

李圣娇,舒清态,徐云栋,等. 基于偏最小二乘回归模型的高山松蓄积量遥感估测[J]. 江苏农业科学,2015,43(8):182-185.  
doi:10.15889/j.issn.1002-1302.2015.08.060

# 基于偏最小二乘回归模型的高山松蓄积量遥感估测

李圣娇,舒清态,徐云栋,张 焱

(西南林业大学林学院,云南昆明 650224)

**摘要:**以 Landsat 8 影像为遥感数据源,以遥感因子、GIS 因子、林分因子、郁闭度等为自变量,在前期野外样地调查的基础上,采用偏最小二乘法(PLS),建立香格里拉县高山松蓄积量遥感估测模型。试验结果表明,郁闭度对香格里拉县高山松蓄积量估测的影响极其显著,第 5、6 波段对其影响较为显著;运用偏最小二乘法建立的样地蓄积量估测模型,调整决定系数  $R^2$  为 0.777 5,均方根误差  $RMSE$  为 36.90  $\text{m}^3/\text{hm}^2$ ,总预报偏差的相对误差  $RE$  为 23.18%,模型精度为 73.08%。以像元为单位提取高山松林所对应的自变量因子,利用估测模型得到研究区高山松林总蓄积量为 1 372.406 万  $\text{m}^3$ 。

**关键词:**高山松;蓄积量遥感估测;偏最小二乘法;香格里拉

**中图分类号:**S127 **文献标志码:**A **文章编号:**1002-1302(2015)08-0182-04

森林是地球上最大的生态系统,森林蓄积则是表征森林数量最重要的指标之一<sup>[1]</sup>。传统的森林蓄积量调查方法费时费力,且仅适用于小范围的研究。随着遥感技术的不断发展,基于遥感技术的森林蓄积量估测已成为国内外学者研究的热点,利用遥感影像数据结合少量的地面样地实测数据,建立以遥感因子、GIS 因子、林分立地条件为自变量因子的蓄积量估测模型已成为一种趋势<sup>[2-7]</sup>。张友静等将 K-T 变换得到的绿色、湿度、郁闭度作为自变量,构造出具有物理意义的森林蓄积量遥感估测模型,精度高达 90%<sup>[2]</sup>。据存勇等利用 TM 影像和 129 个实测样地进行了蓄积量估测模型的研究,结果表明泛化改进的 BP 神经网络比普通 BP 神经网络具有更高的预报精度<sup>[4]</sup>。近年来得以发展的偏最小二乘回归(PLS)方法也逐渐被应用于蓄积量估测领域的研究中<sup>[5-7]</sup>,而运用 PLS 对高海拔地区的高山松进行遥感蓄积量估测尚未见报

道。高山松林一般分布于云杉林、冷杉林下限,海拔 2 800 ~ 3 500 m 之间,林分外貌整齐,成片分布,以同龄单层林常见。高山松适应性广,更新能力强,是山地寒温带向山地亚热带过渡的喜光、耐旱、耐瘠薄的先锋树种。本研究采用 Landsat 8 影像作为遥感数据源,结合云南省香格里拉县 2006 年森林资源二类调查数据,采用偏最小二乘法建立森林蓄积量估测模型,研究结果可为低纬度、高海拔地区遥感地学的研究提供依据。

## 1 研究区概况

香格里拉县(99°20' ~ 100°19'E,26°52' ~ 28°52'N)位于云南省西北部、迪庆州东北部,地处云南亚热带常绿阔叶林植被区向青藏高原高寒植被区过渡地带,森林覆盖率为 74.99%,植被分布南北差异明显,在环县境的东、南、西 3 面山体垂直分布完整而典型。在垂直分布上,海拔 4 500 ~ 4 700 m 为雪线带,有高山草甸、灌丛植被生长;海拔 3 000 ~ 4 500 m 为亚高山、高山寒温性针叶林类型;海拔 3 000 m 以下为暖温性针叶林,其间有多种温凉性针叶树种、落叶树种与其组成各种复杂的森林类型;金沙江边则出现干暖河谷气候下形成的多种灌丛类型。香格里拉县主要的优势树种有云南松(*Pinus yunnanensis*)林、高山松林、云冷杉林、高山栎(*Quercus*

