

# 思茅松天然林单木生物量地理加权回归模型构建

欧光龙<sup>1,2</sup>, 王俊峰<sup>2</sup>, 肖义发<sup>2</sup>, 胥 辉<sup>2\*</sup>

(1. 东北林业大学林学院, 黑龙江 哈尔滨 150040; 2. 西南林业大学西南地区生物多样性保育国家林业局 重点实验室, 云南 昆明 650224)

**摘要:**通过调查云南省思茅区思茅松天然林 63 株思茅松单木的地上部分干、枝、叶生物量数据,并测定其中 30 株的根系生物量数据。基于普通最小二乘模型选型,采用地理加权回归的方法构建思茅松单木树干生物量、树枝生物量、树叶生物量和地上部分生物量,以及根系生物量和整株生物量模型。结果表明:(1)地理加权回归模型(GWR)的决定系数( $R^2$ )大于普通最小二乘(OLS)模型,且 GWR 模型拟合的  $R^2$  值除树叶生物量模型外,其余生物量维量模型均大于 0.950; Akaike 信息指数(AIC)值小于普通最小二乘(OLS)模型,平均相对误差(EE)和平均相对误差绝对值(RMA)的绝对值除树枝生物量外均小于 OLS 模型,说明 GWR 模型拟合效果优于 OLS 模型;(2)地理加权回归模型拟合在一定程度上克服了 OLS 在拟合生物量模型中存在的异方差问题。

**关键词:**生物量;地理加权回归;普通最小二乘;思茅松;云南省思茅区

中图分类号:S711

文献标示码:A

## Modeling Individual Biomass of *Pinus kesiya* var. *langbianensis* Natural Forests by Geographically Weighted Regression

OU Guang-long<sup>1,2</sup>, WANG Jun-feng<sup>2</sup>, XIAO Yi-fa<sup>2</sup>, XU Hui<sup>2</sup>

(1. School of Forestry, Northeast Forestry University, Harbin 150040, Heilongjiang, China; 2. Key Laboratory of Biodiversity Conservation in Southwest China of State Forest Administration, Southwest Forestry University, Kunming 650224, Yunnan, China)

**Abstract:** The stem, branch, leaf biomass of 63 sampling trees, and root biomass of 30 trees at Simao pine (*Pinus kesiya* var. *langbianensis*) natural forest were investigated in Simao district of Yunnan Province. Based on the model selected by ordinary least square (OLS), the models of the tree stem, branch, leaf biomass, aboveground biomass, root biomass and whole tree biomass were built by geographically weighted regression (GWR). The results showed that: (1) the values of the coefficient of determination ( $R^2$ ) of GWR were greater than that of OLS models, and the  $R^2$  of the GWR models were greater than 0.950 except the leaf biomass model. Akaike's information criterions (AIC) of GWR were less than that of OLS models, the absolute value of the mean relative error (EE) and the mean absolute relative error (RMA) were less than that of OLS model except the branch biomass model. So the fitting effect of GWR outperforms OLS models. (2) For individual tree biomass models, GWR overcame the heteroscedasticity of the OLS models at a certain extent.

**Key words:** biomass; geographically weighted regression (GWR); ordinary least square(OLS); *Pinus kesiya* var. *langbianensis*; Simao; Yunnan Province

由于很多观测数据与地理位置有关,地理位置的邻近关系使得数据间具有空间相关性<sup>[1-2]</sup>,为了

收稿日期:2013-10-20

基金项目:国家自然科学基金项目(31160157)和云南省基金应用基础研究计划项目(2012FD027)

