

DOI: 10.5846/stxb202102260545

龙依,蒋馥根,孙华,王天宏,邹琪,陈川石.基于带宽优选地理加权回归模型的深圳市植被碳储量反演.生态学报,2022,42(12):4933-4945.

Long Y, Jiang F G, Sun H, Wang T H, Zou Q, Chen C S. Estimating vegetation carbon storage based on optimal bandwidth selected from geographically weighted regression model in Shenzhen City. Acta Ecologica Sinica, 2022, 42(12): 4933-4945.

基于带宽优选地理加权回归模型的深圳市植被碳储量反演

龙 依^{1,2,3}, 蒋馥根^{1,2,3}, 孙 华^{1,2,3,*}, 王天宏^{1,2,3}, 邹 琪⁴, 陈川石⁴

1 中南林业科技大学 林业遥感信息工程研究中心,长沙 410004

2 林业遥感大数据与生态安全湖南省重点实验室,长沙 410004

3 南方森林资源经营与监测国家林业与草原局重点实验室,长沙 410004

4 国家林业和草原局中南调查规划设计院,长沙 410004

摘要:植被碳储量估测是自然资源监测的重要内容,遥感技术结合地面样地进行反演可以获得区域范围内植被碳储量的空间连续分布,弥补了传统人工抽样调查估测的不足。然而,现有的参数和非参数遥感估测模型大多忽略了样地数据的变异与空间自相关关系。研究以 Landsat 8 OLI 影像为数据源提取遥感变量,结合植被碳储量实测调查数据,利用最小信息准则(AICc)、最大空间自相关距离(MSAD)和交叉验证(CV)分别确定最优带宽,组合 Gaussian、Bi-square 和 Exponential 核函数构建地理加权回归(GWR)模型估算深圳市植被碳储量,并与多元线性回归(MLR)进行比较,选择最优模型绘制深圳市植被碳储量空间分布图。研究结果表明,GWR 模型整体精度优于 MLR 模型,GWR 模型的决定系数(R^2)均高于 MLR 模型,且均方根误差(RMSE)和平均绝对误差(MAE)显著降低。带宽和核函数的选择对 GWR 模型估测结果产生了显著影响。以 CV 确定带宽、Exponential 为核函数组合构建的 GWR 模型效果最佳,其 R^2 为 0.697, RMSE 为 10.437 Mg C/hm²,相比其它模型精度上升了 13.87%—32.28%,且变量回归参数均存在显著空间非平稳性,能较好反映空间异质性。结合最优 GWR 模型和植被类型空间分布估算深圳市植被碳储量,得到其碳储量值在 1.63—60.95 Mg C/hm² 之间,其中高值和低值主要分布于森林和草地区域,与深圳市植被覆盖情况基本一致。GWR 模型考虑了变量空间异质性,基于带宽优选的 GWR 模型一定程度上能获得更合理的碳储量空间分布,能为深圳市植被碳储量遥感估算提供方法与技术参考。

关键词:遥感反演;植被碳储量;局部模型;地理加权回归;Landsat 8

Estimating vegetation carbon storage based on optimal bandwidth selected from geographically weighted regression model in Shenzhen City

LONG Yi^{1,2,3}, JIANG Fugen^{1,2,3}, SUN Hua^{1,2,3,*}, WANG Tianhong^{1,2,3}, ZOU Qi⁴, CHEN Chuanshi⁴

1 Research Center of Forestry Remote Sensing & Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China

2 Key Laboratory of Forestry Remote Sensing Based Big Data & Ecological Security for Hunan Province, Changsha 410004, China

3 Key Laboratory of National Forestry & Grassland Administration on Forest Resources Management and Monitoring in Southern Area, Changsha 410004, China

4 Central South Survey and Planning Design Institute of State Forestry and Grassland Administration, Changsha 410004, China

Abstract: Vegetation carbon storage estimation is an important part of natural resources monitoring. Remote sensing technology combined with ground sample can obtain the spatial continuous distribution of vegetation carbon reserves in the

基金项目:湖南省教育厅重点项目(17A225);国家自然科学基金面上项目(31971578);湖南省林业科技创新专项(XLK201986);长沙市科技计划项目(kq2004095)

收稿日期:2021-02-26; 网络出版日期:2022-02-11

* 通讯作者 Corresponding author. E-mail: sunhua@csuft.edu.cn

region, which makes up for the deficiency of the traditional artificial sampling survey estimation. However, most of the existing parameter and nonparameter remote sensing estimation models ignore the relationship between sample data variation and spatial autocorrelation. In this study, Landsat 8 OLI images were used as data sources to extract remote sensing variables, combined with the survey data of vegetation carbon storage in Shenzhen. Three metrics, including Akaike Information Criterion corrected (AICc), maximum spatial autocorrelation distance (MSAD), and Cross-Validation (CV), were used to determine the optimal bandwidth. Different geographically weighted regression (GWR) models inversing carbon reserve were constructed using Gaussian, Bi-square and Exponential as kernel function, respectively. Compared with the multiple linear regression (MLR), the optimal model was selected to make the spatial distribution map of vegetation carbon storage in Shenzhen. The results showed that the overall precision of the GWR models were better than that of the MLR model. The coefficient of determination (R^2) of the GWR models were higher than that of MLR model, and the root mean square error (RMSE) and mean absolute error (MAE) were reduced significantly. The selection of bandwidth and kernel function can have significant influence on the estimation results of GWR model. Among the constructed GWR models, the GWR model with cross-validation determined bandwidth and exponential as the kernel was the best, with R^2 of 0.697 and RMSE of 10.437 Mg C/hm², which increased by 13.87% to 32.28% on model accuracy compared with other models. And the variable regression parameters of this model had significantly spatial non-stationarity, which could better reflect spatial heterogeneity. Combined with the optimal GWR model and the spatial distribution of vegetation types, the carbon storage of vegetation in Shenzhen was estimated, and the carbon storage values were between 1.63 and 60.95 Mg C/hm². The high and low carbon storage were mainly distributed in forest and grassland areas, which was basically consistent with the vegetation coverage in Shenzhen. Considering the spatial heterogeneity of variables, the GWR model based on bandwidth optimization can obtain more reasonable spatial distribution of vegetation carbon storage to some extent, which can provide a method and technical reference for remote sensing estimation of vegetation carbon storage in Shenzhen.

Key Words: remote sensing inversion; vegetation carbon storage; local model; geographically weighted regression; Landsat 8

城市植被作为城市生态系统的重要组成部分,对维持城市碳氧平衡、改善城市小气候至关重要^[1]。随着碳达峰、碳中和等目标的提出,城市植被在碳汇功能中的作用得到广泛关注^[2—3]。植被碳储量作为评估植被碳收支的重要参数,能直接衡量植被固碳增汇能力,是生态系统服务功能的直接体现^[4]。迅速、准确地估算城市植被碳储量对评估植被碳汇价值、管理城市生态环境和可持续发展具有重要意义^[5]。常用的碳储量估测方式是样地清查法,即利用样地清查数据直接或间接测定生物量,再乘以相应的碳率系数推算得到碳储量^[6—7]。该方法虽能得到较为准确的碳储量数据,但对植被破坏较大、费时费力且获取信息有限。遥感技术的迅速发展为大尺度、多时相的植被信息获取提供了便捷,使快速、大区域地估算城市植被碳储量成为可能^[8—9]。

