

# 基于 SPOT5 影像的杉木胸高断面积估测探讨

陈柏海 林辉 孙华

(中南林业科技大学 林业遥感信息工程研究中心,长沙 410004)

**摘要:** 采用角规实地调查黄丰桥林场 90 个杉木人工纯林样地胸高断面积,利用样地 SPOT5 遥感信息与地理信息,建立了杉木胸高断面积多元线性回归估测模型。首先对样地采用 GIS 软件进行缓冲处理,缓冲后每个样地的面积为 1 hm<sup>2</sup>;然后提取样地遥感光谱信息与纹理信息等 21 个因子和 4 个 GIS 因子,采用逐步回归分析法筛选出 6 个因子作为模型自变量;最后分别采用普通最小二乘法(OLS)和偏最小二乘法(PLS)建立了杉木胸高断面积多元回归模型。研究结果表明:OLS 回归模型的预测精度为 82.2%,均方根误差(RMSE)为 5.12 m<sup>2</sup>/hm<sup>2</sup>;PLS 回归模型的预测精度为 83.9%,均方根误差(RMSE)为 4.21 m<sup>2</sup>/hm<sup>2</sup>,PLS 和 OLS 回归模型在杉木胸高断面积估测中均取得了较好的效果,用中高分辨率遥感影像在估测森林结构参数上是可行的。

**关键词:** 胸高断面积;多光谱;SPOT5;多元统计分析

中图分类号:TP 79;S 757.2 文献标识码:A 文章编号:1003-6075(2012)01-0044-05

## Study on Basal Area Estimation of Chinese Fir Based on SPOT5 Images

CHEN Baihai ,LIN Hui ,SUN Hua

(Research Center of Forestry Remote Sensing & Information Engineering , Central South University of Forestry & Technology , Changsha 410004 , Hunan , China)

**Abstract:** The surveying of Chinese fir basal area by fielding 90 sample plots with angle gauges had been carried out in Huangfengqiao forest farm ,the multiple linear regression estimation model of basal area was set up based on remote sensing and geographic information. First ,each sample plot was buffered by GIS software to 1 hectare; Then from which 21 RS index factors such as spectral and texture information and 4 GIS index factors were extracted ,in which 6 index factors were screened out as independent model variables through stepwise regression analysis; Last the multiple regression model was built by using OLS and PLS respectively. The results showed that: the model predicted accuracy was 82.2% and RMSE was 5.12 m<sup>2</sup>/hm<sup>2</sup> by using OLS; the model predicted accuracy was 83.9% and RMSE was 4.21 m<sup>2</sup>/hm<sup>2</sup> by using PLS; The adoption of OLS and PLS services well in basal area estimation ,to estimate forest structural parameters can achieve good effects by using high resolution remote sensing images.

**Key words:** basal area; multispectral; SPOT5; multivariate statistical analysis

近年来,随着高空间分辨率遥感影像的广泛应用, GIS 和 GPS 技术在森林资源调查和管理中应用的普及,在给外业工作者带来极大方便的同时,也为借助 3S 技术对森林结构参数进行定量估测提供了所需的

必要条件<sup>[1]</sup>。国外学者对此做了大量的研究<sup>[2-6]</sup>。研究表明,利用中高分辨率遥感影像各波段的灰度值线性或非线性的组合,能较好地反映森林的生长状况和分布关系,与蓄积量、结构参数都有很好的相关关系,

\* 收稿日期:2011-11-21 修回日期:2011-12-13

基金项目:湖南省高等学校科学研究项目“高分辨率遥感影像森林结构参数反演研究”(11C1313)。

作者简介:陈柏海(1987-),男,湖南双牌人,硕士,主要研究方向为遥感与林业信息工程。

可作为定量因子,建立杉木胸高断面面积估测模型。

采用 SPOT5 影像作为遥感数据源,以研究区域 1:1 万地形图经矢量化生成的 DEM 图作为 GIS 数据源,对影像数据进行预处理,并提取相应的 RS 和 GIS 因子,结合地面实测数据,利用 SPSS 统计分析软件,用普通线性回归分析方法和偏最小二乘法,建立研究区的杉木胸高断面面积估测模型,为快速、高效、准确地估测杉木胸高断面面积提供理论依据。

