

## 基于 Rapideye 影像的林木地上生物量估测

吴平, 黄庆丰\*, 唐雪海, 陆宁辛

(安徽农业大学林学与园林学院, 合肥 230036)

**摘要:** 以石台县为研究地, 结合 Rapideye 高分遥感影像和不同森林类型样地林木地上生物量调查数据, 采用 Pearson 双变量相关分析方法筛选模型变量, 分别用多元线性回归和随机森林算法建立不同森林类型的遥感地上生物量估测模型, 并进行模型估测精度对比分析。结果表明, 叶绿素红边模型 (CRM) 与叶绿素绿波模型 (CGM) 2 个指数与针叶林、阔叶林生物量在 0.01 水平上的相关性极显著, 且在其多元线性回归模型和随机森林模型中两者均被挑选为建模变量。另外, 与生物量相关性较强的纹理特征主要集中的红光波段和红边波段, 且仅 MEAN、VAR、SM3 个滤波对生物量估测贡献较大, 可作为建模变量。阔叶林、针叶林和针阔混交林 3 种森林类型的地上生物量模型估测精度均表现为随机森林模型优于多元线性回归模型。随机森林模型生物估测绝对均方误差在 12.8760~36.5363 之间, 相对均方误差在 20.20%~45.95% 之间; 多元线性回归生物量估测绝对均方误差在 22.0425~46.4494 之间, 相对均方误差在 34.58%~58.42% 之间。

**关键词:** 林木地上生物量; Rapideye 影像; 随机森林; 逐步线性回归

中图分类号: S758.4

文献标识码: A

文章编号: 1672-352X (2016)06-0891-08

### Estimation of aboveground tree biomass based on Rapideye Imagery

WU Ping, HUANG Qingfeng, TANG Xuehai, LU Ningxin

(School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei 230036)

**Abstract:** Based on Pearson bivariate correlation analysis for model variables selection together with the analysis of data from the Rapideye image and sample plot aboveground biomass survey of different forest types in Shitai County, two remote sensing aboveground forest biomass estimation models of different forest types were built using the multiple linear regression and random forest algorithm method, respectively. The estimation precision of the two models was compared. The results showed that both CRM and CGM had significant correlation with the biomass of the coniferous forest and broad-leaved forest at the 0.01 level, indicating that CRM and CGM can be used for forest biomass estimation. In addition, the texture measurements which had strong correlation with biomass were mainly contained in the band of red or red edge and only MEAN, VAR and SM filter greatly contributed to biomass estimation, which can be selected to building model variables. The forest aboveground biomass estimation precision of broad-leaved forest, coniferous forest and coniferous and broad-leaved mixed forest using the random forest biomass model was better than the multiple linear regression model. The absolute mean square error of the random forest biomass estimation was between 12.8760 and 36.5363, while the relative mean square error was between 20.20% and 45.95%. The absolute mean square error of multiple linear regression biomass estimation was between 22.0425 and 46.4494, while the relative mean square error was between 34.58% and 58.42%.

**Key words:** forest biomass; Rapideye Image; vegetation indices; random forest; stepwise linear regression

森林生物量的研究起始于德国, 早期的研究主要依靠实地调查数据建立生物量相对生长方程的方法来估算主要森林植被类型、树种的生物量、碳储

量及其地理分布规律<sup>[1-3]</sup>。这种生物量研究方法的估测结果受实地调查数据的限制大, 且不能够及时全面的反映大面积区域的生态系统变化及生态环境状