从现有研究来看,光学传感器、激光雷达和雷达数据均可以单独或联合用于植被碳储量估算。激光雷达和合成孔径雷达具有直接测量森林垂直结构的能力,能克服光学遥感中的光谱饱和现象,对植被生物物理和结构参数表现出更高的灵敏度,但较高的使用成本和复杂的数据处理步骤限制了其在大尺度的应用^[10—11]。光学遥感数据以其长时间序列、全球区域尺度覆盖和高重访周期的特点,在大区域尺度的植被监测中有不可替代的作用。常见的光学遥感影像有空间分辨率较低的 MODIS、AVHRR 数据,中空间分辨率的 Landsat 系列以及高空间分辨率的 GF、QuickBird 数据^[12]。其中, Landsat 8 卫星数据具备全球覆盖能力,获取质量稳定、公开免费,提取植被参数及物种群落特征等信息较为准确,成为估算植被碳储量的主要光学遥感数据源之一^[3,13]。

基于遥感影像的植被碳储量估测模型以参数或非参数模型为主^[14—15],参数模型如多元线性回归(Multiple Linear Regression, MLR)、逻辑回归等,简单高效但通常难以应对复杂问题。常用的非参数模型有随机森林、k 最近邻、支持向量机等,其模型稳健、可变性高,但需要较多训练样本且预测结果难以解释。在实际

调查中,由于地形、海拔、气候等因素的影响,植被调查数据往往与其地理位置有关,在空间上表现出明显的异质性^[16—17]。城市景观的复杂性使得植被分布具有更高的破碎化程度,进一步加剧了这种空间异质性。上述模型大多忽略了样地数据的空间变异,掩盖了变量间的局部差异性,从而增加实际估测中的误差,最终造成不合理的局部空间分布估测结果^[18]。为了探索数据的空间特性,地理加权回归(Geographically Weighted Regression, GWR)模型应运而生^[19—20]。GWR 将数据的空间特性纳入模型估测中,为分析回归关系的空间特征创造了条件^[21]。近年来,GWR 模型被广泛用于气象学、生态学、林学等多个领域^[22—24],并取得较好的应用效果。GWR 模型为局部模型,它考虑了变量的空间异质性,将全局参数分解为局部参数进行估计,具有比传统全局模型获得更合理的碳储量局部空间分布的潜力。空间核函数和带宽是决定 GWR 模型估测效果的重要参数^[25—26]。然而,在利用 GWR 进行植被碳储量反演时,多选用单一核函数及带宽确定方式,少有研究对比不同核函数及不同带宽选择方式在 GWR 模型拟合、系数估计及模型残差空间特征上的差异。

深圳市是我国 7 大碳排放权交易试点之一,近几年全面启动了“国家森林城市”高质量建设工作,准确估算其城市植被碳储量对深圳市城市建设、生态发展规划及实现碳达峰、碳中和目标意义重大^[27—28]。研究以广东省深圳市为研究区,基于 Landsat 8 OLI 遥感影像和植被碳储量野外调查数据,采用多个带宽确定方法并结合不同核函数分别构建植被碳储量遥感反演 GWR 模型,并与 MLR 进行比较。最终选取最优模型进行研究区植被碳储量反演制图,以期为我国城市植被碳储量遥感估算提供方法和技术参考。

1 材料与方法

1.1 研究区概况

深圳市地处广东省中南沿海地区,位于 $113^{\circ}43'—114^{\circ}38'E$, $22^{\circ}24'—22^{\circ}52'N$ 之间(图 1)。全市面积 1997.47 km^2 ,平均海拔 70—120 m。全境地势东南高、西北低,东南部主要为低山,中部和西北部为丘陵,西南部为冲击平原。境内母岩以花岗岩为主,东部和北部有较大面积砂页岩分布。研究区属亚热带海洋性气候,年平均气温 22.4°C ;雨量充沛,年平均降水量约 1933 mm,年平均湿度 72.3%;日照时间长,平均年日照时数约 2120 h。热带常绿季雨林与南亚热带季风常绿阔叶林为该市的地带性植被。深圳市自然环境优美,全市建成区绿化覆盖率 45%,人均公共绿地面积 16.01 m^2 ,森林面积 797 km^2 ,森林覆盖率达 40.21%。

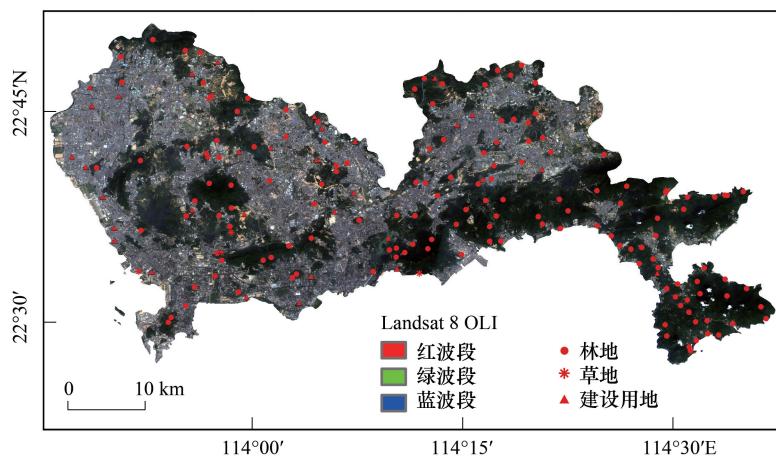


图 1 研究区位置及样地分布图

Fig.1 Geographical location and sample distribution of the study area

1.2 数据获取及处理

1.2.1 植被调查样地数据

参考 2014 年深圳市森林资源规划设计调查数据,根据各地类面积及其比例,按照分层随机抽样原则,在

研究区内确定大小为 25.8 m×25.8 m 的样地 180 个。其中,林地 140 个、草地 7 个及建设用地 33 个(图 1)。外业调查分别在 2014 年、2016 年和 2018 年完成。在实际调查中,地类为林地的样地分为乔木层、灌木层和草本层分别计算碳储量,最终求和得到样地碳储量。乔木层调查样方大小为 25.8 m×25.8 m,对样地内胸径大于 5 cm 的单木进行每木检尺,测量其胸径、树高及冠幅;灌木层调查样方大小为 2 m×2 m,均匀布设 5 个样方,记录树种、树高、地径及盖度;草本层调查样方大小为 1 m×1 m,均匀布设 5 个样方,记录其种类、平均高及盖度。地类为草地的样地调查方法及碳储量计算方法与林地中的草本层一致。为了获得更准确的实际植被分布,建设用地中包含了部分林木,如行道树、绿化带等以及部分草本,也将被调查以及统计。

1.2.2 Landsat 8 OLI 遥感数据

综合考虑影像与采样时间的邻近、云量等因素,通过美国地质调查局官网(<http://glovis.usgs.gov/>)获取了覆盖研究区的 2 景 Landsat 8 OLI 遥感影像,其成像时间分别为 2014 年 10 月 8 日和 2014 年 10 月 15 日,轨道号分别为 PATH 121/ROW 44 和 PATH 122/ROW 44,空间分辨率为 30 m。利用 ENVI 5.3 软件对 2 景 Landsat 8 OLI 遥感影像进行辐射定标、FLAASH 大气校正和影像镶嵌等预处理后,运用深圳市行政矢量边界裁剪得到研究区多光谱影像。

1.2.3 植被类型空间分布

根据 2014 年深圳市森林资源规划设计调查数据(二类调查数据),结合 0.5 m 空间分辨率遥感影像进行人工解译,得到植被类型空间分布矢量数据。利用 ArcGIS 10.6 软件提取研究区林地、草地等植被覆盖区域矢量数据,用于后续制作研究区植被碳储量空间分布图。