作者简介:欧光龙(1983—)男,云南镇雄人,在读博士,主要从事森林经理及森林生态方面的研究。Email:olg2007621@gmail.com

\* 通讯作者:博士,教授,博导,主要从事森林测计学方面的研究。Email:zyxy213@126.com

将数据的空间特性纳入回归模型中予以分析, Brunson 等<sup>[2]</sup>提出了地理加权回归模型(Geographically Weighted Regression Model, GWR), 该模型在环境学、地理学、气象学、生态学以及林学等学科中得到广泛应用<sup>[1-7]</sup>。GWR 在林业的应用上, Zhang 等用地理加权回归构建树木的生长和直径之间关系的局域多元模型, 研究树木在环境竞争中生长的变化情况<sup>[8]</sup>; Wang 等应用地理加权回归模型来估计中国森林净初级生产力<sup>[9]</sup>; 顾凤岐和赵倩<sup>[10]</sup>利用地理加权回归模型建立林木生长模型, 并认为其表现优于一般线性回归模型。但目前应用地理加权回归模型构建林木生物量模型的研究却鲜有报道。

本研究以云南省思茅区思茅松天然林为研究对象, 采用地理加权回归的方法构建其单木树干生物量、树枝生物量、树叶生物量和地上部分生物量, 以及根系生物量和整株生物量模型。

## 1 材料与方法

### 1.1 数据来源

研究对象为思茅松 (*Pinus kesiya* var. *langbi-*

*anensis*), 自然分布于云南热带北缘和亚热带南部半湿润地区<sup>[11]</sup>, 是我国亚热带西南部山地的代表种<sup>[12]</sup>, 因用途广泛, 生长迅速, 近年来已成为云南重要的人工造林树种。思茅松林作为云南特有的森林类型, 主要分布于云南哀牢山西坡以西的亚热带南部, 其分布面积和蓄积量均占云南省有林地面积的 11%<sup>[11]</sup>, 具有重要的经济价值、森林生态服务功能和碳汇效益<sup>[13-14]</sup>。

数据为 2011 年和 2012 年在云南省普洱市思茅区调查数据, 调查思茅松天然林样地 15 个, 标准木 63 株。记录样木位置的经纬度坐标值, 以及样木树龄 ( $t$ )、树高 ( $H$ )、胸径 ( $D$ )、冠长 ( $CL$ )、冠幅 ( $CW$ ) 等数据, 并计算其材积 ( $V$ ); 分别测定单木树干生物量 ( $W_s$ ) (木材生物量和树皮生物量之和)、树枝生物量 ( $W_b$ )、树叶生物量 ( $W_l$ ) 和树木地上部分生物量 ( $W_a$ ), 并调查其中 30 株根系生物量 ( $W_r$ ), 并得到 30 株样木的总生物量 ( $W_t$ )。调查数据见表 1。

表 1 样木数据基本特征汇总

变量	样本数	平均值	最大值	最小值	标准差
树龄/年	63	41.03	76.00	14.00	17.60
胸径/cm	63	22.36	51.30	5.90	11.39
树高/m	63	16.51	27.40	6.10	5.39
冠长/m	63	7.69	16.40	2.30	3.59
冠幅/m	63	7.19	16.94	2.76	2.91
材积/m <sup>3</sup>	63	0.52	2.34	0.01	0.58
树干生物量/kg	63	239.11	1 307.25	4.80	285.97
树枝生物量/kg	63	44.82	286.36	1.07	55.26
树叶生物量/kg	63	5.07	32.65	0.11	5.35
地上部分生物量/kg	63	290.85	1 553.79	6.84	343.23
根系生物量/kg	30	45.87	141.59	2.22	34.92
总生物量/kg	30	272.63	958.77	19.38	223.41

### 1.2 研究方法

1.2.1 基于 OLS 模型选型 采用 SAM v4.0<sup>[15]</sup> 进行普通最小二乘 (Ordinary least square, OLS) 选型, 采用一般线性回归分析, 选取胸径 ( $D$ )、树高 ( $H$ )、冠长 ( $CL$ )、冠幅 ( $CW$ )、材积 ( $V$ ), 以及  $D^2H$  和  $CW^2CL$  2 个复合变量共 7 个变量代入模型计算。分别构建思茅松单木各维量生物量模型 (含树干、树枝、树叶、地上部分、根系和总生物量), 以 Akaike 信息指数 (Akaike information criterion, AIC) 准则进行模型基本形式选型, AIC 值越小模型