## 1 材料与方法

### 1.1 数据获取与处理

#### 1.1.1 研究区概况

研究区位于株洲攸县黄丰桥国有林场内,总面积 351.16 km<sup>2</sup>。林区内森林多样性比较丰富,有木本植物 400 余种,主要以杉木、马尾松人工林为主,各龄组分布合理。

#### 1.1.2 SPOT5 数据

研究采用的数据拍摄时间为 2009 年 11 月 28 日 3 时 6 分 13 秒,它包含空间分辨率为 10 m 的 4 个多光谱波段,即绿波段(500~590 nm)、红波段(610~680 nm)、近红外波段(780~890 nm)与短波红外波段(1580~1750 nm)和分辨率为 2.5 m 的全色波段<sup>[7]</sup>。影像的拍摄太阳高度角和方位角分别为 45.11°和 158.74°。利用 ENVI 软件对 SPOT5 影像进行预处理,主要包括辐射定标、大气校正、几何校正等。

#### 1.1.3 外业调查与内业整理

外业调查时间为 2009 年 10 月至 2010 年 11 月,采用角规抽样的方法对林分进行调查。在研究区范围内,首先对影像进行初步解译和判读,对杉木人工纯林所在的位置和面积进行分析和计算,以确定样地布设的位置和数量要求。外业调查主要针对杉木中龄林和近熟林,并用带坡度改正的角规调查林分每公顷断面面积,用激光测距仪或皮尺测量水平距离来确定边界木;用 GPS 定位样地坐标。在本次外业调查和相应的数据处理过程中都采用 WGS-84 坐标系统,以保证样地的位置准确。

外业调查资料录入计算机统计整理,形成与样地相对应的每公顷杉木断面面积;然后样点在 SPOT5 影像上缓冲面积为 1 hm<sup>2</sup> 的样圆,以减少因 GPS 定位所带来的误差。由此,建立了地面调查点对应像元、像元灰度值和胸高断面面积之间的一一对应关系,为杉木胸高断面面积估测模型的建立提供依据。

### 1.2 变量设置

利用 SPOT 5 遥感影像,选取一系列可能与断面面积估测相关的 RS 因子及 DEM 图中提取的 GIS 因子,作为断面面积估测的自变量因子,以地面实测样点断面面积为因变量,建立多元回归模型。原始波段信息有多光谱波段 B1、B2、B3、B4(也就是 SWIR 短波红外波段)和全色波段 PAN 及相应波段的标准差;波段比值有可见光的 B1/B2(GRENRED)、B4/(B1+B2)(SVR)、B4/B3(MSI)、(B3-B2)/(B3+B2)(NDVI)、B3/B2(RVI);纹理信息主要是能量(ASM)、对比度(CON)、相关(COR)、熵(ENT)、逆差距(IDM)、方差(VAR);在 GIS 因子方面,设置的主要因子有海拔、坡度、坡向、阴影等。共设置 25 个自变量。

### 1.3 解算方法

#### 1.3.1 逐步回归分析

多元逐步回归法的基本思路在于,自动地从大量可供选择变量中选取最重要的变量,根据自变量对因变量作用程度的大小,保留作用程度大的变量,剔除作用小的变量<sup>[8]</sup>,变量的引入和剔除是同时进行的。循环反复计算,直到所有该引入的变量都被引入,该剔除的变量均被剔除为止,保留的变量组合即为最优变量组合。

#### 1.3.2 偏最小二乘法

偏最小二乘法是在建模过程中,集主成分分析、典型相关分析和线性回归分析的特点,实现回归建模、数据结构简化以及变量之间的相关性分析。因此,它在分析结果中,可以提供一个更为合理的回归模型<sup>[9]</sup>。一般情况下,偏最小二乘法回归分析法并不需要选取所有  $m$  个成分来建立回归方程,只需选用前  $h$  个成分,即可得到满意的回归模型。同时对结果进行交叉有效性检验( $Q^2$ ),当  $Q_h^2 \geq 0.0975$  时,可认为引进新成分  $t_h$  会对模型预测能力有明显改善作用。

#### 1.3.3 精度评价方法

确定系数( $R^2$ )、均方根差(RMSE)、相对误差(RE)是常用的对模型预测值和实测值之间进行评价和验证的方法,其定义如下<sup>[8]</sup>:

确定系数( $R^2$ )可以用来检验估计值与实测值之间的线性显著程度,变化范围在  $0 \leq R^2 \leq 1$ ,其中  $R^2$  越接近于 1,表明模型的预测能力越强,模型效果显著。