收稿日期: 2016-03-02

基金项目: 中德合作 BMBF-Lin2Value-033L049C; Lin2Value-CAFYBB2012013 资助。

作者简介: 吴平, 硕士研究生。E-mail: wuping90828@163.com

\* 通信作者: 黄庆丰, 博士, 教授。E-mail: huangqf@ahau.edu.cn

况。遥感估测法因其快速、高效的特性使其成为获得大范围及不同尺度林分地上生物量的一种重要方法。不同遥感数据源的发展促使人们对如何充分利用其中所包含遥感信息的思考, 遥感参数的提出和选取成为遥感生物量估测的重要研究内容。目前用于森林生物量估算的遥感参数主要包括植被指数和纹理信息 2 种类型。植被指数主要包括基于波段的线性组合或比值植被指数、消除土壤背景植被指数以及针对高光谱遥感及热红外遥感而发展的植被指数 3 类<sup>[4]</sup>。基于二阶(共生矩阵)概率统计的滤波是纹理特征值计算比较常用的方法, 主要包括均值、方差、协同性、对比度、相异性等 8 个纹理测度<sup>[5-6]</sup>, 但大部分纹理测度与生物量的相关性并不强<sup>[7-8]</sup>, 只有方差和均值等少数纹理特征有利于提高生物量估算精度。波谱信息与纹理信息的结合应用可以使遥感图像的分类取得更好的成果<sup>[9]</sup>。而很多遥感数据算法都是以多元统计方法的原理和方法为基础的, 在遥感探测植被散热层和热辐射信息处理过程中, 往往存在着互相耦合的高度非线性化过程<sup>[10]</sup>, 这种复杂的非线性反演关系难以通过传统的线性回归的方法实现。动态的神经网络能够有效解决遥感反演过程中存在的复杂的非线性问题, 适用于大量高维遥感数据的参数反演过程。但其“黑箱”操作并不利于揭示森林生物量形成的内在机理。随机森林算法主要可用于分类和回归分析<sup>[11]</sup>。作为一种快捷有效的统计分析方法, 它不仅打破了多元线性回归的局限性, 还避免了神经网络回归估测无法解释生物量形成内在机理的弊端。大量理论与研究证明随机森林对异常值和噪声具有很好的容忍程度, 在数据处理过程中不易出现过拟合, 预测准确度高。

本研究以安徽省石台县部分区域为研究区, 按系统抽样方法设置圆形样地, 以样地林木生物量调查数据为基础, 结合 Rapideye 高分遥感影像纹理特征、波谱信息及与对森林生物量分布有影响的地形因子, 构建不同森林类型地上生物量遥感反演模型, 为该区域森林生物量与碳储量估算提供科学依据。

## 1 研究区概况

研究地为安徽省石台县, 地理坐标北纬 29°59′~30°24′, 东经 117°12′~117°59′, 其中不包括横渡镇和七都镇。海拔高度 50~1000 m; 中亚热带湿润气候区, 光照充足, 四季分明, 年平均气温为 16℃, 最高气温 38.8℃。年平均降水量为 1626.4 mm; 主要土壤类型有红壤、黄壤、黄棕壤等; 地带性植被为中亚热带北缘常绿、落叶阔叶混交林,

由于人为破坏等方面原因, 原始地带性植被已不复存在, 现存的多为人工林和天然次生林, 人工林主要树种有杉木 (*Cunninghamia lanceolata* (Lamb.) Hook.)、马尾松 (*Pinus massoniana* Lamb)、枫香 (*Liduidambar formosanna* Hance)、毛竹 (*Phyllostachys heterocyclus cv pubescens*) 等; 天然次生林主要树种有苦槠 (*Castanopsis sclerophylla* (Lindl.) Schott.)、甜槠 (*Castanopsis. eyrei* (Champ. ex Benth.) Tutch.)、青冈栎 (*Cyclobalanopsis glauca* (Thunb.) Oerst.)、茅栗 (*Castanea seguinii* Dode) 等。

## 2 研究方法

### 2.1 遥感数据获取和处理

以 2012 年 10 月的 Rapideye3A 遥感数据(UTM 投影, WGS84 坐标系, 表 1)为信息源, 结合 DEM 数据, 利用 ENVI、FEP(自主开发软件)及 QGIS 等软件进行消除阴影、影像增强等处理, 利用样地位置信息对应提取影像中各波段光谱信息, 并计算与生物量相关的植被指数、纹理信息等遥感参数, 用于地上生物量遥感反演模型建立。

遥感变量信息包括从遥感数据中直接提取的波谱如红(R)、红边(RE)和近红外(NIR)等波段灰度值以及依靠内部规律计算获得的植被指数如归一化植被指数(NDVI)、比值植被指数(SR)、增强植被指数(EVI)、红边归一化植被指数(NDVIRE)、叶绿素红边模型(CRM)、叶绿素绿光模型(CGM)、均值(Mean)、方差(Variance)、对比度(Contrast)、信息熵(Entropy)、二阶矩(Second moment)和相关性(Correlation)等。并从 DEM 数据中提取样地坡度、坡向作为变量, 参与模型的建立。

根据各纹理测度计算方法可知, 灰度共生矩阵提取纹理特征时, 提取的结果受不同的窗口大小的影响。对经过阴影消除处理后得到的图像分不同窗口大小 3×3, 5×5, 7×7, ..., 51×51, 滞后距离为 1, 纹理检测方向分别为 0°、45°、90°和 135°, 取 4 个方向的纹理均值进行纹理均值特征(Mean)提取。通过不同窗口提取的方差(Variance)值与样地生物量相关性分析得出窗口大小为 43×43 时提取的纹理特征值与生物量的相关性最强。因此使用 43×43 的窗口大小, 对其他用于建模的纹理特征值进行提取。