越好。

1.2.2 基于 GWR 的单木生物量模型构建 基于 OLS 模型选型后的最佳模型形式, 采用 SAM v4.0<sup>[16]</sup> 进行地理加权回归 (GWR) 拟合思茅松单木各维量生物量模型。

地理加权回归模型是一种对普通线性回归模型的扩展, 它将数据的地理位置嵌入到回归参数之中, 即

$$y_i = \beta_0(\mu_i, \nu_i) + \sum_{k=1}^p \beta_k(\mu_i, \nu_i) x_{ik} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (1)$$

式中: $\beta_0(u_i, v_i)$ 为第*i*个采样点的坐标(经纬度), $\beta_k(u_i, v_i)$ 是第*i*个采样点上的第*k*个回归参数, $x_{ik}$ 是位置在 $(u_i, v_i)$ 处的自变量值; $\varepsilon_i$ 为独立同分布的误差项,通常假定其服从 $N(0, \sigma^2)$ 。

1.2.3 模型评价 模型构建过程中选用 Akaike 信息指数(AIC)、决定系数(Coefficient of determination,  $R^2$ )、总相对误差(Sum relative error, RS)、平均相对误差(Mean relative error, EE)和平均相对误差绝对值(Absolute mean relative error, RMA)5个评价指标对模型进行评价。

$$AIC = -2\ln L(\hat{\theta}_L, x) + 2q \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

$$RS = \frac{\sum (y_i - \hat{y}_i)}{\sum \hat{y}_i} \times 100\% \quad (4)$$

$$EE = \frac{1}{N} \sum \left( \frac{y_i - \hat{y}_i}{\hat{y}_i} \right) \times 100\% \quad (5)$$

$$RMA = \frac{1}{N} \sum \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \times 100\% \quad (6)$$

式中: $\hat{\theta}_L$ 为模型的似然函数 $L(\theta, x)$ 中 $\theta$ 的极大似然估计, $\theta$ 的维数为*p*, $x$ 为随机样本, $q$ 为未知参数个数, $y_i$ 为实测值, $\hat{y}_i$ 为估计值, $\bar{y}_i$ 为样本平均值, $N$ 为样本容量。

## 2 结果分析

### 2.1 基于 OLS 的基本模型选型

通过 OLS 的线性回归,引入胸径( $D$ )、树高( $H$ )、冠长( $CL$ )、冠幅( $CW$ )、材积( $V$ ),以及 $D^2H$ 和 $CW^2CL$ 2个复合变量共7个变量代入模型计算。以AIC值为依据,分别确定各生物量维量的模型形式。最优模型引入变量及拟合参数情况见表2。

表2 基于 OLS 的各变量的最佳线性模型及其参数情况

维度	模型	$R^2$	AIC
树干生物量	$W_s = 46.6488 + 0.0058 D^2H + 0.0447 CW^2CL - 10.7279 CL + 331.5891 V$	0.975	669.668
树枝生物量	$W_b = 11.4291 + 3.3748 D - 2.8149 H - 3.4821 CL + 0.0009 D^2H + 0.0271 CW^2CL$	0.890	557.336
树叶生物量	$W_l = 2.0381 + 0.0003 D^2H - 0.0028 CW^2CL$	0.544	345.523
地上生物量	$W_a = 62.5958 - 13.2901 CL + 0.0079 D^2H + 0.0649 CW^2CL + 373.2571 V$	0.974	691.200
根系生物量	$W_r = -16.5702 + 2.1759 D - 0.0029 D^2H + 101.8151 V$	0.833	256.102
总生物量	$W_t = -83.6282 + 8.7429 D - 0.0172 D^2H + 0.0976 CW^2CL + 698.2446 V$	0.977	311.665