收稿日期:2015-03-18

基金项目:国家自然科学基金(编号:31460194,31060114)。

作者简介:李圣娇(1991—),女,河北邯郸人,硕士研究生,主要从事资源环境遥感的研究。E-mail:jmwzrh@sina.com。

通信作者:舒清态,博士,副教授,硕士生导师,主要从事 3S 技术及森林景观经营的教学与研究。E-mail:shuqt@163.com。

[7]彭晓鹏,赵克飞. 基于 GIS 的山地城镇空间拓展土地适宜性评价——以大埔县为例[J]. 热带地理,2013,33(4):480-488.

[8]类淑霞,郝晋珉,王丽敏. 生态脆弱区宜耕未利用土地开发适宜性评价——以山西省大同市为例[J]. 中国生态农业学报,2011,19(6):1417-1423.

[9]贾树海,白 静,邵日晶,等. 基于 GIS 的瓦房店市农用地适宜性评价研究[J]. 土壤通报,2009,40(5):987-992.

[10]任 玲,马 蓉,芦 帅,等. 基于 GIS 的棉花土地适宜性评价——以新疆兵团农八师 148 团 8 连为例[J]. 石河子大学学报:自然科学版,2012,30(2):193-197.

[11]郭月婷,廖和平,徐建刚. 三峡库区农村居民点用地适宜性评价[J]. 农业工程学报,2012,28(5):252-259.

[12]金 贵,王占岐,胡学东,等. 基于模糊证据权模型的青藏高原区土地适宜性评价[J]. 农业工程学报,2013,29(18):241-250.

[13]胡彩婷,李巧云,关 欣,等. 永兴县发展冰糖橙产业的土地适宜性评价[J]. 水土保持研究,2012,19(6):228-231,236.

[14]王路明,常庆瑞,白雪娇. 基于 GIS 的陕西省定边县耕地种植玉米适宜性评价[J]. 西北农林科技大学学报:自然科学版,2014,42(3):105-110,117.

[15]宋如华,齐 实,孙保平,等. 区域土地资源的适宜性评价和空间布局[J]. 土壤侵蚀与水土保持学报,1997,3(3):24-31.

[16]石淑芹,陈佑启,李正国,等. 基于空间插值分析的指标空间化及吉林省玉米种植区划研究[J]. 地理科学,2011,31(4):408-414.

*semicarpifolia*) 林, 占全县森林面积的 90.8%<sup>[8]</sup>。

## 2 研究内容与方法

### 2.1 数据获取与预处理

本研究采用的遥感数据为 2014 年 3—4 月的 Landsat 8 数据, 7 波段共 3 景, 航带号分别为 131-41、132-40、132-41, 空间分辨率为 30 m (图 1)。采用 ENVI 5.0 软件对其进行预处理, 包括大气校正、裁剪、拼接等。

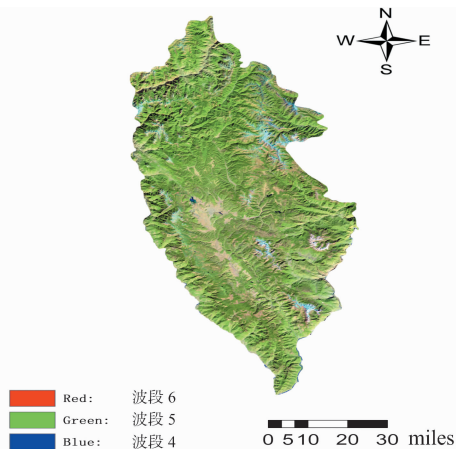


图1 香格里拉县 Landsat 8(6, 5, 4)遥感影像

本研究还参考了以下数据。2006 年森林资源二类调查小班数据, 已校正的香格里拉县 SPOT5 影像, 精度为 30 m 的 DEM 数据, 研究区行政边界矢量图。2014 年香格里拉县高山松蓄积量 30 m × 30 m 样地实测数据, 包括每个样地的 GPS 坐标 ( $X, Y$ )、高山松 30 m × 30 m 样地蓄积量、郁闭度、平均树高、平均胸径等样地因子。