均方根误差(RMSE)表示实测值与估计值之间的离散程度,其值越小表示模型估计值与实测值之间的相对距离比较接近,即模型预测精度就越高。

相对误差( RE )评价是以相对误差绝对值的平均值作为模型精度评价指标,其值越小,模型的预测精度就越高。

计算公式:

$$\begin{cases} R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \\ RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \\ RE\% = \frac{(y_i - \hat{y}_i)}{y_i} \times 100 \end{cases}$$

式中:  $y_i$  为实测值,  $\hat{y}_i$  为模型估计值,  $\bar{y}_i$  为实测平均值,  $n$  为样本数。

## 2 结果与分析

### 2.1 自变量分析与筛选

自变量的选择是建立回归模型的关键。在建立的回归模型中,应尽可能的剔除那些对断面积估测没有重要解释作用的变量,从而达到简而精的要求。

从总体 90 个样本中,按断面积大小分段选取 70 个样本作为建模数据,其中包括最小断面积与最大断面积样本数据;剩余 20 个样本作为检验数据。计算各因子与断面积之间的相关系数及方差扩大因子(VIF),结果见表 1。

分析表 1 可知:与断面积相关系数较大前 3 个变量是 MSI、SVR、能量,说明这 3 个变量与断面积有较好的相关性;负相关性方面居前 3 个变量是 GRENRED、坡向、对比度,且这 3 个变量的绝对值都大于 0.1。在所设置的 25 个自变量中,方差扩大因子(VIF)大于 10 的有 10 个,表明它们之间存在多重相关性,若将这些因子全部纳入方程,估测模型的精度将无法保证。

在模型的筛选过程中,采用逐步回归分析方法,以回归系数显著性检验中各自变量的  $F$  统计量的临界概率值  $F_{Sig} \leq 0.05$  和  $F_{Sig} \geq 0.10$  作为自变量引入或剔除模型的判断标准,即:当  $F_{in} \geq F_{0.05}$ ,该变量被引入模型,否则不引入;如果  $F_{out} \leq F_{0.10}$ ,该变量被剔除,否则保留。共选取了 6 个自变量:MSI,能量,B3MEAN, B4S,熵,GRENRED。

### 2.2 最小二乘法分析

#### 2.2.1 基于最小二乘法回归模型的建立

利用逐步回归分析所筛选出的 6 个自变量,相应的断面积为因变量,用最小二乘法建立多元线性回归

表 1 遥感和 GIS 因子及其相关信息

变量序号	变量	相关系数	方差扩大因子
1	B1MEAN	-0.097	99.25
2	B2MEAN	0.023	6.281
3	B3MEAN	0.12	3.213
4	B4MEAN	0.21	5.458
5	B1S	0.091	96.13
6	B2S	0.12	89.21
7	B3S	0.059	14.04
8	B4S	-0.016	5.325
9	PANMEAN	-0.053	112.9
10	PS	0.17	29.25
11	GRENRED	-0.23	4.255
12	MSI	0.45	2.351
13	NDVI	0.22	20.58
14	RVI	0.24	8.264
15	方差	-0.044	54.37
16	逆差距	0.31	2.268
17	对比度	-0.13	53.39
18	熵	0.21	3.594
19	能量	0.28	5.112
20	相关	0.21	3.336
21	海拔	0.21	51.27
22	坡度	0.090	6.842
23	坡向	-0.12	3.546
24	阴影	0.10	9.108
25	SVR	0.42	1.297

模型。模型统计量见表 2。

表 2 模型统计量

相关系数	确定系数	校正确定系数	估计值标准误
0.736	0.542	0.496	5.091

在表 2 中,复相关系数  $R = 0.736$ ,确定系数  $R^2 = 0.542$ ,校正确定系数  $R_{adj}^2 = 0.496$ ,估计值标准误 = 5.091。

方差分析结果见表 3。

在表 3 中,可知模型回归平方和为 1958.51,剩余平方和为 1634.30,总平方和为 3592.80,检验回归方程的统计量  $F$  值 = 12.58,查  $F$  检验表可得临界值  $F_{0.05}(6, 63) = 2.25$ ,  $F > F_{0.05}(6, 63)$  表明模型在显著性水平  $\alpha = 0.05$  下通过回归方程检验,即所建回归模型有效。也可通过表中显著性值 = 0.0009 < 0.05 来