从表2中可以看出, $D^2H$ 均引入到各个生物量维度模型中, $CW$ 均未引入到各生物量维度模型中;树叶生物量模型引入 $CW^2CL$ ,根系生物量模型还包括了 $D$ 和 $V$ ,总生物量模型包括了上述4个变量;树枝生物量包括了所有变量中除 $CW$ 和 $V$ 外的其余5个变量,树干生物量和地上生物量模型包括了 $CL$ , $D^2H$ , $CW^2CL$ 和 $V$ 4个变量。此外,从决定系数上看,树干生物量、地上生物量和总生物量均高于0.970,而树叶生物量仅为0.544。

### 2.2 基于 GWR 的生物量模型构建

以 OLS 模型选择为基础,采用地理加权回归拟合各生物量维度模型,各模型回归参数见表3。从表2中可以看出,OLS模型给出的拟合参数值是不变的,而表3中的GWR模型的回归参数会随地理位置的变化而变化,SAM程序给出了各个模型的回归参数的变化范围,包括了最小值、下四分位值、中位值、上四分位值和最大值,表4列出了GWR模型回

归分析的拟合参数值的范围及中位值。从表中可以看出,GWR模型的回归参数的分布范围涵盖了OLS模型的参数值。

此外,通过对GWR与OLS拟合的6个指标进行对照(见表4)。GWR回归模型拟合的所有6个生物量维量的AIC值均小于OLS模型,而模型的决定系数值均大于OLS模型;由于OLS回归分析是无偏估计,其残差和为零,因此,OLS模型的总相对误差值为零,而GWR模型拟合的总相对误差(RS)的绝对值均大于零;平均相对误差(EE)和平均相对误差绝对值(RMA)表现相对一致,除树枝生物量的GWR模型的EE值的绝对值和RMA值略高于OLS模型外,其余均是GWR模型的绝对值小于OLS模型,其中总生物量的GWR模型的EE值仅为0.630,树干生物量的GWR模型的EE值为-1.980,地上部分生物量的EE值为-3.889,总生物量的GWR模型的RMA值仅为2.870(远小于OLS模型的16.653)。

表3 基于GWR的线性模型的生物量模型回归参数

维度	截距	<i>D</i>	<i>H</i>	<i>CL</i>	<i>V</i>	<i>D<sup>2</sup>H</i>	<i>CW<sup>2</sup>CL</i>	
树干生物量	最小值	-39.520 6	—	—	-39.740 4	58.797 1	-0.003 5	-0.073 1
	中位值	29.695 4	—	—	-7.194 8	333.055 6	0.008 0	0.030 4
	最大值	176.459 9	—	—	6.091 7	550.390 6	0.018 4	0.203 5
树枝生物量	最小值	-23.825 8	2.423 9	-5.132 1	-4.894 8	—	-0.001 1	-0.031 8
	中位值	10.393 3	3.325 4	-3.896 0	-0.476 3	—	0.002 3	0.008 0
	最大值	31.915 0	5.096 6	-2.162 7	3.036 3	—	0.004 7	0.034 3
树叶生物量	最小值	1.464 1	—	—	—	—	0.000 0	-0.006 4
	中位值	1.903 8	—	—	—	—	0.000 4	-0.002 0
	最大值	3.276 2	—	—	—	—	0.000 7	0.002 4
地上生物量	最小值	-33.777 4	—	—	-41.898 1	172.205 5	0.002 1	-0.095 2
	中位值	41.975 9	—	—	-8.868 7	285.843 0	0.013 1	0.032 1
	最大值	191.2373	—	—	6.132 2	451.878 4	0.020 1	0.262 8
根系生物量	最小值	-72.933 8	-7.233 7	76.754 1	—	—	-0.016 7	—
	中位值	14.355 1	-1.255 0	120.664 3	—	—	-0.000 3	—
	最大值	69.428 9	5.878 9	575.964 9	—	—	0.010 5	—
总生物量	最小值	-348.880 0	-33.958 1	—	—	132.922 1	-33.958 1	-0.082 7
	中位值	-28.283 8	2.873 1	—	—	709.538 4	2.873 1	-0.006 2
	最大值	471.9949	27.233 4	—	—	944.056 7	27.233 4	0.118 7