### 2.2 研究方法

偏最小二乘法是一种新型的多元统计分析方法, 集多元线性回归分析、典型相关分析、主成分分析的功能和优点于一体, 将建模的数据分析与非模型的数据认知方法有机结合起来, 使模型的精度、稳健性、实用性得到提高, 已广泛应用于各个领域<sup>[9-11]</sup>。

设有  $q$  个因变量 ( $Y_1, Y_2, \dots, Y_q$ ) 和  $p$  个自变量 ( $X_1, X_2, \dots, X_p$ ), 观测了  $n$  个样本点, 由此构成自变量与因变量的数据表  $X_{n \times p}$  和  $Y_{n \times q}$ 。偏最小二乘回归分别在  $X, Y$  中提取出成分  $t_1, u_1$  ( $t_1, u_1$  分别是  $X_1, X_2, \dots, X_p$  和  $Y_1, Y_2, \dots, Y_q$  的线性组合), 提取时需满足 2 个条件:  $t_1, u_1$  应尽可能多地携带其各自数据表中的变异信息;  $t_1$  与  $u_1$  的相关程度达到最大。在第 1 个成分  $t_1$  与  $u_1$  被提取后, 偏最小二乘回归分别实施  $X, Y$  对  $t_1$  的回归。若回归方程达到满意的精确度则算法终止; 否则将利用  $X, Y$  分别被  $t_1$  解释后的残余信息进行第 2 轮成分提取, 如此往复, 直到获得满意的精确度为止。若最终对  $X$  提取了  $m$  个成分  $t_1, \dots, t_m$ , 偏最小二乘回归将实行  $Y_k$  对  $t_1, \dots, t_m$  的回归, 并表达为  $Y_k$  关于原变量  $X_1, X_2, \dots, X_p$  的回归方程 ( $k=1, 2, \dots, q$ )。

本研究采用标准差分析法剔除样本中离群值较大的数据, 在 R 环境下建立偏最小二乘法模型, 建模样地共 79 块, 按 3:1 原则, 用 60 个样本进行模型训练, 用 19 个样本进行精度检验。具体步骤为: 从预处理后的影像中提取遥感因子

和 GIS 因子; 对提取的因子数据进行标准化处理; 在 R 环境下建立偏最小二乘法模型; 模型训练及模型精度检验。

### 2.3 高山松空间分布信息及自变量因子的提取

2.3.1 高山松信息的提取 研究区高山松林中的天然林、人工林均基本以纯林出现。在 ENVI 5.0 软件下采用基于面向对象分类技术, 并结合 2006 年香格里拉县森林资源二类调查小班数据, 实现对高山松空间分布信息的提取 (图 2)。

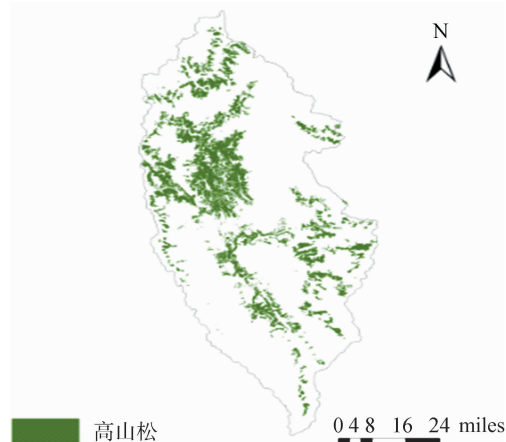


图2 2014年香格里拉县高山松分布情况

2.3.2 样地设置 样地设置需满足以下条件: 样地森林类型为纯林, 在一定范围内连续分布, 且存在蓄积; 遥感图像上样地的灰度值应均匀; 样地应尽量均匀分布, 覆盖整个研究区<sup>[6]</sup>。样地分布见图 3。