1.3 研究方法

1.3.1 样地植被碳储量估算

利用样地调查数据,结合广东省主要树种和软硬阔树种二元材积表^[29]分树种计算单木材积后累加得到该树种蓄积量并转换成单位面积蓄积量;根据《全国林业碳汇计量与监测技术指南》^[30]中蓄积量与生物量转换参数,采用换算因子连续函数法^[18],对同一样地按不同树种(组)换算并累加求和,得到样地单位面积乔木层生物量;最后根据优势树种(组)碳含率^[30]计算各样地的乔木层单位面积碳储量。

$$W_{\text{乔}} = aV + b \quad (1)$$

$$C = W \times CF \quad (2)$$

式中, $W_{\text{乔}}$ 为乔木层生物量 (Mg/hm^2), V 为每公顷蓄积量 (m^3/hm^2), a 和 b 为常数, 主要优势树种(组)取值见表 1; C 为各层碳储量 ($\text{Mg C}/\text{hm}^2$), W 为各层生物量 (Mg/hm^2), CF 为含碳率, 无量纲, 灌木层取 0.4672, 草本层取 0.3270, 主要优势树种(组)取值见表 1。

表 1 主要优势树种(组)对应生物量转换参数及含碳率取值

Table 1 Biomass conversion parameters and carbon fraction of main dominant tree species

树种(组) Tree species group	a /(Mg/m^3)	b /(Mg)	含碳率 Carbon fraction
杉木 <i>Cunninghamia lanceolata</i>	0.4652	19.1410	0.5127
马尾松 <i>Pinus massoniana</i>	0.5034	20.5470	0.5271
桉树 <i>Eucalyptus robusta</i>	0.8873	4.5539	0.4748
樟树 <i>Cinnamomum camphora</i>	0.9292	6.4940	0.4916
相思 <i>Acacia confusa</i>	0.9292	6.4940	0.4666
木荷 <i>Schima superba</i>	0.9292	6.4940	0.5115
栎类 <i>Quercus</i>	1.1453	8.5473	0.4798
其它松类 Other pines	0.5292	25.0870	0.4963
其它软阔类 Other soft broad-leaved trees	0.9292	6.4940	0.4502
其它硬阔类 Other hard broad-leaved trees	0.9292	6.4940	0.4901
针阔混 Mixed needle	0.8136	18.4660	0.4893

采用范文义等^[31]建立的不同区域灌木草本生物量与高度之间的模型计算灌木草本的生物量,乘以盖度得到样方灌草总生物量。

$$W_{灌} = 0.0398 \times h_1 - 0.3326 \quad (3)$$

$$W_{草} = 0.0175 \times h_2 - 0.2888 \quad (4)$$

式中, $W_{灌}$ 为灌木层生物量 (Mg/hm^2), $W_{草}$ 为草本层生物量 (Mg/hm^2), h_1 和 h_2 分别表示灌木和草本的平均高度 (m)。

为保证样地点与遥感影像像元的匹配, 获取更准确的影像信息, 分别对灌木层和草本层的 5 个样方总生物量取均值后乘以样地面积比, 得到样地灌草总生物量并转换成样地单位面积灌草生物量。利用灌木、草本层的平均碳含率^[30]换算得到各样地单位面积灌草总碳储量。最后将乔木层、灌木层及草本层碳储量相加得到样地单位面积总碳储量。

由于遥感影像成像时间和多数样地调查时间为 2014 年, 为减少估测误差, 依据树木生长方程将 2016 年及 2018 年的样地的平均胸径和树高反推至 2014 年的生长状态, 从而将所有样地植被碳储量换算至 2014 年水平。

1.3.2 遥感变量提取

植被指数由不同遥感光谱波段经线性或非线性组合构成, 对植被具有一定指示意义, 已被广泛用于定性和定量评价植被生长状况^[32—33]。对预处理后的影像数据进行波段计算与提取, 得到 7 个单波段反射率、42 个两波段比值植被指数、105 个三波段比值植被指数以及土壤调节植被指数、增强植被指数、归一化植被指数、大气阻抗植被指数、修正归一化差值植被指数、红绿植被指数等常用的植被指数共 160 个遥感变量, 其具体计算公式见表 2。

表 2 所采用的遥感变量计算方法

Table 2 Expressions of the adopted remote sensing variables

植被指数 Vegetation index	缩写 Abbreviation	计算公式 Calculation formula	参考文献 Reference
土壤调节植被指数 Soil adjusted vegetation index	SAVI	$(NIR-Red) (1+0.5) / (NIR+Red+0.5)$	[13]
增强植被指数 Enhanced vegetation index	EVI	$2.5(NIR-Red) / (NIR+6.0Red-7.5Blue+1)$	[33]
归一化植被指数 Normalized difference vegetation index	NDVI	$(NIR-Red) / (NIR+Red)$	[33]
大气阻抗植被指数 Atmospherically resistant vegetation index	ARVI	$[NIR-(Red-Blue)] / [NIR+(Red-Blue)]$	[33]
修正归一化差值植被指数 Modified normalized difference vegetation index	MNDVI	$\frac{NIR-Red}{NIR+Red} \left(1 - \frac{SWIR1-SWIR1_{min}}{SWIR1_{max}-SWIR1_{min}} \right)$	[33]
红绿植被指数 Red-green vegetation index	RGVI	$(Red-Green) / (Red+Green)$	[33]
两波段比值植被指数 Two-band simple ratio vegetation index	SR_{ij}	$Band_i / Band_j, (i,j=1,2,\dots,7, i \neq j)$	[33]
三波段比值植被指数 Three-band simple ratio vegetation index	SR_{ijk}	$Band_i / (Band_j + Band_k), (i,j,k=1,2,\dots,7, i \neq j \neq k, j < k)$	[33]
单波段反射率 Band reflectance	B_i	$Band_i, (i=1,2,\dots,7)$	[33]

1.3.3 遥感变量筛选

计算所有遥感变量与样地植被碳储量之间的 Pearson 相关系数矩阵, 选择与植被碳储量显著相关的变量, 采用逐步回归进行变量筛选。同时, 为了消除变量之间的共线性, 引入方差膨胀因子 (variance inflation factor, VIF)^[28] 进行共线性诊断, 阈值设为 10。最终所得变量用于后续所有模型的构建。

1.3.4 GWR 模型构建

GWR 模型是对普通全局回归模型的扩展,它将数据的空间特性以距离加权的方式纳入模型中,其基本形式为:

$$y_i = \beta_{i0} + \sum_{k=1}^n \beta_{ik}(u_i, v_i) x_{ik} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (5)$$

式中, (u_i, v_i) 为第 i 个样本点的空间坐标; β_{i0} 为第 i 个样本点的常数估计值; $\beta_{ik}(u_i, v_i)$ 为第 i 个样本点第 k 个自变量系数,与其空间位置有关; x_{ik} 为第 k 个自变量在样本 i 的值; ε_i 为残差,通常假定其服从独立正态分布。空间各点回归系数的求解如下:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (6)$$

式中, X 、 y 为各样本点的自变量与因变量矩阵; $W(u_i, v_i)$ 为样本点 i 的空间权重矩阵:

$$W(u_i, v_i) = \text{diag}(W_{i1}, W_{i2}, \dots, W_{in}) \quad (7)$$

对于空间权重核函数,研究利用 R 语言 GWmodel 函数包,分别构建了 Gaussian^[19]、Bi-square^[26] 和 Exponential^[34] 三种核函数。

校正 Akaike 信息准则 (Akaike Information Criterion corrected, AICc) 和最大空间自相关距离 (Maximum Spatial Autocorrelation Distance, MSAD) 是常用的两种带宽确定方式。校正 Akaike 信息准则通过选择 AICc 值最小时对应的带宽为最优带宽^[22]; 最大空间自相关距离通过创建半变异函数获得^[28]。交叉验证 (Cross-Validation, CV)^[34] 是目前主流的最优带宽求解方法之一,多用于气象学、海洋学等研究^[35–36],在植被碳储量反演中运用较少,其表达式如下:

$$CV(b) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (8)$$

式中, $\hat{y}_{\neq i}(b)$ 表示带宽为 b 时的模型预测值,其中 $\neq i$ 表示除 i 以外的其它样本。利用预测值进行计算,通过交叉验证,获得使 $CV(b)$ 最小的最优带宽 b 。