### 2.2 样圆的设计与调查

本研究通过设置不同半径大小圆形嵌套样地获取各径级单位面积生物量来折算林分地上总生物

量, 这不仅免去了传统方形样地在复杂地形中四角定向、定位的操作困难, 而且大大减小大样方内每木检尺的工作量。Zolkos 等在激光雷达的森林生物量估测精度与样点大小的相关性分析中得出在一定范围内, 样地大小与模型估测误差成负相关<sup>[12]</sup>。

Fassnacht 等在样地大小、数据类型以及预测方法在遥感森林生物量估测中重要性研究中也表示, 样地大小对遥感森林生物量估测精度影响不大, 而遥感估测精度却随着样地数量的增加而提高<sup>[13]</sup>。

表 1 Rapideye 影像的数据参数  
Table 1 Parameters of Rapideye Imagery

卫星 Satellite	数据幅宽 The data width	波段号 No. of band	波段名称 Name of the band	光谱范围/nm Spectral range	空间分辨率/ m Spatial resolution
Rapideye	77 km	B1	蓝	440~510	5
		B2	绿	520~590	5
		B3	红	630~685	5
		B4	红边	690~730	5
		B5	近红外	760~850	5

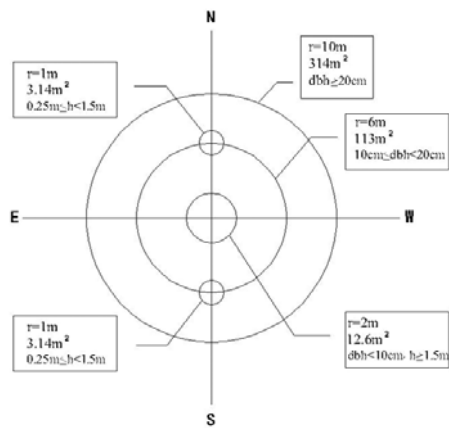


图 1 环形嵌套样圆设置

Figure 1 Configuration of the nested circle sample plots

按照 3 km×3 km、500 m×500 m、100 m×100 m 网格系统布置样点, 共有 227 个样点 (表 2),

剔除落在非林地区的调查样点和毛竹林样点后, 分不同森林类型共计 145 个样点, 其中: 阔叶林 56 个, 针叶林 61 个, 针阔混交林 28 个。

以样点为圆心  $r=6\text{ m}$ ,  $10\text{ cm} \leq dbh < 20\text{ cm}$ ;  $r=10$

$\text{m}$ ,  $dbh \geq 20\text{ cm}$  (如图 1) 设置圆形样地, 记录样地内每株树木的名称、树高、胸径、到样地中心的水平距离。同时在圆心处设置  $r=2\text{ m}$ ,  $dbh < 10\text{ cm}$ ,  $h > 1.5\text{ m}$  圆形样地, 并在其正南正北方向各设置一个  $r=1\text{ m}$ ,  $0.25\text{ m} < h < 1.5\text{ m}$  圆按形样地来调查林下植被更新情况。

### 2.3 样地生物量估算

样地地上部分生物量等于各单木地上生物量之和, 各单木地上部分生物量估算利用中德 Lin4carbon 项目成果<sup>[14]</sup>, 详见表 3, 表中没有列出的树种, 均参照其相近树种生物量模型计算生物量。由于 6~10 m 圆环内只调查胸径  $\geq 20\text{ cm}$  的林木, 圆环内  $10\text{ cm} \leq dbh < 20\text{ cm}$  的林木生物量等于半径  $r=6\text{ m}$  的样圆单位面积生物量乘以圆环面积。内圆与外环生物量合计即样圆地上部分生物量。本文是中德双方伙伴共同承担的 Lin<sup>2</sup> Value 项目的部分内容, 根据项目实施方案样地林木起测直径为 10 cm, 对小于 10 cm 的林木只进行了计数, 并未纳入样地总生物量计算。

表 2 不同尺度生物量调查样地的设置  
Table 2 Different scales of inventory studies

调查类型 Investigation type	调查范围 Field of investigation	抽样网格跨度 Based grid	样地个数 Plot number
I	整个研究区域	3 km×3 km	67
II	森林经营示范区域	500 m×500 m	60
III	森林经营示范小班	100m×100m	100
合计			227

注: 森林经营示范区为石台县德援森林经营示范区域; 森林经营示范小班为石台县德援森林经营示范小班。

Note: Demonstration area of Forest Management: Sustainable forest management (SFM) demonstration area in Shitai County; Forest Management demonstration sub-compartment: SFM sub-compartment in Shitai County.

