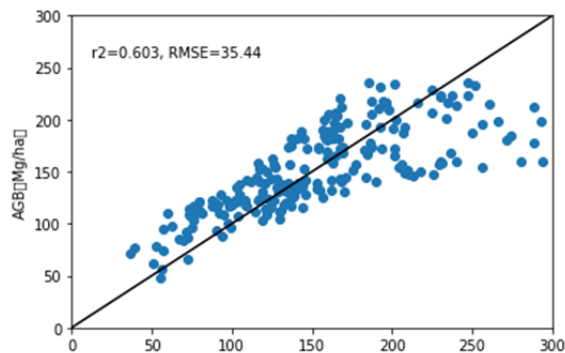


Figure 3. The scatter diagram of GWR regression result.

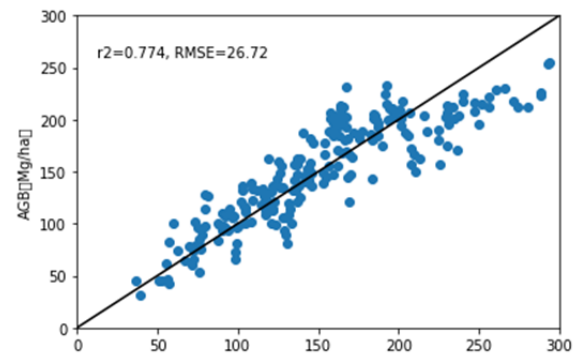


Figure 4. The scatter diagram of RF regression result.

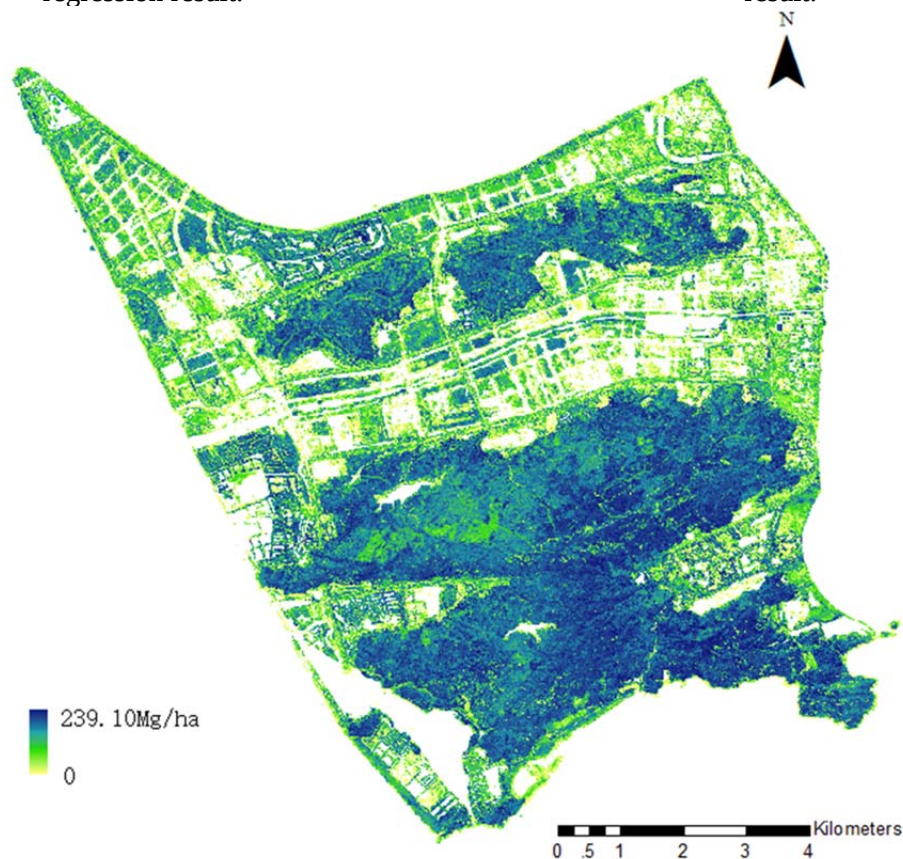


Figure 5. Thematic map of aboveground biomass of trees in Hengqin.

The maximum aboveground biomass of Hengqin is 239.10 Mg/ha, and the average is 88.12 Mg/ha.