表4 基于GWR和OLS的线性模型的生物量模型拟合情况对照

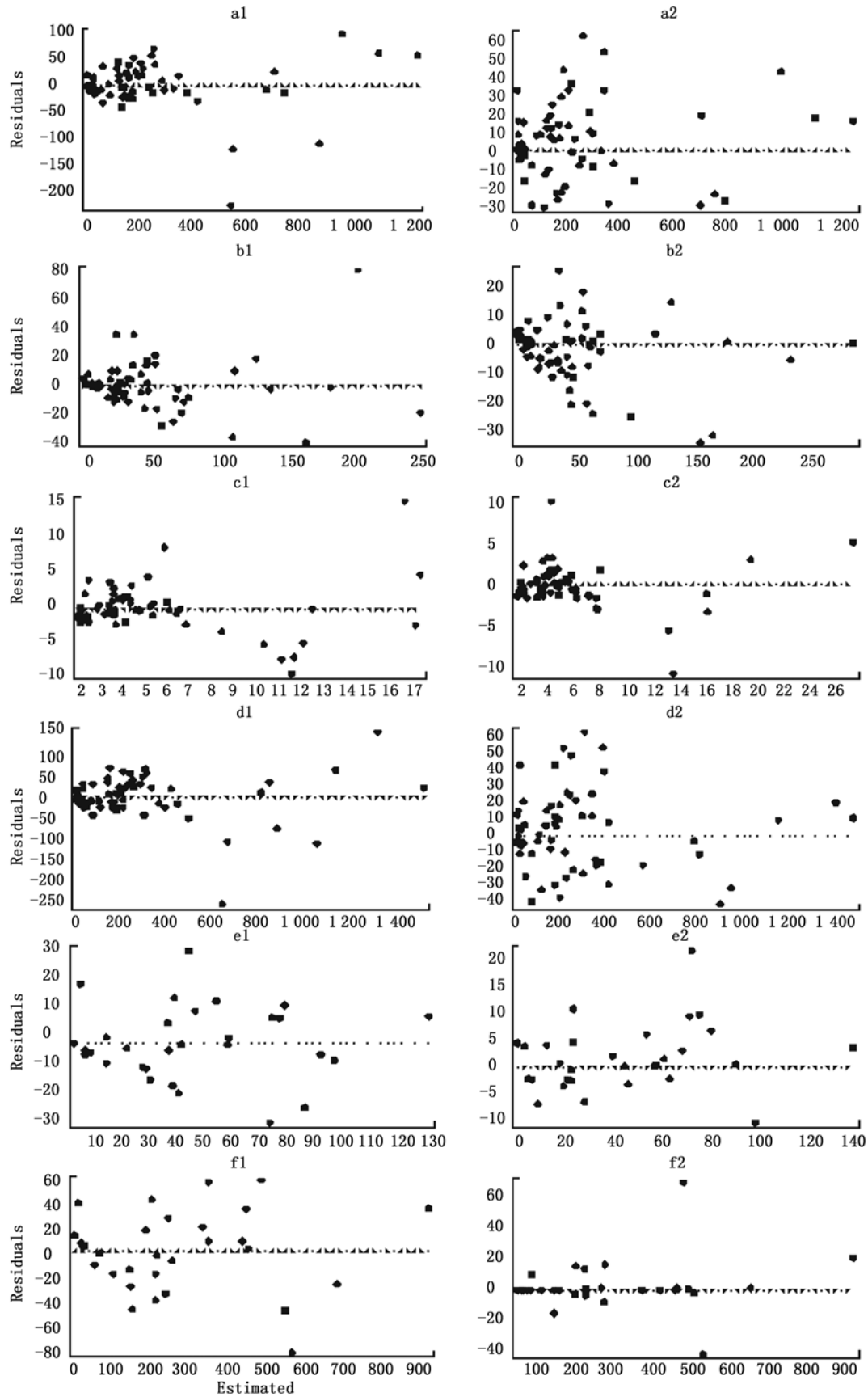
维度	<i>AIC</i>	<i>R<sup>2</sup></i>	<i>RS</i>	<i>EE</i>	<i>RMA</i>	
树干生物量	GWR	617.789	0.994	0.029	-1.980	18.153
	OLS	669.668	0.975	0.000	-11.174	25.004
树枝生物量	GWR	549.262	0.960	-0.092	-8.880	41.963
	OLS	557.336	0.890	0.000	-8.803	40.141
树叶生物量	GWR	334.061	0.731	-0.096	-72.996	90.339
	OLS	345.523	0.544	0.000	-77.677	101.703
地上生物量	GWR	631.243	0.995	0.016	-3.889	17.918
	OLS	691.200	0.974	0.000	-8.825	25.728
根系生物量	GWR	256.102	0.955	0.118	-11.561	28.756
	OLS	327.351	0.833	0.000	-23.494	42.224
总生物量	GWR	499.237	0.994	0.032	0.630	2.870
	OLS	311.665	0.977	0.000	1.863	16.653