表 3 主要树种林木地上生物量生长方程  
Table 3 Allometric biomass equations of the main species

树种 Tree species	生物量生长方程 Allometric biomass equation	$R^2$	RMSE
甜槠 <i>Castanopsis eyrei</i>	$AGB = \exp(-0.9166 + 1.9750 \ln(dbh)) \times 1.0042$	0.9207	0.0293
苦槠 <i>castanopsis sclerophylla</i>	$AGB = \exp(-2.8958 + 0.9427 \ln(dbh^2h)) \times 1.0397$	0.9753	0.0224
杉木 Chinese fir	$AGB = \exp(-2.6751 + 2.4668 \ln(dbh)) \times 1.0093$	0.9768	0.0169
青冈栎 <i>Cyclobalanopsis glauca</i>	$AGB = \exp(-2.2996 + 0.9065 \ln(dbh^2h)) \times 1.0545$	0.9774	0.0302
枫香 <i>Liquidambar formosana</i> Hance	$AGB = \exp(-1.0902 + 2.0568 \ln(dbh)) \times 1.0454$	0.9519	0.0393

## 2.4 建模构建与验证

采用 R3.1.1 软件和逐步回归的方法, 无放回随机抽样的方法将各森林类型样地按 7:3 的比例分为建模样本和验证样本及随机森林算法, 构建不同森林类型林木地上生物量与遥感变量的多元回归模型及随机森林模型, 并进行模型检验<sup>[15-16]</sup>, 各检验指标见表 4。

表 4 模型预测精度评价方法

Table 4 Measures for precision evaluation of predictions

评价指标 Evaluation index	计算公式 Calculation formula
绝对偏差 BIAS	$BLAS = \frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)$
相对偏差 BIAS <sub>r</sub>	$BLAS_r = \frac{BLAS}{\bar{x}} \times 100\%$
绝对均方根误差 RMSE	$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}}$
相对均方根误差 RMSE <sub>r</sub>	$RMSE_r = \frac{RMSE}{\bar{x}} \times 100\%$

注: 上式中  $n$  为验证样本数量,  $x_i$  为生物量地面实测值,  $\hat{x}_i$  为模型估测值,  $\bar{x}$  为实测值的平均数。

Note:  $n$  is the number of verification samples,  $x_i$  is the measured data of above-ground biomass,  $\hat{x}_i$  is the value of model estimation,  $\bar{x}$  is the mean value of measured data.

## 3 结果与分析

### 3.1 变量筛选

在挑选建模自变量时, 不仅要考虑自变量与相关变量间的相关程度, 还要避免同时选择两个相关性过大的变量, 以免较大的信息重叠造成模型的多重共线性, 以致影响模型预估精度。利用 R3.1.1 软件中的相关性分析功能, 分别对阔叶林、针叶林和针阔混交林的样地原始波段平均值、遥感参数平均值、海拔、坡度共 36 个变量和样地平均生物量进行 Pearson 双变量相关分析。由于逐步回归按照偏相关系数的大小次序引进方程, 最终得到与因变量最相关的因素, 在此过程中无法直接处理共线性问题。为避免多重共线性问题对逐步线性回归结果的影响,

本研究通过预先检验 (Pearson 自相关函数), 找到相关性较强的变量, 剔除其中与生物量相关性较小的变量。部分变量间虽然存在较高的相关性, 但考虑到其物理意义的不可替代性, 仍予以保留。

综合考虑以上因素, 最终选择用于建立 3 种不同森林类型反演模型的遥感变量分别为: 阔叶林 (16 个) CGM、CRM、NDVIRE、R、RE、SR、MEAN1、CORR1、VAR1、EN1、EN3、SM3、MEAN4、SM4、海拔、坡度; 针叶林 (12 个) R、CGM、CRM、NDVI、SR、MEAN3、EN1、VAR4、SM3、CORR2、海拔、坡度; 针阔混交林 (12 个) CGM、CRM、R、NDVI、SR、MEAN1、VAR3、EN3、SM3、CORR3、海拔和坡度。

### 3.2 回归模型

**3.2.1 逐步线性回归模型** 根据上述变量初步筛选的结果, 在 R3.1.1 平台上采用逐步回归的方法 (变量显著水平  $P \leq 0.05$  时选入, 变量显著水平  $P \geq 0.1$  时剔除) 选择建立模型的最优变量组合并计算各变量系数, 以建立 3 种不同森林类型地上生物量进行模型, 结果如下:

