

引用格式:刘唐,江涛,李昂,等.基于神经网络和不同立地质量的森林蓄积量遥感估测[J].山东科技大学学报(自然科学版),2019,38(2):25-35.

LIU Tang, JIANG Tao, LI Ang, et al. Remote sensing estimation of forest stock volume based on neural network and different site quality[J]. Journal of Shandong University of Science and Technology (Natural Science), 2019, 38(2): 25-35.

# 基于神经网络和不同立地质量的森林蓄积量遥感估测

刘唐<sup>1</sup>, 江涛<sup>1</sup>, 李昂<sup>1</sup>, 郭连杰<sup>2</sup>

(1. 山东科技大学 测绘科学与工程学院, 山东 青岛 266590; 2. 国家海洋信息中心, 天津 300171)

**摘要:**运用 BP 神经网络技术建立区分立地质量等级森林蓄积量遥感估测模型并探讨其适用性。以 2009 年黑龙江省伊春市凉水自然保护区森林资源二类调查数据为基础数据划分森林的立地质量等级,以森林蓄积量为研究对象,基于该地区 LANDSAT-TM 影像以及 DEM 数据提取遥感因子,采用 BP 神经网络方法构建区分立地质量的森林蓄积量遥感估测模型,并引入回归分析方法和不区分立地质量的模型予以比较。结果表明,基于不同立地质量等级的模型明显好于不区分立地质量等级的估测模型,且 BP 神经网络模型较回归分析模型可以更好地预测森林的蓄积量。经过对比检验,基于不同立地质量等级的 BP 神经网络模型性能优异,验证总体预测精度高达 97% 以上,实测值与预测值的  $R^2$  在 0.94 左右;不区分立地质量等级的 BP 神经网络模型的  $R^2$  为 0.89,预测精度 95% 左右。同等条件下,BP 神经网络模型较多元线性回归模型的预测精度约提高了 3%~5%, $R^2$  值提高了 0.1 左右。

**关键词:**立地质量;森林蓄积量;主成分分析;神经网络

中图分类号:S771.8

文献标志码:A

文章编号:1672-3767(2019)02-0025-11

DOI:10.16452/j.cnki.sdkjzk.2019.02.003

## Remote sensing estimation of forest stock volume based on neural network and different site quality

LIU Tang<sup>1</sup>, JIANG Tao<sup>1</sup>, LI Ang<sup>1</sup>, GUO Lianjie<sup>2</sup>

(1. College of Geomatics, Shandong University of Science and Technology, Qingdao, Shandong 266590, China;

2. National Marine Information Center, Tianjin 300171, China)

**Abstract:** This paper discussed a remote sensing estimation model for the forest stock volume with different site qualities by using BP neural network technology, and its applicability. Based on the data of the second type of forest resources in the Liangshui Nature Reserve of Yichun City, Heilongjiang Province in 2009, the site qualities of the forest were divided. With the forest stock volume as research object, the remote sensing factor was extracted based on the LANDSAT-TM image and DEM data in the region. The BP neural network method was used to construct a remote sensing estimation model for forest stocks with different site qualities. A regression analysis method and a non-separate quality model were introduced for comparison. The results show that the models based on different site qualities are better than the estimation models without the distinction of site qualities and that the BP neural network

收稿日期:2018-03-18

基金项目:国家自然科学基金项目(41706194);山东省自然科学基金项目(ZR2016DB23)

作者简介:刘唐(1994—),男,山东泰安人,硕士研究生,主要从事高光谱遥感分类等方面的研究。

E-mail:sddpltwanqiu@163.com

江涛(1961—),男,山东济南人,教授,博士生导师,主要从事遥感数字图像处理与分析、资源与环境遥感、地理信息系统开发与应用方面研究,本文通信作者。E-mail:jiangt@126.com

model can better predict the forest accumulation volume than the regression analysis model. The comparison indicates that the BP neural network model based on different site qualities has excellent performance, the overall prediction accuracy being over 97% and the measured value and predicted  $R^2$  being around 0.94, while the  $R^2$  of the BP neural network model without distinguishing site qualities is 0.89 and the prediction accuracy is about 95%. Under the same conditions, the prediction accuracy of the BP neural network model is about 3%-5% higher than that of the multiple linear regression model, and the  $R^2$  value is enhanced by about 0.1.