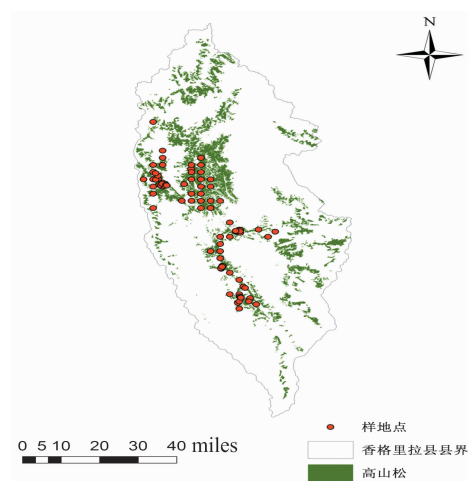


图3 香格里拉县高山松样地分布

2.3.3 自变量因子的提取 关于利用 3S 技术估测森林蓄积量, 国内外学者已作了大量研究<sup>[12-17]</sup>, 本研究在前人研究的基础上选择用于构建蓄积量估测模型的备选变量, 包括遥感因子变量、GIS 因子变量 (表 1)。

森林郁闭度通常需要野外实地测量, 且仅能获得部分点的数据, 不利于研究大范围或区域内郁闭度的空间分布及变化<sup>[18]</sup>。遥感技术的估测为区域范围内郁闭度的反演提供了新思路。本研究利用偏最小二乘法对整个香格里拉县高山松林的郁闭度以像元为单位进行估测及精度检验, 结果如下:

$$Y_{\text{郁闭度}} = -0.0013 \times \text{Dem} - 0.0058 \times B1 - 0.0045 \times B2 - 0.0024 \times B3 + 0.0011 \times B4 + 0.0015 \times B5 - 0.0015 \times B6 -$$

表 1 高山松蓄积量估测模型备选变量

类型	自变量
遥感因子 单波段	B1
	B2
	B3
	B4
	B5
	B6
	B7
植被指数	B5 - B4 (DVI 差值植被指数)
	B5/B4 (RVI 比值植被指数)
	(B5 - B4)/(B5 + B4) (NDVI 归一化植被指数)
比值波段	B5 × B6/B7
	B4/(B1 + B2 + B3 + B4 + B5 + B6 + B7)
GIS 因子	Dem 海拔
	Slope 坡度
	Aspect 坡向
其他因子	郁闭度

注: B4、B5、B6 分别为 Landsat 8 的红光、近红外、短波红外波段的灰度值;令  $B4/(B1 + B2 + B3 + B4 + B5 + B6 + B7)$  为  $X_1$ ,  $B5 \times B6/B7$  为  $X_2$ 。

$0.0042 \times B7 - 3.89 \times 10^{-5} \times DVI - 7.39 \times 10^{-5} \times RVI - 1.13 \times 10^{-5} \times NDVI + 1.05 \times 10^{-5} \times X_1 - 0.0038 \times X_2 - 0.035 \times slope + 0.0023 \times aspect$ 。

最终得到估测模型的相对误差为 15.53%,郁闭度估测模型精度为 81.75%。

3 结果与分析

3.1 偏最小二乘法模型的构建

考虑到遥感因子和 GIS 因子之间的量纲不同,先将所有样本数据进行中心标准化,统一量纲<sup>[19]</sup>。从 60 个建模数据中提取相关遥感因子、GIS 因子,利用 R 中的偏最小二乘函数包建立蓄积量估测模型,交叉验证结果(部分)见表 2。

表 2 60 个样本数据交叉验证结果(部分)

指标	截距	12 个 自变量	13 个 自变量	14 个 自变量	15 个 自变量	16 个 自变量
CV	1.008	0.903 7	0.908 8	0.923 7	0.956 5	0.984 8
adjCV	1.008	0.901 8	0.906 9	0.921 7	0.954 0	0.982 0

注:CV 为不同主成分个数对应的 PRESS(每个样本的预测误差平方和),adjCV 为调整后的 PRESS。

利用 validationplot 函数得到偏最小二乘法模型在不同主成分数下对应的 RMSEP(由留一交叉验证法算出的均方预测误差根)(图 4)。

根据成分数  $m = 12$  得到回归系数并建立最终模型(图 5)。

$Y = 4.50308749 \times \text{郁闭度} + 0.00100862 \times \text{Dem} - 0.11731689 \times B1 + 0.10231844 \times B2 + 0.08665483 \times B3 - 0.12036277 \times B4 - 0.03976178 \times B5 + 0.08052672 \times B6 - 0.02852109 \times B7 - 0.16971486 \times DVI + 0.20360285 \times RVI - 0.00191010 \times NDVI + 0.01856442 \times X_1 + 0.08454423 \times X_2 + 0.00904814 \times slope + 0.00062518 \times aspect$ 。