阔叶林:  $W = 586.69 - 234.03CGM + 100.59SR + 227.38VAR1 - 72.25MEAN4 + 521.99SM4$

针叶林:  $W = 283.474 + 168.193CGM - 43.809SR - 45.862MEAN3 + 42.842EN1 - 46.988SM3 - 26.265VAR4$

针阔混交林:  $W = 467.73 + 75.94VAR3 - 205.68EN3 - 415.00SM3$

上述公式中  $W$  代表乔木地上生物量, 单位为  $t \cdot hm^{-2}$ 。

经回归诊断, 不同森林类型地上生物量遥感估测模型符合线性假设、残差正态性、同方差性和因变量独立性, 且不存在多重共线性问题。各森林地上生物量回归结果表明 (表 6) 拟合效果较好。

**3.2.2 随机森林预估模型** 由于建模数据分布在不同的数量级, 为此, 在回归前应进行数据归一化处理, 利用 R3.1.1 中 party 功能包, 计算随机森林中各特征变量的重要性, 根据变量个数和袋外误差图剔除不重要的特征变量, 再利用新的特征集重复上

表 5 地上生物量观测值与遥感变量间 Pearson 相关系数  
Table 5 Pearson correlation coefficients  $r$  of observed AGB and satellite image variables

变量 Variable	阔叶林		针叶林		针阔混交林		
	相关性 $r$ Correlation	显著性 $P$ Significant	相关性 $r$ Correlation	显著性 $P$ Significant	相关性 $r$ Correlation	显著性 $P$ Significant	
原始波 —	$R$	-0.4050**	0.0020	-0.4800**	0.0001	-0.3262	0.0903
	$RE$	0.3300*	0.0130	0.0629	0.6299	0.0589	0.7658
	$NIR$	-0.0074	0.9566	0.0693	0.5957	0.0379	0.8482
植被指 —	$CGM$	-0.3510**	0.0079	0.0091	0.9445	-0.0124	0.9500
	$CRM$	0.2186	0.1056	0.3670**	0.0036	0.1922	0.3271
	$NDVIREE$	-0.3320*	0.0124	0.0316	0.8090	-0.0006	0.9975
	$EVI$	0.0006	0.9966	0.0551	0.6735	0.1219	0.5368
	$NDVI$	0.2469	0.0666	0.3090*	0.0154	0.2530	0.1939
	$SR$	0.2660*	0.0477	0.3570**	0.0047	0.2430	0.2128
	纹理信 —	$MEAN1$	-0.6900**	0.0000	-0.7960**	0.0000	-0.5440**
$EN1$		-0.2860*	0.0324	-0.2337	0.0698	-0.2750	0.1566
$SM1$		0.1664	0.2203	0.2138	0.0981	0.1820	0.3539
$CORR1$		0.3480**	0.0087	0.0951	0.4661	0.3880*	0.0413
$VAR1$		-0.3460**	0.0089	-0.2590*	0.0442	-0.3830*	0.0441
$MEAN2$		-0.5390**	0.0000	-0.6380**	0.0000	-0.3980*	0.0359
$VAR2$		-0.3170*	0.0173	-0.1898	0.1429	-0.3441	0.0729
$EN2$		-0.2353	0.0809	-0.1746	0.1783	-0.3405	0.0762
$SM2$		0.1384	0.3091	0.0765	0.5577	0.3151	0.1024
$CORR2$		0.2820*	0.0355	0.4590**	0.0002	0.3116	0.1065
$MEAN3$		-0.5070**	0.0001	-0.5870**	0.0000	-0.5010**	0.0066
$VAR3$		-0.3030*	0.0230	-0.1576	0.2250	-0.3980*	0.0360
$EN3$		-0.3200*	0.0162	-0.1092	0.4021	-0.4350*	0.0207
$SM3$		0.3200*	0.0163	0.0396	0.7617	0.3970*	0.0363
$CORR3$		0.0401	0.7694	0.0972	0.4563	0.4450*	0.0177
$MEAN4$		0.3490**	0.0085	-0.0225	0.8635	-0.1117	0.5714
$VAR4$		0.0858	0.5294	0.1333	0.3057	-0.0977	0.6210
$EN4$		0.2429	0.0713	0.1586	0.2223	-0.0602	0.7610
$SM4$		-0.3900**	0.0030	-0.1795	0.1662	0.0261	0.8950
$CORR4$		-0.0562	0.6807	-0.1834	0.1571	0.2391	0.2205
$MEAN5$	0.3540**	0.0075	0.1094	0.4012	0.2746	0.1573	
$VAR5$	-0.0846	0.5355	-0.0272	0.8354	0.2247	0.2503	
$EN5$	-0.0725	0.5953	-0.0850	0.5151	-0.0003	0.9988	
$SM5$	0.0675	0.6209	0.0268	0.8376	0.0553	0.7800	
$CORR5$	0.2254	0.0949	-0.0285	0.8275	0.0831	0.6742	
地形因 —	海拔 Elevation	-0.2229	0.0987	-0.0608	0.6414	-0.2120	0.2789
	坡度 Gradient	0.0688	0.6141	-0.0611	0.6402	-0.2450	0.2090