**Key words:** site quality; forest stock volume; principal component analysis; neural network

森林被誉为“地球之肺”,在全球生态系统的平衡中发挥着重要作用,是人类以及多种物种赖以生存和发展的基础<sup>[1]</sup>。森林蓄积量是衡量森林健康与否的重要指标,也是政府掌握国家森林资源状况和制定计划采伐、森林经营管理措施的重要依据。传统的蓄积量测定方法是一、二类森林调查,人力、物力投入巨大,而且调查周期长,难以掌握森林的动态发展状况。20世纪70年代以来,遥感技术如雨后春笋,迅速应用于各个领域,利用遥感技术进行森林蓄积量的定量估测也取得了一系列的重要进展<sup>[2-4]</sup>。

基于遥感技术对森林蓄积量进行估算,主要是利用遥感影像数据以及少量实地调查数据,采用数学方法构建森林蓄积量的估算模型,从而使得模型可以直接应用在森林蓄积量估算中<sup>[5]</sup>。国外很早就有遥感数据定量估算森林蓄积量的应用。早在1988年,Nelson等<sup>[6]</sup>就利用Laser数据估算了森林的蓄积量和生物量。Gemmell等<sup>[7]</sup>基于专题制图仪(thematic mapper, TM)数据,研究了各波段影像以及其他实测数据对估测森林蓄积量的影响度,为TM影像数据在定量估测蓄积量方面提供了理论依据。Fazakas等<sup>[8]</sup>使用TM影像数据结合最近邻法估算了森林的蓄积量。国内蓄积量的遥感估算相对国外起步较晚,刘琼阁等<sup>[1]</sup>提出了基于偏最小二乘法的森林蓄积量遥感估测模型,程武学等<sup>[9]</sup>对基于小班的各种蓄积量预测的合理性进行了诊断,证实了其合理性。近年来,得益于神经网络超强的非线性处理能力,国内外研究人员将其成功地运用到森林蓄积量的估测中,取得了一定的成果。王臣立等<sup>[10]</sup>运用3层的后向传输(back propagation, BP)神经网络建立了基于植被指数的热带森林蓄积量估算模型,同等条件下较回归分析模型的预测效果更好。许炜敏等<sup>[11]</sup>构建了结构为10:3:1的杉木林蓄积量BP神经网络模型。吴达胜等<sup>[12]</sup>建立了涵盖地形、地貌、气候气象、土壤、林分结构特征及光谱特征的森林资源蓄积量预测自变量因子集,并应用改进的BP神经网络模型对森林资源蓄积量进行了预测,所取得的模型充分利用了多源数据,且具备很强的泛化能力。

森林地位级表是森林管理中的常用数表之一。早在上世纪90年代,林昌庚等<sup>[13]</sup>就提出了适合于中国森林调查数据的地位级表编制方法,周春国等<sup>[14-15]</sup>对其进行了完善,提出了最佳导线曲线的测定方法。研究表明,区分立地质量等级进行预测建模可以显著提高森林蓄积量的估算精度。王艳婷等<sup>[16]</sup>运用岭估计法,分不同地类分别建立森林蓄积量估测方程,结果表明,当影响蓄积量估测的自变量间存在复共线性时,分地位级建立模型的精度明显优于统一建模的精度。刘俊等<sup>[5,17]</sup>建立了基于不同立地质量的杉木和松树的森林蓄积量遥感估测模型,采用了更为准确的地位级划分方法,但在建模时只利用了遥感影像主成分分析的第一主成分进行线性回归建模,并没有充分利用遥感影像和遥感辅助数据,由于森林蓄积量受到多方面因素的影响,单一变量难以准确估算蓄积量,因此估算精度不高。

本研究以2009年TM影像数据和凉水自然保护区森林二类小班数据为数据源,采用地位级法划分林分的立地质量等级,结合回归分析与神经网络分析算法,对红松树种以不区分立地质量和区分不同立地质量等级分别建立森林蓄积量的遥感估测模型,旨在为森林蓄积量的遥感监测提供一种新的思路。

## 1 研究基础

### 1.1 研究区概况

凉水自然保护区位于黑龙江省伊春市带岭区,小兴安岭山脉的东南段。地理坐标为东经 $128^{\circ}47'8''$ — $128^{\circ}57'19''$ ,北纬 $47^{\circ}6'49''$ — $47^{\circ}16'10''$ ,东西宽13.0 km,南北长17.0 km,总面积为12 133 km<sup>2</sup>(图1)。该地区深受海洋性气候的影响,具有明显的温带大陆性季风气候特征,总的特点是冬长夏短,冬季严寒干燥,夏季

湿凉多雨。年平均气温  $-0.3\text{ }^{\circ}\text{C}$ , 年平均最高气温  $7.5\text{ }^{\circ}\text{C}$ , 年平均最低气温  $-6.6\text{ }^{\circ}\text{C}$ ; 年平均降水量  $676\text{ mm}$ , 年平均相对湿度  $78\%$ , 年平均蒸发量  $805\text{ mm}$ 。土壤类型主要有暗棕壤、沼泽土、草甸土、泥炭土 4 类。该地区森林资源丰富, 其中红松是这里的优势树种。