此外,从残差图(图1)可以看出。各生物量维度的GWR拟合模型的残差分布区间明显小于OLS拟合模型。从残差的分布情况还可以看出拟合模型的方差齐性,6个生物量的OLS拟合模型均存在一定的异方差现象,而GWR模型在一定程度上克服了OLS模型的异方差。

### 3 结论与讨论

地理加权回归(GWR)模型通过在线性回归模

型中假定回归系数是观测点地理位置的任意函数,将数据的空间特性纳入模型,为分析回归关系的空间特征创造了条件<sup>[7]</sup>。忽略空间异质性可能会带来有偏参数估计,无意义的显著性检验和次佳的估计<sup>[8]</sup>。本文通过采用GWR模型对思茅松单木地上部分各维度、地上部分总生物量、根系生物量和总生物量进行线性拟合。



注:a 表示树干生物量,b 为树枝生物量,c 为树叶生物量,d 为地上生物量,e 为根系生物量,f 为总生物量;1 表示 OLS 模型,2 表示 GWR 模型。

图 1 基于 GWR 与 OLS 的生物量模型残差

通过 GWR 线性拟合 6 个生物量维量的线性模型,其拟合模型的决定系数( $R^2$ )均大于 OLS 模型,且 GWR 模型拟合的  $R^2$  值除树叶生物量模型外,其余生物量维度模型均大于 0.950,而树干生物量、地上部分生物量和总生物量模型都大于 0.990;6 个生物量维量的 GWR 模型的 AIC 值均小于 OLS 模型;平均相对误差(EE)的绝对值和平均相对误差绝对值(RMA)除树枝生物量模型外,均小于 OLS 模型,尤其是在树干生物量和总生物量的 GWR 拟合模型明显优于 OLS 模型;由于普通最小二乘(OLS)是无偏估计,其拟合的残差和为零,因此,在总相对误差(RS)的绝对值上,GWR 均大于 OLS 回归模型。通过这 5 个模型评价指标的分析可以看出,GWR 模型在拟合生物量树干、树叶、地上部分生物量、根系和总生物量模型上优于 OLS 模型,这与顾凤岐和赵倩<sup>[11]</sup>通过 GWR 拟合林木生长数据时得出的结果是一致的。

树枝生物量的 GWR 模型决定系数(0.960)要高于 OLS 模型(0.890),通过残差图可以看出,该维量的 GWR 模型的残差分布区间小于 OLS 模型,但其模型的平均相对误差(EE)的绝对值和平均相对误差绝对值(RMA)却高于 OLS 模型,造成这一现象的原因可能与误差结构有关,从而影响模型精度。本研究假设基于地理加权回归(GWR)的生物量模型随机误差项服从正态分布,且误差项是独立分布的,但随机误差的自相关性普遍存在<sup>[4,16]</sup>,因此,在采用 GWR 模型拟合生物量的过程中,将进一步研究如何降低随机误差自相关性带来的影响。此外,GWR 和 OLS 线性模型拟合树叶生物量的决定系数均不高,分别为 0.731 和 0.544,这可能是线性模型不能很好拟合树叶生物量,而基于 GWR 的非线性生物量模型将在今后研究中进一步探索。