注：“\*”表示在 0.05 水平上（双侧）显著相关，“\*\*”表示在 0.01 水平上（双侧）显著相关；R.红光波段；RE.红边波段；NIR.近红外波段；CGM.叶绿素绿光模型；CRM.叶绿素红边模型；NDVIRE.红边归一化植被指数；EVI.增强植被指数；NDVI.归一化植被指数；SR.比值植被指数；MEAN1~5.各波段均值纹理滤波；EN1~5.各波段信息熵纹理滤波；SM1~5.各波段二阶矩纹理滤波；CORR1~5.各波段相关性纹理滤波；VAR1~5.各波段方差纹理滤波。

Note: “\*” stands for (double side) significantly correlated at the 0.05 level, “\*\*” express a (double side) significantly correlated at the 0.01 level; R. Red Band; RE. Red Edge Band; NIR. Near-infrared Band; CGM. Chlorophyll Green Model; CRM. Chlorophyll Red Edge Model; NDVIRE. Red Edge Normalized Difference Vegetation Index; EVI. Enhanced Vegetation Index; NDVI. Normalized Difference Vegetation Index; SR. Simple Ratio Index; MEAN1~5. Mean filtered texture of bands1~5; EN1~5. Entropy filtered texture of bands1~5; SM1~5. Second Moment filtered texture of bands1~5; CORR1~5. Correlation filtered texture of bands1~5; VAR1~5. Variance filtered texture of bands1~5.

表 6 多元回归模型参数统计

Table 6 Statistics of multiple regression model parameters

森林类型 Forest type	决定系数 $R^2$	调整决定系数 $AdjR^2$	标准残差 $RSE$	$F$	$P$
阔叶林 Broad-leaved forest	0.7082	0.6626	37.59	15.54	<0.001
针叶林 Coniferous forest	0.6360	0.5914	21.41	14.27	<0.001
针阔混交林 Theropencedrymion	0.7485	0.6982	19.32	14.88	<0.001

表 7 随机森林模型参数统计

Table 7 Statistics of RF model parameters

森林类型 Forest type	变量 Variable	残差平方和 Residual sum of squares	变量解释率/% Var. explained
阔叶林 Broad-leaved forest	CGM,NDVIRE,RE,MEAN1,VARI1, CORR1,EN3,MEAN4,海拔	3056.283	63.24
针叶林 Coniferous forest	CGM,NDVIRE,RE,CORR3,MEAN4,CORR4	781.3387	67.16
针阔混交林 Theropencedrymion	RE,SM1,VARI1,EN1,SM2, VAR2, MEAN4,CORR4	954.6212	68.57

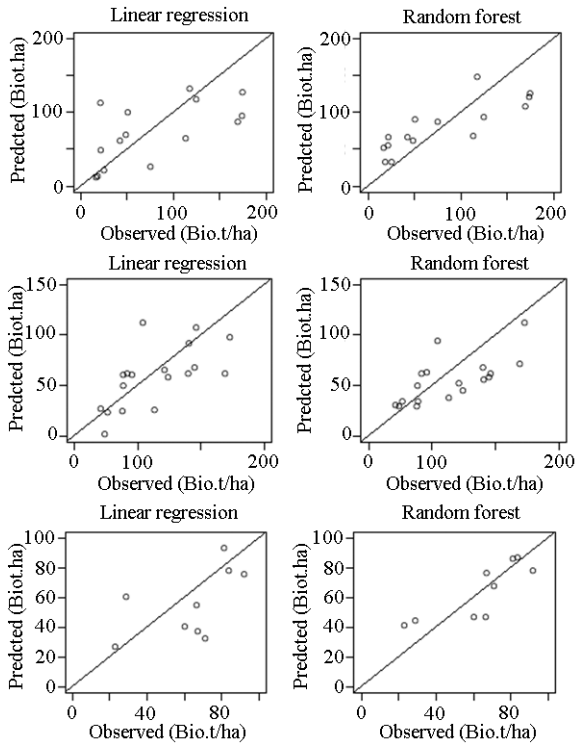


图 4 不同森林类型预测值与实测值(上左:阔叶林,上右:针叶林,下:针阔混交林)