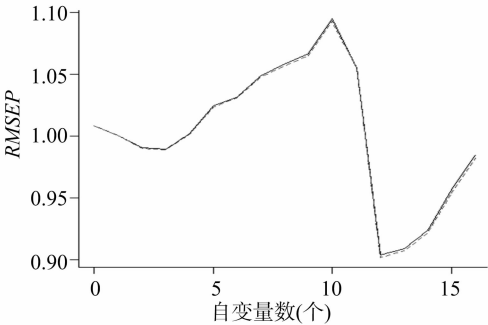


图 4 不同成分数对应的均方误差

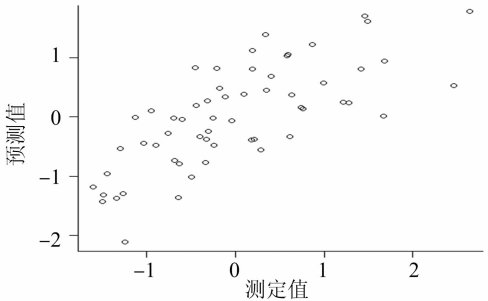


图 5 预测结果

利用 jack.test 函数进行回归系数的显著性检验(表 3)。郁闭度对所取样地高山松林蓄积量的影响极其显著,第 5、6 波段的灰度值对其影响较为显著。

3.2 香格里拉县高山松林蓄积量估测

本研究中实测样地大小为  $30\text{ m} \times 30\text{ m}$ ,与 Landsat 8 影像上 1 个像元的大小一致;因此,对香格里拉县高山松林总蓄积量进行估测时,应以像元为单位读取每个像元对应的线性组合。利用 ERDAS 的 Model maker 模块将数据带入已建立的模型中,得到香格里拉县高山松林总蓄积量为  $1\,372.406\text{ 万 m}^3$ 。

3.3 模型精度检验

将未参与建模的 19 个样地数据代入上述模型中,还原归一化的预测值,进行模型精度检验和适应性评价,采用调整决定系数( $R^2$ )、均方根误差(RMSE)、总预报偏差的相对误差(RE)作为评价指标,结果见表 4。

4 结论与讨论

由于偏最小二乘模型对自变量的选择要求较低,无需最优,且较多自变量因子有利于对主成分进行累计解释能力分析;因此,本研究引入所有变量参与森林蓄积量估测模型的建立,提高了模型的拟合效果,最终模型  $R^2$  为 0.777 5,拟合精度为 73.08%,可用于进行区域尺度高山松蓄积量的遥感估测。根据偏最小二乘模型估算得到 2014 年香格里拉县高山松林蓄积量为  $1\,372.406\text{ 万 m}^3$ 。研究中存在的不足之处仍有待改进。通过增加自变量因子的数量来估测蓄积量,虽效果令人满意,但无法真正解决问题。删除部分影响不显著因子后,模型精度急剧下降,表明影响不显著因子对蓄积量的估测也起到了重要作用。郁闭度是样地实测的结果,由此建立的模型较为可靠,而对香格里拉县高山松总蓄积量进行预测时,研究区像元对应的郁闭度数值仅能通过反演得到。本研究利