最终,将三种带宽选择方式与三种核函数进行组合,构建共 9 种 GWR 模型进行植被碳储量反演和精度评价。

1.3.5 精度评价

采用留一交叉验证^[37]对模型结果进行精度验证,即每次只留 1 个样本作为验证样本,余下的样本作为建模样本,直到所有样本都做过验证样本,最后对验证结果取平均作为泛化误差的估计。选用决定系数 R^2 (coefficient of determination)、均方根误差 (root mean square error, RMSE) 及平均绝对误差 (mean absolute error, MAE) 对模型进行精度评价^[37]。

采用 Moran 指数 (Moran I) 对不同模型预测值和实测值之间所得的残差进行空间自相关分析。Moran $I > 0$ 表示空间正相关,值越大空间相关性越明显; Moran $I < 0$ 表示空间负相关值,值越大空间差异越大; Moran I 接近 0 且在统计上不显著,则说明空间模式呈随机性。

1.3.6 非平稳性检验

对 GWR 模型参数估计值的空间非平稳性进行显著性检验,以判断这种非平稳性是空间数据本身固有,还是由于随机因素干扰。利用 Brunsdon 等^[19]提出的置信区间检验法对 GWR 模型各回归参数进行空间非平稳性检验,将 GWR 模型的局部参数与 MLR 模型的全局参数进行对比,若 GWR 模型参数估计的第 1 分位和第 3 分位值变化范围大于 MLR 模型的二倍标准误值,则可认为各回归参数具有显著非平稳性。

1.3.7 植被碳储量空间分布制图

以 Landsat 8 OLI 影像为数据源,选取最优 GWR 模型进行碳储量反演得到碳储量空间分布,利用植被类型空间分布数据对植被区域进行掩膜提取,得到深圳市植被碳储量空间分布图。

2 结果与分析

2.1 遥感变量选择及带宽确定

遥感变量与植被碳储量之间的 Pearson 相关系数值在 -0.378—0.383 之间,共有 111 个变量与碳储量显著相关($P<0.05$)。其中,相关性最高的前 3 个变量是两波段比值植被指数 SR_{64} 、三波段比值植被指数 SR_{436} 和三波段比值植被指数 SR_{324} ,与碳储量的相关系数分别为 0.383、-0.378 和 0.377。从各遥感变量的波段组成来看,单波段对碳储量的敏感度一般,但与红(Red)和短波红外波段(SWIR1)组合而成的比值植被指数对碳储量的敏感度相对较高。在相关性分析基础上,引入方差膨胀因子 VIF,采用逐步分析,最终筛选出三波段比值植被指数 SR_{324} 、 SR_{657} 及两波段比值植被指数 SR_{35} 共计 3 个遥感变量(表 3)。

表 3 变量筛选结果

Table 3 Variable selection results

变量 Variable	系数 Coefficient	标准误 Standard error	膨胀因子 Variance inflation factor
常数项 Constant	237.602	40.761	
三波段比值植被指数 SR_{324}	-189.193	27.947	4.117
两波段比值植被指数 SR_{35}	-201.556	51.110	2.692
三波段比值植被指数 SR_{657}	-39.908	12.776	2.475

SR :比值植被指数 Simple ratio vegetation index

图 2 为分别利用 AICc 与 CV 选择带宽,结合不同核函数构建 GWR 模型时所得的带宽曲线图,图 2 中虚线所对应带宽即为基于最小 AICc 值或 CV 值的最优带宽。最终,以 Gaussian、Bisquare 和 Exponential 为核函数,利用 AICc 确定的最优带宽分别为 11536、28131 和 11770,利用 CV 确定的最优带宽分别为 6347、17081 和 4519。通过创建半变异函数最终确定 MSAD 的最优带宽均为 28000,基于不同带宽选择方法所得最优带宽相差较大。

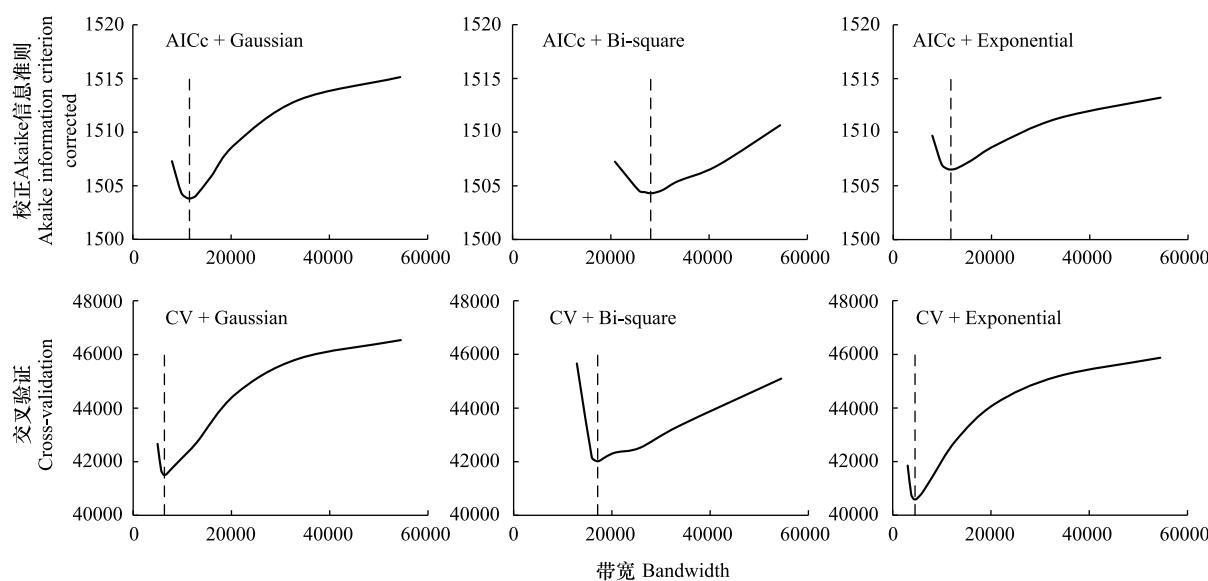


图 2 AICc 与 CV 带宽选择结果

Fig.2 Bandwidth selection results of using AICc and CV

AICc:校正 Akaike 信息准则 Akaike information criterion corrected; CV:交叉验证 Cross-validation

2.2 碳储量预测模型精度评价

由表 4 可知,MLR 模型的决定系数(R^2)为 0.298,均方根误差(RMSE)为 $15.886 \text{ Mg C}/\text{hm}^2$,平均决定误差

(MAE) 为 $11.721 \text{ Mg C}/\text{hm}^2$, 其模型残差的 Moran I 为 $0.172 (P < 0.01)$, 说明模型残差在空间中的分布并非完全随机, 且呈现显著正自相关。考虑局部差异的 GWR 预测效果均优于全局的 MLR ($P < 0.05$), 但不同带宽选择方法与不同核函数组合体现出不同的效果。利用 AICc 确定带宽, 以不同核函数分别构建的 3 个 GWR 模型在精度上无显著差异, 但以 Exponential 为核函数构建的 GWR 模型残差存在显著正相关; 利用 MSAD 确定带宽时, 以 Bi-square 为核函数所构建的 GWR 模型效果最佳, 且其模型残差呈随机分布; 利用 CV 确定带宽构建的 GWR 模型预测效果普遍优于另两种带宽确定方法所构建的 GWR 模型, 所构建的 3 个 GWR 模型, 其 Moran I 绝对值均小于 $0.05 (P > 0.1)$, 模型残差均无显著自相关。其中又以 CV 与 Exponential 核函数组合构建的 GWR 模型效果最佳, 其 R^2 为 0.697 , RMSE 为 $10.437 \text{ Mg C}/\text{hm}^2$, MAE 为 $7.401 \text{ Mg C}/\text{hm}^2$, 残差 Moran I 为 $-0.036 (P > 0.1)$, 模型的预测结果更接近实测值, 且模型残差之间相互独立。整体来看, 构建的 GWR 模型估计误差及残差自相关程度普遍低于 MLR 模型。这是由于样本地理位置信息以距离权重的方式参与了建模, 使得每个样地都具有一个独立的加权回归方程, 这能在一定程度上提高预测精度并有效减弱残差的空间自相关性。