Figure 4 Predicted versus observed for different forest types (upper left, broad-leaved forest; upper right, coniferous forest; below, coniferous and broad-laved mixed forest)

述过程,直到剩下的特征值子集建立的模型袋外误差率最低为止。根据上述方法最终选取的随机森林特征向量集,使用 Random Forest 包建立阔叶林、针叶林和针阔混交林回归模型,在此过程决策树的数量(ntree)和回归节点分裂的特征向量个数(mtry)

的设置很关键,ntree 回归树的数量可以通过回归误差的大小来确定,而 mtry 往往为输入变量的 1/3,当输入变量小于 3 时,设为 1<sup>[17]</sup>。

由表 7 可知,3 种森林类型生物量的随机森林模型参量中均有红边及红边归一化植被指数被选中,因此,遥感影像的红边信息能够很好的反映出植被生物量的情况。而在多元线性回归中,红边信息对生物量估测的作用并没有很好的体现,这可能与红边波段受植被覆盖度大小和叶绿素含量变化的影响来间接反映生物量变化有关<sup>[18]</sup>。

与多元线性回归相同,叶绿素模型与纹理信息均对生物量的变化表现出良好的反映,但随机森林模型选择的变量保留了物理意义不可替代的所有变量,而不仅仅通过与生物量的相关性大小来筛选变量,因此,随机森林模型较多元线性回归模型对生物量形成内在机理的解释更加全面。

**3.2.3 不同建模方法估测精度比较** 利用每个森林类型预留的验证样本数据,分别对以上两种生物量模型的预估精度进行评价,结果见图 4 和表 8。

由图 4 可以看出,3 种森林类型随机森林预测的结果值围绕斜线波动的幅度较多元线性回归的结果小,且变化平稳。

由表 8 可知,阔叶林和针阔混交林随机森林模型较多元线性回归模型的估测结果的整体绝对偏差和相对偏差小;阔叶林整体估测结果偏高,而针阔混交林整体估测结果偏低;针叶林的随机森林绝对偏差和相对偏差较多元线性模型大。3 种森林类型的随机森林模型生物量估测的绝对均方根误差和相对均方根误差均低于多元线性回归模型。阔叶林的随机森林模型估测结果的绝对均方误差和相

对均方误差分别为 36.5363 和 45.95%; 针叶林的随机森林模型估测结果的绝对均方误差和相对均方误差分别为 24.0510 和 37.64%; 针阔混交林的随机

森林模型估测结果的绝对均方误差和相对均方误差分别为 12.8760 和 20.20%。

表 8 不同建模方法的模型预测精度结果  
Table 8 Prediction precision of different algorithms

森林类型 Forest type	模型 Model	绝对偏差 Absolute deviation	相对偏差/% Relative deviation	绝对均方根 Absolute root mean square	相对均方根/% Relative root mean square
阔叶林 Broad-leaved forest	多元线性回归 MLR	-7.1061	-8.94	46.4494	58.42
针叶林 Coniferous forest	随机森林 Random forest	1.0539	1.33	36.5363	45.95
针阔混交林 Theropencedrymion	多元线性回归 MLR	-5.2062	-8.15	26.5481	41.54
	随机森林 Random forest	-8.0209	-12.55	24.0510	37.64
	多元线性回归 MLR	-8.4196	-13.21	22.0425	34.58
	随机森林 Random forest	-0.0537	-0.08	12.8760	20.20

从模型的拟合效果和预估精度来看, 随机森林模型估测林分地上生物量的结果优于多元线性回归模型, 能够更加全面解释生物量形成的内在机理, 较准确的反映生物量的真实情况。

## 4 讨论与结论

### 4.1 讨论

本研究采用 Pearson 相关分析法从大量遥感参数中选取具有较高代表性的变量子集进行模型拟合, 而遥感图像特征与生物量之间存在着多维复杂的非线性关系, 通过分析遥感变量与生物量之间的线性相关程度来进行变量筛选的方法是否会造成有用变量被剔除仍需进一步的研究论证。随着数据挖掘方法的研究进展, 越来越多的模型拟合变量并不受多重共线问题的限制, 变量选择机制也在不断进步, 引用新的变量筛选方法也是未来模型拟合的重要研究方向。

通过多元线性回归和随机森林算法(非线性方法)两种建模方法效果的对比研究, 最终随机森林算法的回归估测效果更好, 但随机森林算法与其他建模方法(多元自适应回归、遗传算法、联立方程法等)的预测效果对比仍有待探索。