表 3 偏最小二乘法模型回归系数的显著性

自变量	估计值	标准误	df	t 值	P 值	
郁闭度	4.503 087 49	1.223 243 62	59	3.681 3	0.000 504 8	***
Dem	0.001 008 62	0.000 892 14	59	1.130 6	0.262 812 0	
B1	-0.117 316 89	0.147 613 92	59	-0.794 8	0.429 941 5	
B2	0.102 318 44	0.141 388 36	59	0.723 7	0.472 128 3	
B3	0.086 654 83	0.099 901 51	59	0.867 4	0.389 235 7	
B4	-0.120 362 77	0.098 186 03	59	-1.225 9	0.225 120 3	
B5	-0.039 761 78	0.022 354 97	59	-1.778 7	0.080 448 3	o
B6	0.080 526 72	0.043 600 81	59	1.846 9	0.069 777 0	o
B7	-0.028 521 09	0.059 140 05	59	-0.482 3	0.631 403 1	
DVI	-0.169 714 86	0.144 893 43	59	-1.171 3	0.246 183 7	
RVI	0.203 602 85	1.062 092 17	59	0.191 7	0.848 635 6	
NDVI	-0.001 910 10	0.112 246 51	59	-0.017 0	0.986 480 5	
B4_all	0.018 564 42	0.023 473 61	59	0.790 9	0.432 191 4	
B4_5_7	0.084 544 23	0.072 913 26	59	1.159 5	0.250 916 3	
slope	0.009 048 14	0.019 921 22	59	0.454 2	0.651 354 3	
aspect	0.000 625 18	0.001 703 75	59	0.366 9	0.714 973 2	

注：\*\*\*、\*\*、\*、o 分别代表极其显著、非常显著、很显著、较显著。

表 4 偏最小二乘法模型的精度

<i>R</i> <sup>2</sup>	<i>RMSE</i>	<i>RE</i>	<i>Accuracy</i>
0.777 5	36.90	23.18%	73.08%

用偏最小二乘法进行估测以减小误差,虽然估测精度为 81.75%,但误差传递问题可能导致香格里拉县高山松林总蓄积量的预测精度受到一定影响。

参考文献:

[1]程武学,杨存建,周介铭,等. 森林蓄积量遥感定量估测研究综述[J]. 安徽农业科学,2009,37(16):7746-7750.

[2]张友静,方有清,陈钦峦. 南方山地森林蓄积量遥感估算研究[J]. 国土资源遥感,1993(2):39-47.

[3]冯仲科,杨伯钢,罗旭,等. 应用 LIDAR 技术预测林分蓄积量[J]. 北京林业大学学报,2007,29(增刊2):45-51.

[4]琚存勇,蔡体久. 用泛化改进的 BP 神经网络估测森林蓄积量[J]. 林业科学,2006,42(12):59-62.

[5]杜晓明,蔡体久,琚存勇. 采用偏最小二乘回归方法估测森林郁闭度[J]. 应用生态学报,2008,19(2):273-277.

[6]洪奕丰,林辉,严恩萍,等. 基于偏最小二乘法的平南县森林蓄积量估测模型研究[J]. 中南林业科技大学学报,2011,31(7):80-85.

[7]刘琼阁,彭道黎,涂云燕. 基于偏最小二乘回归的森林蓄积量遥感估测[J]. 中南林业科技大学学报,2014,34(2):81-84,132.

[8]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.

[9]陈楚,关泽群,张鹏林,等. 利用 RS 和 GIS 的森林蓄积量偏最

小二乘估测研究[J]. 湖北林业科技,2004(4):25-28.

[10]罗批,郭继昌,李铨,等. 基于偏最小二乘回归建模的探讨[J]. 天津大学学报:自然科学与工程技术版,2002,35(6):783-786.

[11]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.

[12]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,2006:15-26.

[13]张凝,冯仲科,冯跃文,等. 旺业甸实验林场针叶林蓄积量估测模型的研究[J]. 中南林业科技大学学报,2013,33(11):83-87,108.

[14]张彦林. 基于 3S 技术的山东省森林蓄积量定量估测研究[D]. 北京:北京林业大学,2008.

[15]黄伟平,谭三清,张贵,等. 估测森林蓄积量的遥感因子选择研究[J]. 中南林业科技大学学报,2010,30(4):112-115.

[16]Tomppo E, Nilsson M, Rosengren M, et al. Simultaneous use of Landsat-TM and IRS-1 CWIFS data in estimating large area tree stem volume and aboveground biomass[J]. Remote Sensing of Environment,2002,82(1):156-171.

[17]Mäkelä H, Pekkarinen A. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data[J]. Forest Ecology and Management,2004,196(2/3):245-255.

[18]李崇贵,蔡体久. 森林郁闭度对蓄积量估测的影响规律[J]. 东北林业大学学报,2006,34(1):15-17.

[19]吴喜之. 复杂数据统计方法——基于 R 的应用[M]. 北京:中国人民大学出版社,2012.