表 3 方差分析

项目	平方和	自由度	平均平方和	F 值	显著性
回归平方和	1 958.51	6	326.42	12.58	0.000 9
剩余平方和	1 634.30	63	25.95		
总平方和	3 592.80	69			

直接判断回归方程通过 F 检验(这里显著性表示弃真概率,含义下同)。

回归系数见表 4。

在表 4 中,各自变量的回归系数: B4S = -3.176, B3MEAN = 1.621, GRENRED = -0.2541, MSI = 19.25,

熵 = 9.214, 能量 = 32.90; 同时可知各变量所对应的 t 值及显著性值,在表中各自变量显著性值均小于 0.05,表明在显著性水平  $\alpha = 0.05$  下模型自变量各个回归系数通过 t 检验,具有显著性。

表 4 模型回归系数

变量	非标准回归系数		标准回归系数	t 值	显著性
	回归系数	回归系数标准误			
(常量)	-13.73	5.328		-1.635	0.11
B4S	-3.176	1.235	-0.240	-2.090	0.039
B3MEAN	1.621	0.1901	0.355	2.977	0.003 2
GRENRED	-0.254 1	0.065 23	-0.245	-2.522	0.012
MSI	19.25	4.495	0.428	3.450	0.009 4
熵	9.214	3.375	0.393	3.306	0.009 8
能量	32.90	10.99	0.523	4.814	0.007 6

综上所述,用筛选所得的 6 个自变量建立的拟合回归方程为:

$$y = -13.73 + 19.25 x_1 + 32.90 x_2 + 1.621 x_3 - 3.176 x_4 + 9.214 x_5 - 0.2541 x_6$$

式中:  $x_1$  表示 MSI,  $x_2$  表示能量,  $x_3$  表示 B3MEAN,  $x_4$  表示 B4S,  $x_5$  表示熵,  $x_6$  表示 GRENRED。

2.2.2 模型精度检验

利用 20 组检验数据对建立的断面积估测模型进行预测精度检验,检验结果见表 5。

表 5 模型精度检验结果

实测值与预测值 拟合方程	检验指标		
	确定系数	均方根误差	相对误差/%
$y = 1.059 x + 1.799$	0.673	5.12	17.8

以 20 组检验数据的预测值为横坐标,实测值为纵坐标,绘制散点图,如图 1 所示。

由表 5 和图 1 可知:以多光谱波段的绿红波段比值、水分胁迫指数、纹理信息中的能量、熵、全色波段的标准差为自变量,用最小二乘法建立的杉木胸高断面面积估测模型,它的确定系数值为 0.673 1,均方根误差

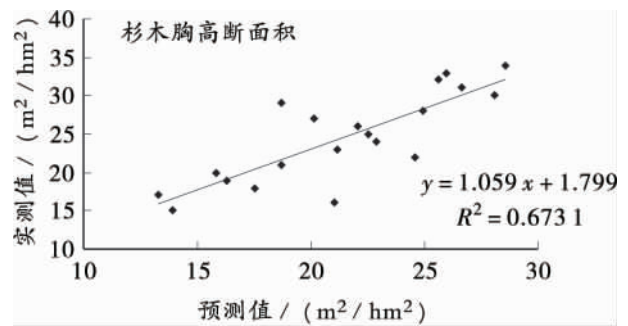


图 1 断面积估计值与实测值比较

值为 5.12  $m^2/hm^2$ , 预测精度达到 82.2%。

2.3 偏最小二乘法分析

2.3.1 基于偏最小二乘法回归模型的建立

利用逐步回归分析所筛选出的 6 个自变量,相应的断面积为因变量,用偏最小二乘法建立多元线性回归模型。

偏最小二乘法的前 5 个成分的统计信息见表 6。

由表 6 可知:校正确定系数  $R_{adj}^2$  随着成分的增加而增大,当成分个数为 3 时,校正确定系数  $R_{adj}^2$  达到最大,其值为 0.515,同时交叉检验因子  $Q_3^2 = 0.101 >$

0.097 5,表明只需提取自变量的前3个成分就可建立精度较高的回归模型。

表6 偏最小二乘法统计信息

成分个数	统计信息	
	校正确定系数	Q <sup>2</sup>
1	0.476	0.433
2	0.501	0.247
3	0.515	0.101
4	0.509	-0.086 3
5	0.503	-0.054 2