3.3. Discussion

The largest contribution of Hengqin's surface biomass is the trees of Hengqin Mountain. The aboveground biomass values near Poji Mountain and Nuobei Mountain are generally high. Aboveground biomass is distributed between 120Mg / ha and 239Mg / ha, which is the densest vegetation. The overall distribution of surface biomass in Santang Mountains ranges from 60 Mg / ha to 166 Mg / ha. The aboveground biomass of Santang Mountain is lower than the surrounding area, indicating that the virgin forest in this area has been destroyed by human activities. The aboveground

biomass of Mangzhou wetland and coastal wetland is generally distributed between 45 Mg / ha and 115 Mg / ha. The mangrove planted in large area has a positive impact on the ecology of Hengqin. Huandao Road and coastline has made significant contribution, which not only provides leisure space for citizens and tourists, but also plays an important role in urban carbon sink. In the central business district and port service area in the northeast of Hengqin, the aboveground biomass value is relatively low, and the aboveground biomass value is generally distributed between 20 Mg / ha and 40 Mg / ha, which is caused by the damage of artificial buildings to vegetation in the built-up area.

For GWR Model and RF model, the prediction accuracy of RF model is higher than that of GWR Model, and the mining ability of feature parameters is stronger. However, the aboveground biomass values of the two models are underestimated when they are greater than 200 Mg / ha due to the saturation of optical data with dense trees. In order to solve the saturation phenomenon of optical data, Lidar data can be introduced in the future [14].

4. Conclusion

In this paper, we use Worldview-03 data to extract tree areas. Based on 210 field measurement data and 15 feature parameters extracted from remote sensing images, we established GWR Model and RF model to estimate the aboveground biomass of trees. The results show that the accuracy of RF model is higher, R^2 is 0.774. Then we use RF model to estimate the aboveground biomass of trees in Hengqin. The maximum aboveground biomass of Hengqin is 239.10 Mg/ha, and the average is 88.12 Mg/ha. In general, the random forest model with the vegetation index and texture features extracted from worldview-03 satellite as input has obtained satisfactory estimation accuracy of urban tree biomass. The proposed workflow will greatly promote the future urban tree growth monitoring and urban forest resources assessment.

Acknowledgments

This study was supported by Science and technology innovation plan for college students of Chinese Academy of Sciences(20184000026).

Reference

- [1] Shao Z F, Fu H Y, Li D L, Orhan A, Cheng T 2019 Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation *Remote Sensing of Environment* **232** 111338
- [2] Godwin C , Chen G , Singh K K 2015 The impact of urban residential development patterns on forest carbon density: An integration of LiDAR, aerial photography and field mensuration *Landscape and Urban Planning* **136** 97-109
- [3] Nowak D J, Crane D E 2002 Carbon storage and sequestration by urban trees in the USA *Environmental Pollution* **116(3)** 381-389
- [4] O. W. Tsui, N. C. Coops, M. A. Wulder, and P. L. Marshall 2013 Integrating airborne LiDAR and space-borne radar via multivariate Kriging to estimate above-ground biomass *Remote Sens. Environ.* **139** 340-352
- [5] ZHENG D, RADEMACHER J, CHEN J 2004 Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA *Remote Sensing of Environment* **93(3)** 402-411
- [6] L. Korhonen, J. Heiskanen, and I. Korpela 2013 Modelling Lidar-derived boreal forest canopy cover with SPOT 4 HRVIR data *Int. J. Remote Sens* **34** 8172-8181
- [7] HEALEY S P, YANG Z, COHEN W B 2006 Application of two regression-based methods to estimate the effects of partial harvest on forest structure using Landsat data *Remote Sensing of Environment* **101(1)** 115-126
- [8] COHEN W B, SPIES T A, ALIG R J 2002 Characterizing 23 Years (1972-95) of Stand Replacement Disturbance in Western Oregon Forests with Landsat Imagery *Ecosystems* **5(2)** 122-137

- [9] He Q S, Cao C X, Chen E X 2012 Forest stand biomass estimation using ALOS PALSAR data based on LiDAR-derived prior knowledge in the Qilian Mountain, western China *International Journal of Remote Sensin* **33(3)** 710-729
- [10] DUBE T, MUTANGA O 2015 Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas *Isprs Journal of Photogrammetry & Remote Sensing* **108** 12-32
- [11] Zhang L J, Cheng Q M, Li C M 2015 Improved model for estimating the biomass of *Populus euphratica* forest using the integration of spectral and textural features from the Chinese high-resolution remote sensing satellite GaoFen-1 *JOURNAL OF APPLIED REMOTE SENSING* **9** 096010
- [12] KELSEY K, NEFF J **2014** Estimates of Aboveground Biomass from Texture Analysis of Landsat Imagery *Remote Sensing* **6(7)** 6407-6422
- [13] Lu B, Yang W, Ge Y 2018 Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths *Computers, Environment and Urban Systems* S0198971517303447
- [14] Zhang L J, Shao Z F, Liu J C, Cheng Q M 2019 Deep Learning Based Retrieval of Forest Aboveground Biomass from Combined LiDAR and Landsat 8 Data *Remote Sensing* **11(12)** 1459