表 4 模型估测精度及残差空间自相关

Table 4 Estimation accuracy and residual spatial autocorrelation of models

模型 Models	带宽选择 Bandwidth selection	核函数 Kernel	R^2	均方根误差 RMSE/ (Mg C/hm ²)	平均绝对误差 MAE/ (Mg C/hm ²)	Moran 指数 Moran index	P	Z 得分 Z score
多元线性回归 MLR			0.298	15.886	11.721	0.172	0.001	4.290
地理加权回归 GWR	AICc	Gaussian	0.451	14.051	10.228	0.059	0.071	1.550
		Bisquare	0.448	14.084	10.278	0.055	0.085	1.457
	MSAD	Exponential	0.462	13.904	10.072	0.076	0.038	1.974
CV	AICc	Gaussian	0.339	15.412	11.281	0.145	0.001	3.637
		Bisquare	0.449	14.073	10.271	0.055	0.086	1.443
	MSAD	Exponential	0.363	15.130	11.026	0.134	0.001	3.371
CV	MSAD	Gaussian	0.591	12.118	8.749	-0.025	0.364	-0.424
		Bisquare	0.563	12.536	9.087	-0.019	0.421	-0.293
	CV	Exponential	0.697	10.437	7.401	-0.036	0.251	-0.684

MLR: 多元线性回归 Multiple linear regression; GWR: 地理加权回归 Geographically weighted regression; AICc: 校正 Akaike 信息准则 Akaike information criterion corrected; CV: 交叉验证 Cross-validation; MSAD: 最大空间自相关距离 Maximum spatial autocorrelation distance; RMSE: 均方根误差 Root mean square error; MAE: 平均绝对误差 Mean absolute error

图 3 为碳储量实测值与所有模型预测值的拟合图, 各模型均存在一定低值高估与高值低估现象。对于碳储量大于 $50 \text{ Mg C}/\text{hm}^2$ 的样地, 各模型均存在明显的高值低估; 对于碳储量低值则普遍存在高估现象, 误差较大。此外, 对于碳储量极低值, 其模型估测结果存在少数预测值为负值。对比不同模型拟合曲线与理想曲线的偏离程度, MLR 模型偏离程度远大于各 GWR 模型的偏离程度; 从回归拟合曲线两侧的散点分布情况来看, 相比于 MLR 模型, GWR 模型的散点更为紧凑地聚集于拟合曲线的两侧。利用 AICc 确定带宽构建的 3 个 GWR 模型, 实测值与预测值散点分布相似度较高, 无显著性差异。利用 MSAD 确定带宽、以 Gaussian 和 Exponential 为核函数构建的 GWR 模型, 虽 R^2 显著高于 MLR 模型, 但从拟合图来看, 其预测结果并不存在明显的差异。利用 CV 确定带宽构建的 GWR 模型, 其曲线拟合及散点分布情况明显优于其它模型。其中, 又以 CV 与 Exponential 核函数组合构建的 GWR 模型, 其散点分布最为紧凑, 预测结果与实测结果偏离程度最小。

2.3 非平稳性检验

对所有 GWR 模型进行回归参数的空间非平稳性检验, 同一带宽选择方法所构建的 GWR 模型, 其参数非平稳性具有相似性(图 4)。采用 AICc 确定带宽构建的 GWR 模型, 常数项、变量 SR_{35} 和变量 SR_{324} 的系数估计值变化幅度均大于 MLR 的二倍标准误值, 变量 SR_{657} 虽存在一定的空间非平稳性, 但并不显著; 采用 MSAD 确

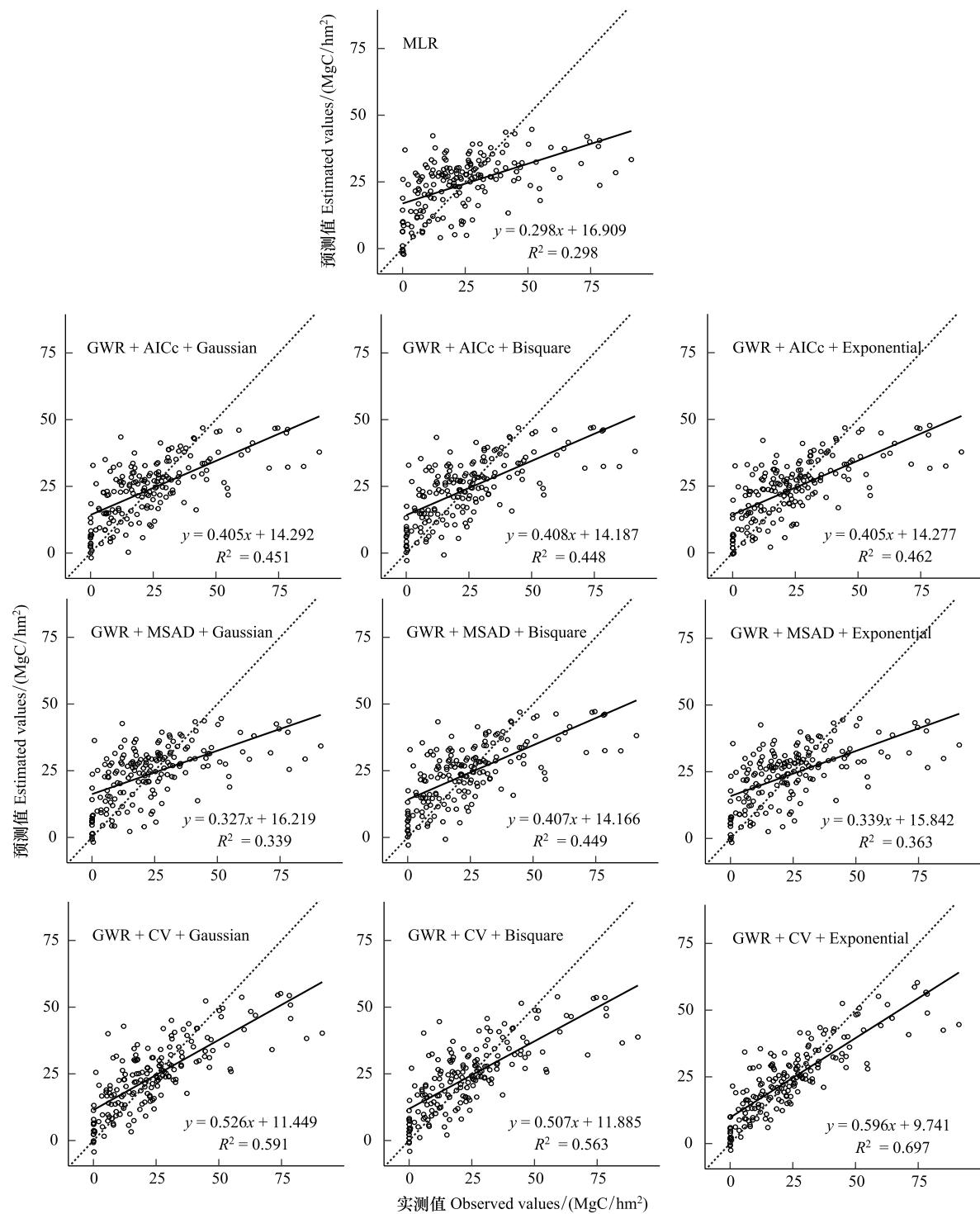


图3 植被碳储量估测模型拟合结果

Fig.3 Fitting results of vegetation carbon storage estimation models

MLR:多元线性回归 Multiple linear regression; GWR:地理加权回归 Geographically weighted regression; MSAD:最大空间自相关距离 Maximum spatial autocorrelation distance