因此,设定自变量提取成分数为3,其回归系数和变量投影重要性指标见表7。

表7 回归系数与变量投影重要性表

自变量	回归系数	变量投影重要性指标		
		1	2	3
(常量)	-11.13			
B4S	-3.124	0.060 24	0.422 3	0.478 1
B3MEAN	1.512	0.648 2	0.641 0	0.687 2
GRENRED	-0.229 1	0.879 1	0.868 3	0.866 0
MSI	19.27	1.743	1.703	1.682
熵	9.424	0.762 3	0.751 1	0.829 4
能量	32.83	1.089	1.102	1.090

变量投影重要性指每一个自变量对因变量集合的解释能力,可通过变量投影重要指标值来测度。由表7可知:在所提取的3个成分中,按变量投影重要性指标值大小排序,可得6个自变量在估测模型中重要程度次序为:MSI > ASM > GRENRED > ENT > B3MEAN > B4S。同时,由此可得回归模型为:

$$y = -11.13 + 19.27 x_1 + 32.83 x_2 + 1.512 x_3 - 3.124 x_4 + 9.424 x_5 - 0.2291 x_6$$

式中:  $x_1$  表示 MSI  $x_2$  表示能量  $x_3$  表示 B3MEAN  $x_4$  表示 B4S  $x_5$  表示熵  $x_6$  表示 GRENRED。

### 2.3.2 模型精度检验

利用20组检验数据对建立的断面积估测模型的预测精度进行检验,检验结果见表8。

表8 模型精度检验结果表

实测值与预测值 拟合方程	检验指标		
	确定系数	均方根误差	相对误差/%
$y = 1.075 x - 1.177$	0.699	4.21	16.1

以20组检验数据的预测值为横坐标,实测值为纵坐标,绘制散点图如图2所示。

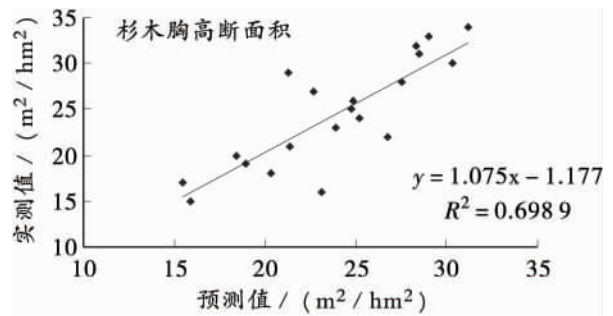


图2 断面积估计值与实测值比较

由表8和图2可知:以多光谱波段的绿红波段比值、水分胁迫指数、纹理信息中的能量、熵、全色波段的标准差为自变量,用偏最小二乘法建立的杉木胸高断面积估测多元线性回归模型,其确定系数0.698 9、均方根误差值为4.21 m<sup>2</sup>/hm<sup>2</sup>、预测精度达到83.9%。

## 3 结论与讨论

本研究是在较小范围内利用中高分辨率遥感影像对杉木胸高断面积进行估测分析,主要是探索利用中高分辨率在森林结构参数建模中的应用能力。

1) 利用SPOT5影像灰度值线性与非线性组合等25个自变量因子,经过逐步回归筛选出6个自变量,同时按断面积大小分段选取70个样本作为建模数据,其中最小断面积与最大断面积必取,建立多元线性回归模型;将剩余的20个样本数据代入估测模型进行精度评价。其中最小二乘法估测精度为82.2%、偏最小二乘法为估测精度为83.9%,达到了较为理想的预测效果。

2) 在建立回归模型时,通过两种不同的解算方法——最小二乘法(OLS)与偏最小二乘法(PLS),分别建立了两种杉木胸高断面积估测模型。比较可知,用PLS解算方法建立的估测模型比OLS建立的估测模型在估测精度上提高1.7%,每公顷断面积估测的均方根误差提高约1 m<sup>2</sup>/hm<sup>2</sup>,说明用PLS建立的回归模型更能接近实测值。

总体来说,用中高分辨率的遥感影像来估测杉木胸高断面积是可行的。同时,研究也存在着一定的不足:首先,因研究区域范围不大,实测的总体样本数据比较少,具有很强的地域性;其次,本研究对杉木胸高断面积估测模型的自变量的筛选,仅使用逐步回归分析来筛选自变量,没有考虑其它可能存在对估测模型影响较大的自变量因子。(下转第56页)