### 4.2 结论

本研究基于森林生物量形成的内在机理, 引入叶绿素红边模型(CRM)及叶绿素绿波模型(CGM), 结合植被指数、纹理信息和 DEM 数据, 与林木地上生物量进行关联分析, 但由于 DEM 数据精度不足以及 Rapideye 数据本身的局限, 本研究无法对林分的垂直结构生物量进行区分。最终研究结果表明 CRM、CGM 与针叶林、阔叶林生物量在 0.01 水平上的相关性极显著, 且在其多元回归模型及随机森

林模型中 CGM 均被挑选为建模的变量, 说明 CRM 和 CGM 可以用于森林生物量遥感估测。此外, 与生物量相关性较强的纹理特征主要集中的红光波段和红边波段, 且仅 MEAN、VAR、SM3 个滤波对生物量估测贡献较大, 其他滤波并未与林分地上生物量表现出明显的相关性。

多元线性回归与随机森林算法回归结果对比表明, 3 种森林类型的回归拟合结果均表现为随机森林模型的精度优于多元线性回归模型: 阔叶林绝对均方误差 (RMSE) 为 36.5363, 相对均方误差 (RMSEr) 为 45.95%; 针叶林 RMSE 为 24.0510, RMSEr 为 37.61%; 针阔混交林 RMSE 为 12.8760, RMSEr 20.20%。随机森林模型的变量的解释能力也优于多元线性回归模型: 阔叶林 63.24%, 针叶林 67.16%, 针阔混交林 68.57%。

致谢: 本文是中德双方伙伴共同承担的 Lin<sup>2</sup>Value 项目的部分内容, 在研究方案制定、外业调查与数据分析过程中得到了德国哥廷根大学 Chripston Klein 教授及其团队的支持和帮助, 在此一并致谢。

### 参考文献:

- [1] 郭娜, 刘剑秋. 植物生物量研究概述(综述)[J]. 亚热带植物科学, 2011, 40(2): 83-88.
- [2] HOUGHTON R A, SKOLE D L, NOBRE C A, et al. Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon[J]. Nature, 2000, 403(6767): 301-304.
- [3] FANG J, CHEN A, PENG C, et al. Changes in forest biomass carbon storage in China between 1949 and 1998[J]. Science, 2001, 292(5525): 2320-2322.
- [4] 范文义, 张海玉, 于颖, 等. 三种森林生物量估测模型的比较分析[J]. 植物生态学报, 2011, 35(4): 402-410.

- [5] 岳彩荣. 香格里拉县森林生物量遥感估测研究[D]. 北京: 北京林业大学, 2012.
- [6] 曾文华. 基于灰度共生法和小波变换的遥感影像纹理信息提取[D]. 长春: 东北师范大学, 2006.
- [7] 黄丽梅. 纹理信息在遥感影像分类中的应用[D]. 济南: 山东师范大学, 2009.
- [8] 王昆, 张晓丽, 王珊, 等. 鹭峰地区 QuickBird 影像纹理特征与生物量估测关系初探[J]. 地理与地理信息科学, 2013, 29(3): 52-55.
- [9] 周成虎, 骆剑承, 杨晓梅, 等. 遥感影像地学理解与分析[M]. 北京: 科学出版社, 1999.
- [10] 张锋. 基于遥感信息估测森林生物量的研究[D]. 哈尔滨: 东北林业大学, 2003.
- [11] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
- [12] ZOLKOS S G, GOETZ S J, DUBAYAH R. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing[J]. Remote Sens Environ, 2013, 128: 289-298.
- [13] FASSNACHT F E, HARTIG F, LATIFI H, et al. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass[J]. Remote Sens Environ, 2014, 154: 102-114.
- [14] RUBÉN GUIASOLA RODRÍGUEZ. Allometric biomass equations and crown architecture in mixed-species forests of subtropical China [D]. Freiburg: Albert-Ludwigs Universität, 2014.
- [15] 张敏, 袁辉. 关于标准误差应用问题的讨论[J]. 郑州工业大学学报, 1996, 17(3): 95-100.
- [16] FUCHS H, MAGDON P, KLEINN C, et al. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory[J]. Remote Sens Environ, 2009, 113(3): 518-531.
- [17] LIAW A, WIENER M. Classification and regression by random forest[J]. R News, 2002, 2(3): 18-22.
- [18] 孙雪莲, 舒清态, 欧光龙, 等. 基于随机森林回归模型的思茅松人工林生物量遥感估测[J]. 林业资源管理, 2015(1): 71-76.