定带宽,以 Gaussian 和 Exponential 为核函数构建的 GWR 模型,其变量均不具有显著非平稳性;采用 CV 选择带宽所构建的 GWR 模型,所有变量的系数估计第 1 分位数和第 3 分位数变化幅度均大于 MLR 的二倍标准误差,说明它们均具有显著的空间非平稳性,能较好反映空间异质性。

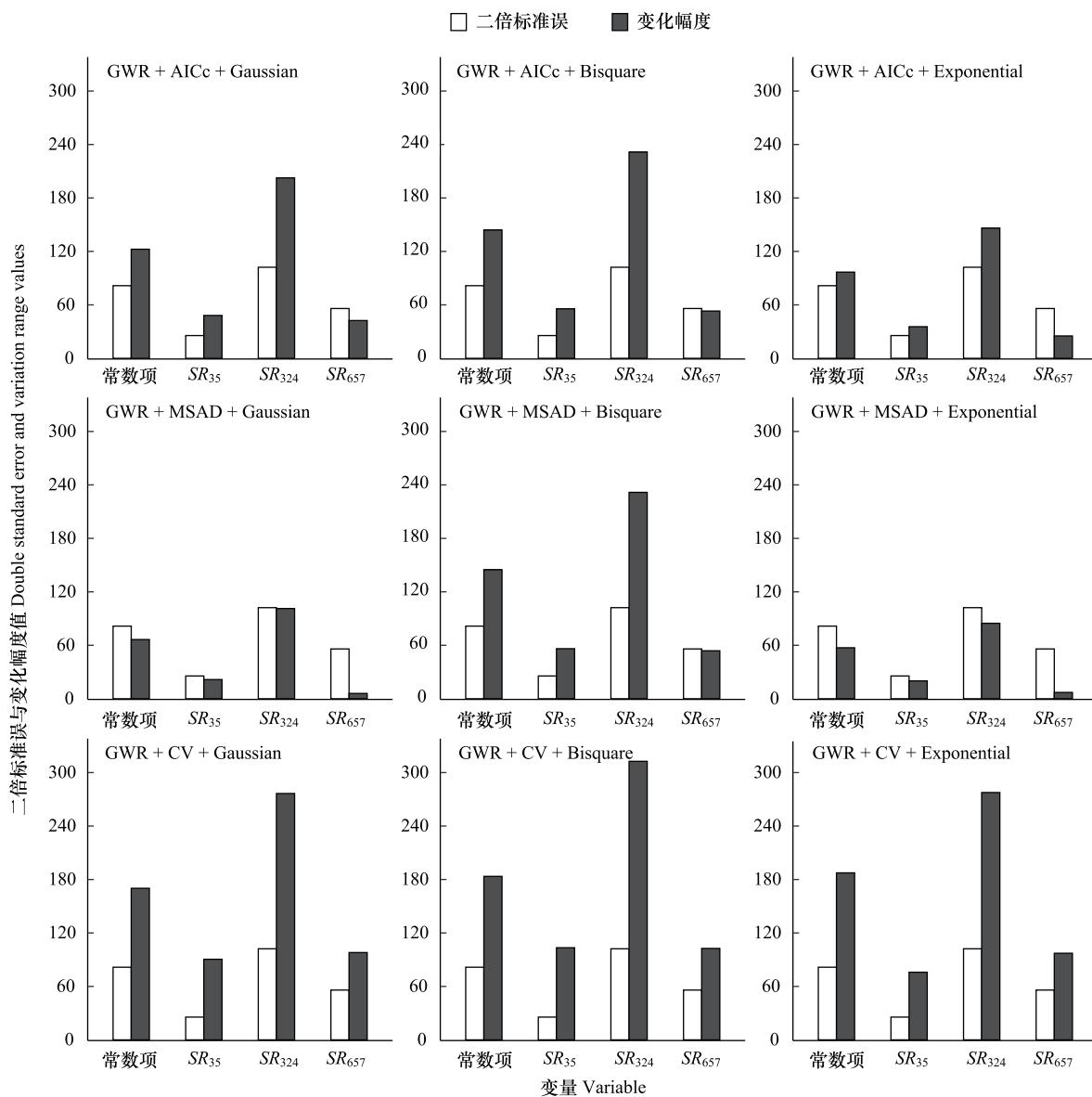


图 4 空间非平稳性检验

Fig.4 Stationary test of relationship

SR: 比值植被指数 Simple ratio vegetation index

图 5 为最优 GWR 模型得到的回归参数空间分布。变量回归系数的正负和数值在空间分布上的差异性，反映了各变量在不同区域对碳储量表现出的不同影响。常数项体现的主要原因是地理位置的影响，地理位置对大部分地区碳储量呈现不同程度的正向影响，东南部低山地区碳储量受地理位置的影响相比其它地区较大。SR₃₂₄ 变量回归系数大部分为负值，系数绝对值高值主要集中在东南部及北部的碳储量高值区，这表明碳储量高的地区，SR₃₂₄ 变量回归系数相对较高，SR₃₂₄ 变量对高碳储量地区的敏感度高于其它地区。除中部及南部地区外，SR₃₅ 变量大部分回归系数为负值，且回归系数由研究区中心区域向四周逐渐递减。SR₆₅₇ 回归系数均为负值，能负向反映植被碳储量，在保持其它条件不变时，较高植被碳储量处具有较低的 SR₆₅₇ 指数值。中部及南部的城市居民区的 SR₆₅₇ 回归系数较大，表明 SR₆₅₇ 变量更能反映低植被碳储量。

2.4 植被碳储量空间分布

深圳市植被碳储量空间分布值在 1.63—60.95 Mg C/hm² 之间(图 6)。碳储量值小于 10 Mg C/hm² 的区域

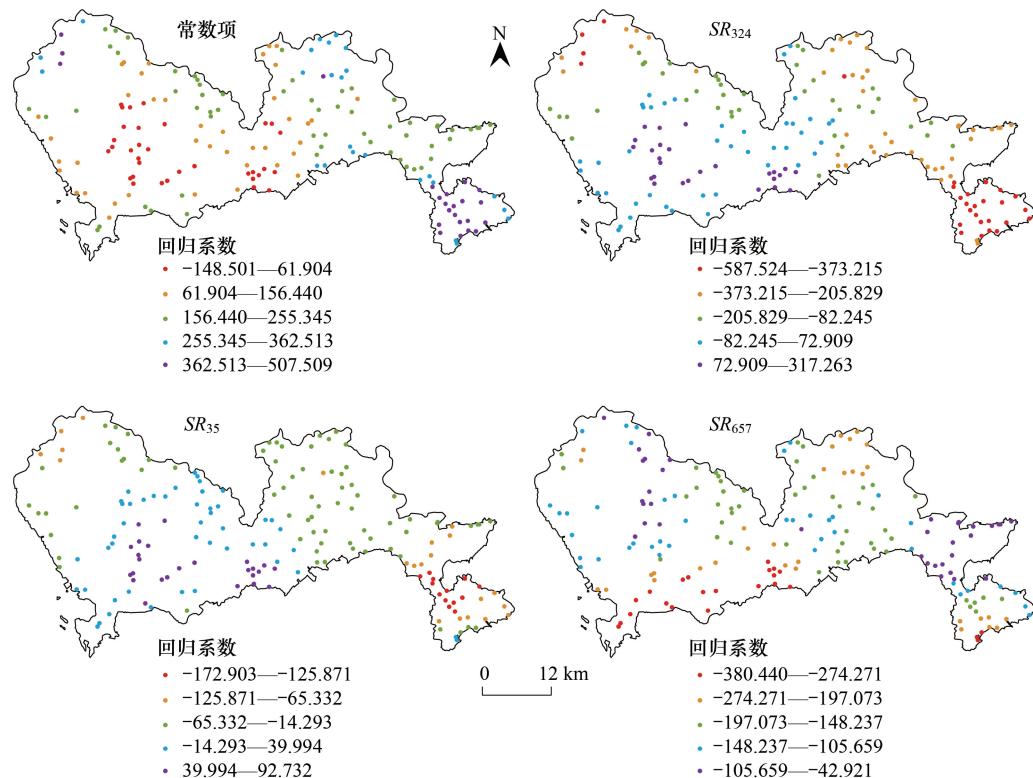


图 5 最优 GWR 模型回归参数系数空间分布图

Fig.5 The spatial distribution of regression parameter coefficients of the optimal GWR model