表2 西藏砂生槐种群分布格局

地点	平均数	方差	空间分布指数	分布格局
卡嘎镇	36.0	36.5	1.01	约随机分布
多颇章	44.8	179.3	4.00	集群分布
朗镇	31.0	19.7	0.64	均匀分布

注: 空间分布指数 = 方差/平均数; 空间分布指数 = 1 为随机分布型、<1 为均匀分布型、>1 为集群分布型。

多颇章区砂生槐呈集群分布, 成群、成簇、成块或成斑点地密集分布, 各群的大小、群间距离、群内个体密度等都不等。该区地处雅鲁藏布河江谷冲积扇, 地势凹凸不平, 凸地上层流沙被风吹走, 土壤表层干燥, 而凹地沉积了较多的流沙, 有利于土壤保墒, 也利于砂生槐种子的传播和保藏, 当土壤水分和温度在一定时段达到有效偶合时, 脱落在凹地的种子会大量萌发成苗, 从而使种群扩大, 密度增加, 形成集群分布。

朗镇区的砂生槐呈均匀分布。其原因主要是该区位于坡地, 地势较高, 流沙较少, 土壤干燥, 土壤质地为砾质沙土, 植株不受或轻微受到沙埋, 地面环境不利于砂生槐种子的保藏和土壤水分的保持, 当土壤温度和水分达到有效偶合时, 只有母株附近脱落的种子才能

萌发成苗。

## 4 结论

研究区西藏砂生槐天然种群分布在干旱荒漠地区, 常形成单优群落, 结构单一, 砂生槐为群落中的优势种。在砾质沙地交错带常呈随机分布, 在半固定沙丘、流动沙丘区常呈集群分布, 而在砾质坡地上分布较均匀。由于西藏砂生槐种皮坚硬、透水性差、发芽率低, 这一特性限制了天然种群的扩大和更新; 虫害和人畜活动等导致其天然种群不断衰减。因此, 建议采取封育、人工抚育等措施保护砂生槐资源, 充分发挥砂生槐防沙固沙的作用, 防治荒漠化, 促进西藏生态环境的改善。

参考文献:

- [1] 刘志明. 西藏日喀则固沙植物引种的比较研究[J]. 中国沙漠, 1996, 16(3): 326-330.
- [2] Greig-Smith P. Quantitative Plant Ecology[M]. Berkeley and Los Angeles University of California Press, 1983: 54-104.
- [3] Gitting R. Canonical analysis, a review with application in ecology [M]. Berlin: Springer Verlag, 1985.

(上接第48页)

参考文献:

- [1] 李崇贵, 赵宪文, 李春干, 等. 森林蓄积量遥感估测理论与实现 [M]. 北京: 科学出版社, 2006. 43-48.
- [2] W Roberts S, Tesfamichael M, Gebreslasie et al. Forest structural assessment using remote sensing technologies: an overview of the current state of the art [J]. Southern Hemisphere Forestry Journal, 2007, 69(3): 183-203.
- [3] Ingram JC, Dawson TP and Whittaker RJ. Mapping tropical forest structure in south-eastern Madagascar using remote sensing and artificial neural networks [J]. Remote Sensing of Environment, 2005 (94): 491-507.
- [4] Maselli F, Chirici G, Bottai L et al. Estimation of Mediterranean forest attributes by the application of k-NN procedures to multitemporal

Landsat ETM+ images [J]. International Journal of Remote Sensing, 2005, 26(17): 3781-3796.

- [5] Wolter P, P. Townsend, and B. Sturtevant. Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data [J]. Remote Sensing of Environment, 2009, 113: 2019-2036.
- [6] Song, C. Estimating tree crown size with spatial information of high resolution optical remotely sensed imagery [J]. International Journal of Remote Sensing, 2007, 28(15): 3305-3322.
- [7] 林辉, 孙华, 熊育久, 等. 林业遥感 [M]. 北京: 中国林业出版社, 2011. 64-69.
- [8] 袁志发, 周静芋. 多元统计分析 [M]. 北京: 科学出版社, 2006. 138-145.
- [9] 王惠文. 偏最小二乘回归方法及其应用 [M]. 北京: 国防工业出版社, 1999: 200-220.