林业应用中的许多回归模型,包括生物量模型,其误差项的方差一般都随自变量的变化呈现有规律的增大或减小,出现异方差问题<sup>[17-18]</sup>。异方差的存在,使利用普通回归进行的参数估计有偏,导致误差和参数的变动系数增大,使模型的正确性受到质疑<sup>[17]</sup>。通常采用加权最小二乘回归可以在一定程度上消除异方差<sup>[17]</sup>,但是权函数的选取是其消除异方差的关键。本文采用 GWR 模型拟合生物量模型,通过图示法可以直观看出,GWR 模型在一定程度上克服了 OLS 拟合模型中存在的异方差问题,但图示法仅能对异方差进行定性判定,对 GWR 模型

和 OLS 模型异方差的改进效果进行定量评价将在今后的研究中进一步探索。

#### 参考文献:

- [1] Leung Y, Mei C L, Zhang W X. Statistical tests for spatial nonstationarity based on the geographically weighted regression model [J]. *Environment and Planning - Part A*, 2000, 32(1): 9-32
- [2] Brunsdon C, Fotheringham A S, Charlton M E. Geographically weighted regression: a method for exploring spatial nonstationarity [J]. *Geographical Analysis*, 1996, 28(4): 281-298
- [3] Zhang L, Bi H, Cheng P, et al. Modeling spatial variation in tree diameter - height relationships [J]. *Forest Ecology and Management*, 2004, 189(1): 317-330
- [4] Zhang L, Gove J H, Heath L S. Spatial residual analysis of six modeling techniques [J]. *Ecological Modelling*, 2005, 186(2): 154-177
- [5] Zhang L, Ma Z, Guo L. An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models [J]. *Forest Science*, 2009, 55(6): 533-548
- [6] Jetz W, Rahbek C, Lichstein J. Local and global approaches to spatial data analysis in ecology [J]. *Global Ecology and Biogeography*, 2005, 14(1): 97-98
- [7] 玄海燕,黎锁平,刘树群. 地理加权回归模型及其拟合[J]. *甘肃科学学报*, 2009, 19(1): 51-52
- [8] Zhang L J, Shi H J. Local modeling of tree growth by geographically weighted regression [J]. *Forest Science*, 2004, 50(2): 225-244
- [9] Wang Q, Ni J, Tenhunen J. Application of a geographically weighted regression analysis to estimate net primary production of Chinese forest ecosystems [J]. *Global Ecology and Biogeography*, 2005, 14(4): 379-393
- [10] 顾凤岐,赵倩. 林木生长关系的 GWR 模型 [J]. *东北林业大学学报*, 2012, 40(6): 129-130, 140
- [11] 云南森林编写委员会. 云南森林 [M]. 昆明: 云南科技出版社 & 中国林业出版社, 1988
- [12] 西南林学院, 云南省林业厅. 云南树木图志 [M]. 昆明: 云南科技出版社, 1988
- [13] 温庆忠,赵远藩,陈晓鸣,等. 中国思茅松林生态服务功能价值动态研究 [J]. *林业科学研究*, 2010, 23(5): 671-677
- [14] 李江,孟梦,邱琼,等. 思茅松中幼龄人工林生物量及生产力动态 [J]. *东北林业大学学报*, 2010, 38(8): 36-38
- [15] Rangel T F, Diniz - Filho J A F, Bini L M. SAM: A comprehensive application for Spatial Analysis in Macroecology [J]. *Ecography*, 2010, 33(1): 1-5
- [16] 黄视玲. 地理加权空间经济计量模型的 GMM 估计及区域金融发展收敛性实证研究 [D]. 广州: 华南理工大学, 2012
- [17] 张会儒,唐守正,胥辉. 关于生物量模型中的异方差问题 [J]. *林业资源管理*, 1999(1): 46-49
- [18] 曾伟生. 关于加权最小二乘法中权函数的选择问题 [J]. *中南林业调查规划*, 1996, 15(1): 54-55