占深圳市植被区 7.61%, 其分布较为零散, 与深圳市草地地类区域重合度较高。在城市东南部低山、中西部丘陵、中南部和北部地区, 碳储量估计值大部分在 10—30 Mg C/hm²之间, 该部分地区植被类型主要为森林。碳储量值大于 30 Mg C/hm²的区域面积约为 276.7 km², 占深圳市植被区 30.47%, 分散分布于城市各地区, 主要地类为林地, 且植被类型多为阔叶林。从整体来看, 反演所得的碳储量空间分布与深圳市各植被类型空间分布大体一致, 与实际情况较符合。

3 讨论

碳储量能反映植被固碳增汇功能, 快速准确地估算植被碳储量对评估区域碳汇价值具有重要意义。本研究以 Landsat 8 OLI 影像为数据源, 构建了多个基于不同带宽选择方式及核函数的 GWR 模型进行深圳市植被碳储量估测研究。

结果表明 GWR 模型预测效果优于 MLR 模型, 这与 Hu 等^[23]和 Kupfer 等^[26]的研究结果一致。由于 MLR 模型为全局模型, 得到的回归参数估计是在整个研究区的平均值, 反映碳储量空间异质特征的能力有限; 而 GWR 模型为局部模型, 它考虑了变量的空间异质性, 将全局参数估计分解为局部参数进行估计, 在估测精度与保留样本空间特征上都显著优于 MLR, 且能有效降低模型残差的空间自相关性^[22,24]。Jiang 等^[3]利用改进 GWR 模型估算深圳市植被碳空间分布, 获得了最小 RMSE 值 13.280 Mg C/hm²。本研究中通过对带宽选择方

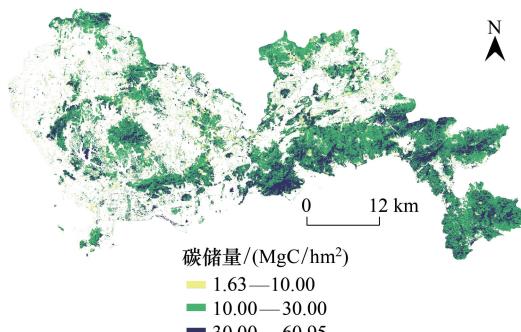


图 6 研究区植被碳储量空间分布图

Fig.6 Spatial distribution of vegetation carbon stocks in the study area

式和不同核函数进行优选得到最优 GWR 模型 RMSE 为 10.437 Mg C/hm^2 , 表明估测效果有明显改善。但从建模结果来看, 不同的带宽选择方式结合不同核函数所构建的 GWR 模型预测效果有显著差异。在实际操作中需结合具体情况综合考虑, 针对不同研究区和数据, 选择合适的权重函数确定最优带宽, 以保证 GWR 模型的估测效果^[38]。

由于样地调查时间跨度较大, 为尽可能降低因时间差异带来的影响, 依据树木生长方程将不同年度样地植被碳储量反推换算至同一水平, 但难以避免造成碳储量样地调查数据与实际情况的偏差, 从而影响碳储量估测。在构建碳储量估测模型时, 各模型均存在少数预测值为负值, 类似的问题在 Sun 等^[13]的研究中也存在, 其主要原因是植被碳储量与选定光谱变量之间的非线性关系。作为一种线性方法, GWR 模型使用局部最小二乘得到局部参数估计值, 虽在探索空间异质性方面具有优势, 但对于非线性关系解释能力有限。罗小波等^[39]提出一种局部非线性地理加权回归模型, 并将其应用于地表温度研究, 获得了比 GWR 线性模型更优的结果。可以考虑引入局部非线性地理加权回归模型用于植被碳储量等参数的反演研究。

植被碳储量空间分布图所得碳储量值在 $1.63\text{--}60.95 \text{ Mg C/hm}^2$ 之间, 这与 Sun 等^[13]所得的深圳市森林碳储量数值范围相似。但反演制图所得碳储量最大值与样地碳储量最大观测值 91.247 Mg C/hm^2 相差较大, 这很大程度上与样本数据的分布有关。在使用 GWR 作为样本外空间预测工具时, 是基于先前估计的带宽从未观测位置周围借来已知样本数据, 获得样本外预测。因 90% 以上样本数据植被碳储量值位于 $0\text{--}60 \text{ Mg C/hm}^2$ 之间, 对样本外预测结果造成了影响, 使数值集中于此区间。可结合以往的碳储量调查数据综合考虑, 使所布设样地尽可能包含多个地类多个级别碳储量值分布。

GWR 模型的特点在于局部参数的估计, 局部参数的非平稳性能体现模型反映空间异质性的能力^[26,38]。通过 CV 选择带宽构建的 GWR 模型效果相对较好, 其变量回归参数存在显著空间非平稳性。但通过 AICc 及 MSAD 选择带宽构建的 GWR 模型, 均存在回归参数空间非平稳性不显著的变量, 需将其作为常参数考虑。这种既包含变参数又包含常参数的 GWR 扩展模型, 通常称为混合 GWR 模型^[20-21]。混合 GWR 可提高常参数估计的精度和稳定性, 从而降低因常参数估计值而造成的模型预测误差, 具有获得高精度、稳定的植被碳储量估测结果的潜力。

4 结论

本研究以 Landsat 8 OLI 影像为数据源, 结合植被碳储量样地实测数据, 构建了 MLR 模型和多个基于不同带宽选择方式及核函数的 GWR 模型, 对深圳市植被碳储量进行估测和空间分布制图。得出以下主要结论:

(1) 局部的 GWR 模型优于全局的 MLR 模型, 不同的带宽选择方式结合不同核函数所构建的 GWR 模型, 在模型拟合效果上具有较大差别。以 CV 确定带宽、Exponential 为核函数组合构建的 GWR 模型效果最佳, 其 R^2 为 0.697, RMSE 为 10.437 Mg C/hm^2 , 相比其他 GWR 模型 RMSE 下降了 13.87%—32.28%, 估测效果有明显改善。

(2) 同一带宽选择方法所构建的 GWR 模型, 其参数非平稳性具有相似性。通过 CV 选择带宽构建的 GWR 模型, 其变量回归参数均存在显著空间非平稳性, 能较好反映空间异质性。

(3) 由最优 GWR 模型获得的深圳市植被碳储量空间分布表明植被碳储量高值和低值主要分布于森林和草地区域, 与深圳市植被覆盖情况基本一致, 能为深圳市植被碳储量遥感估算提供方法与技术参考。

参考文献(References) :

- [1] Jaman S, Zhang X, Islam F. Carbon storage and tree diversity in the urban vegetation of Dhaka city, Bangladesh: a study based on intensive field investigation. Arboricultural Journal, 2020, 42(2): 76-92.
- [2] Sun Y, Xie S, Zhao S Q. Valuing urban green spaces in mitigating climate change: a city-wide estimate of aboveground carbon stored in urban green spaces of China's Capital. Global Change Biology, 2019, 25(5): 1717-1732.
- [3] Jiang F G, Chen C S, Li C J, Kutia M, Sun H. A novel spatial simulation method for mapping the urban forest carbon density in southern China by the Google earth engine. Remote Sensing, 2021, 13(14): 2792.

- [4] Powell P A, Nanni A S, Názaro M G, Loto D, Torres R, Gasparri N I. Characterization of forest carbon stocks at the landscape scale in the Argentine dry Chaco. *Forest Ecology and Management*, 2018, 424: 21-27.
- [5] Li L, Zhou X S, Chen L Q, Chen L G, Zhang Y, Liu Y Q. Estimating urban vegetation biomass from sentinel-2A image data. *Forests*, 2020, 11(2): 125.
- [6] Sun W L, Liu X H. Review on carbon storage estimation of forest ecosystem and applications in China. *Forest Ecosystems*, 2020, 7(1): 4.
- [7] 赵苗苗, 赵娜, 刘羽, 杨吉林, 刘熠, 岳天祥. 森林碳计量方法研究进展. *生态学报*, 2019, 39(11): 3797-3807.
- [8] 汤煜, 石铁矛, 卜英杰, 石羽. 城市绿地碳储量估算及空间分布特征. *生态学杂志*, 2020, 39(4): 1387-1398.
- [9] Mitchell M G E, Johansen K, Maron M, McAlpine C A, Wu D, Rhodes J R. Identification of fine scale and landscape scale drivers of urban aboveground carbon stocks using high-resolution modeling and mapping. *Science of the Total Environment*, 2018, 622-623: 57-70.
- [10] 宋洁, 刘学录. 郴连山国家公园森林地上碳密度遥感估算. *干旱区地理*, 2021, 44(4): 1045-1057.
- [11] Sinha S, Mohan S, Das A K, Sharma L K, Jeganathan C, Santra A, Mitra S S, Nathawat M S. Multi-sensor approach integrating optical and multi-frequency synthetic aperture radar for carbon stock estimation over a tropical deciduous forest in India. *Carbon Management*, 2020, 11(1): 39-55.
- [12] 赵亚杰, 王立辉, 孔祥兵, 阴海明, 池泓, 黄进良. 基于 Sentinel-2 和 Landsat8 OLI 数据融合的土地利用分类研究. *福建农林大学学报: 自然科学版*, 2020, 49(2): 248-255.
- [13] Sun H, Qie G P, Wang G X, Tan Y F, Li J P, Peng Y G, Ma Z G, Luo C Q. Increasing the accuracy of mapping urban forest carbon density by combining spatial modeling and spectral unmixing analysis. *Remote Sensing*, 2015, 7(11): 15114-15139.
- [14] 巫明焱, 董光, 王艺积, 熊瑞东, 李悦, 程武学, 付志玺, 范曙峰. 川西米亚罗自然保护区森林地上碳储量遥感估算. *生态学报*, 2020, 40(2): 621-628.
- [15] 秦立厚, 张茂震, 袁振花, 杨海宾. 基于人工神经网络与空间仿真模拟的区域森林碳估算比较——以龙泉市为例. *生态学报*, 2017, 37(10): 3459-3470.
- [16] 张凌宇, 刘兆刚. 基于地理加权回归模型的大兴安岭中部天然次生林更新分布. *林业科学*, 2019, 55(11): 105-116.
- [17] Zhang L J, Shi H J. Local modeling of tree growth by geographically weighted regression. *Forest Science*, 2004, 50(2): 225-244.
- [18] 邱新彩, 郑冬梅, 王海宾, 安天宇, 许等平, 郝月兰, 彭道黎. 结合地统计学与 Landsat 8 影像的乔木林地上碳储量估算. *中南林业科技大学学报*, 2020, 40(11): 138-146.
- [19] Brunsdon C, Fotheringham S, Charlton M. Geographically weighted regression-modelling spatial non-stationarity. *The Statistician*, 1998, 47(3): 431-443.
- [20] Brunsdon C, Fotheringham A S, Charlton M. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 1999, 39(3): 497-524.
- [21] 覃文忠. 地理加权回归基本理论与应用研究[D]. 上海: 同济大学, 2007.
- [22] Yu H R, Gong H L, Chen B B, Liu K S, Gao M L. Analysis of the influence of groundwater on land subsidence in Beijing based on the geographical weighted regression (GWR) model. *Science of the Total Environment*, 2020, 738: 139405.
- [23] Hu X S, Zhang L Y, Ye L M, Lin Y H, Qiu R Z. Locating spatial variation in the association between road network and forest biomass carbon accumulation. *Ecological Indicators*, 2017, 73: 214-223.
- [24] 刘源, 邱振戈, 栾奎峰, 侍炯, 朱卫东, 刘鲁燕, 沈蔚, 曹彬才. 基于地理加权回归模型的水深反演算法研究. *海洋学研究*, 2018, 36(4): 35-42.
- [25] Zhang L J, Gove J H, Heath L S. Spatial residual analysis of six modeling techniques. *Ecological Modelling*, 2005, 186(2): 154-177.
- [26] Kupfer J A, Farris C A. Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. *Landscape Ecology*, 2007, 22(6): 837-852.
- [27] 邹琪, 孙华, 王广兴, 林辉, 谭一凡, 马中刚. 基于 Landsat 8 的深圳市森林碳储量遥感反演研究. *西北林学院学报*, 2017, 32(4): 164-171.
- [28] 陈川石. 基于局部样本量优化方法的城市植被碳密度空间模拟[D]. 长沙: 中南林业科技大学, 2019.
- [29] 刘琪璟. 中国立木材积表. 北京: 中国林业出版社, 2017.
- [30] 国家林业局林业碳汇计量监测中心. 全国林业碳汇计量监测技术指南(试行). 北京: 国家林业局, 2011.
- [31] 范文义, 张海玉, 于颖, 毛学刚, 杨金明. 三种森林生物量估测模型的比较分析. *植物生态学报*, 2011, 35(4): 402-410.
- [32] Zang Z, Wang G X, Lin H, Luo P. Developing a spectral angle based vegetation index for detecting the early dying process of Chinese fir trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 171: 253-265.
- [33] Sun H, Wang Q, Wang G X, Lin H, Luo P, Li J P, Zeng S Q, Xu X Y, Ren L X. Optimizing kNN for mapping vegetation cover of arid and semi-arid areas using Landsat images. *Remote Sensing*, 2018, 10(8): 1248.
- [34] Zhang W, Liu D, Zheng S J, Liu S Y, Loaiciga H A, Li W K. Regional precipitation model based on geographically and temporally weighted regression kriging. *Remote Sensing*, 2020, 12(16): 2547.
- [35] 李广超, 李如仁, 卢月明, 赵阳阳, 余博. 利用主成分分析法及地理加权回归模型分析 AOD 数据. *测绘通报*, 2018, (4): 50-56.
- [36] Wang Y A, Chen W, Kang Y Q, Li W, Guo F. Spatial correlation of factors affecting CO₂ emission at provincial level in China: a geographically weighted regression approach. *Journal of Cleaner Production*, 2018, 184: 929-937.
- [37] Jiang F G, Smith A R, Kutia M, Wang G X, Liu H, Sun H. A modified KNN method for mapping the leaf area index in arid and semi-arid areas of China. *Remote Sensing*, 2020, 12(11): 1884.
- [38] 郭含茹, 张茂震, 徐丽华, 袁振花, 陈田阁. 基于地理加权回归的区域森林碳储量估计. *浙江农林大学学报*, 2015, 32(4): 497-508.
- [39] 罗小波, 王书敏, 高阳华, 陈圆. 基于局部非线性地理加权回归模型的地表温度降尺度算法研究. *重庆邮电大学学报: 自然科学版*, 2020, 32(6): 1003-1011.