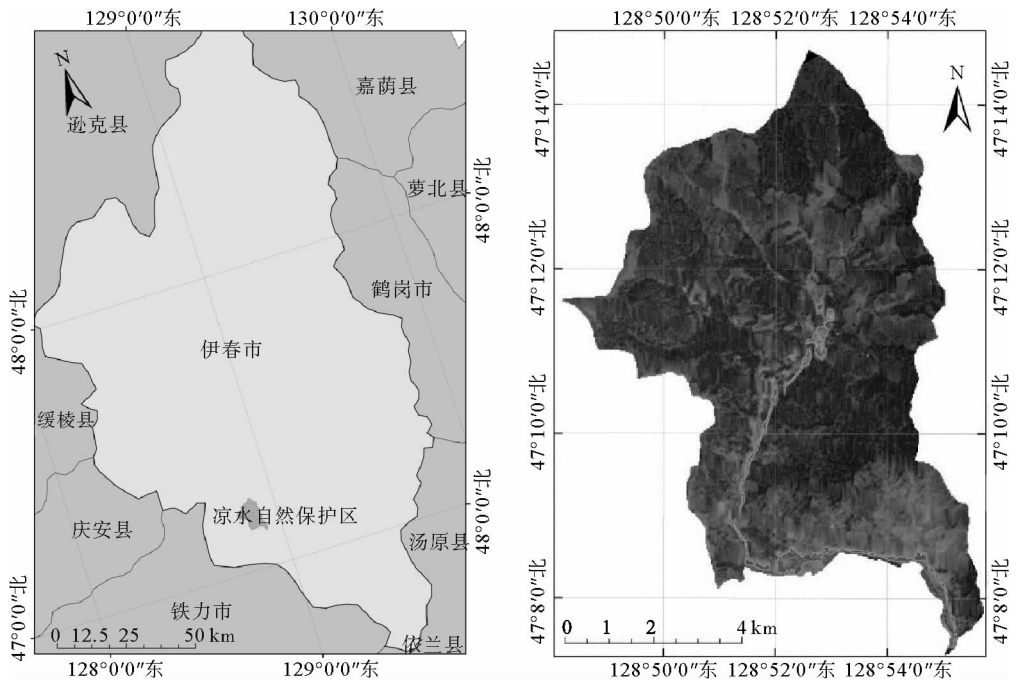


图 1 研究区地理位置与 TM 影像数据图

Fig. 1 Geographical location and TM image data of the study area

## 1.2 数据获取

### 1.2.1 遥感影像数据

实验使用 Landsat TM 遥感影像, 数据在美国地质勘探局 (United States Geological Survey, USGS) 官网 (<https://earthexplorer.usgs.gov>) 上获取, 综合考虑时相和云量等因素, 选取成像时间为 2009 年 5 月 29 日覆盖凉水自然保护区的 TM 遥感影像。利用 ENVI5.3 对影像进行辐射定标、大气校正、裁剪与镶嵌等预处理。

### 1.2.2 高程数据

凉水自然保护区位于小兴安岭地区, 地形复杂多样, 海拔  $290\sim 550\text{ m}$ , 以山地、丘陵地形为主。实验运用全球数字高程模型 (global digital elevation model, GDEM) 数据, 对应的分辨率精度为  $30\text{ m}$ 。

### 1.2.3 森林资源二类调查数据

获取凉水自然保护区 2009 年森林资源二类调查小班矢量数据, 属性表中包括小班林种、起源、地类、公顷株数、郁闭度、高度、自然度、林分高、优势树高、优势径阶、枯倒树种、下木名称、地权使用权、地被总盖度、地被名称、地貌、部位、坡向、坡度、树种组成、林分类型、树种年龄、树种直径、树高、林层结构、龄级、龄组、小班每公顷蓄积量、小班总蓄积量等。

## 1.3 研究技术路线

运用 2009 年凉水自然保护区二类森林资源调查数据和同年度该区 TM 影像作为数据源, 首先, 对数据进行预处理后, 根据前人研究经验和方法, 寻找最合适立地指数导向曲线来划分森林各小班数据的立地质量等级。其次, 筛选出共线性低且与森林蓄积量相关性高的遥感因子变量作为模型的自变量, 对自变量和蓄积量之间分别进行基于回归分析和基于神经网络分析的模拟, 并进行精度检验, 筛选出最合适的森林蓄积量估测模型。研究技术路线如图 2 所示。

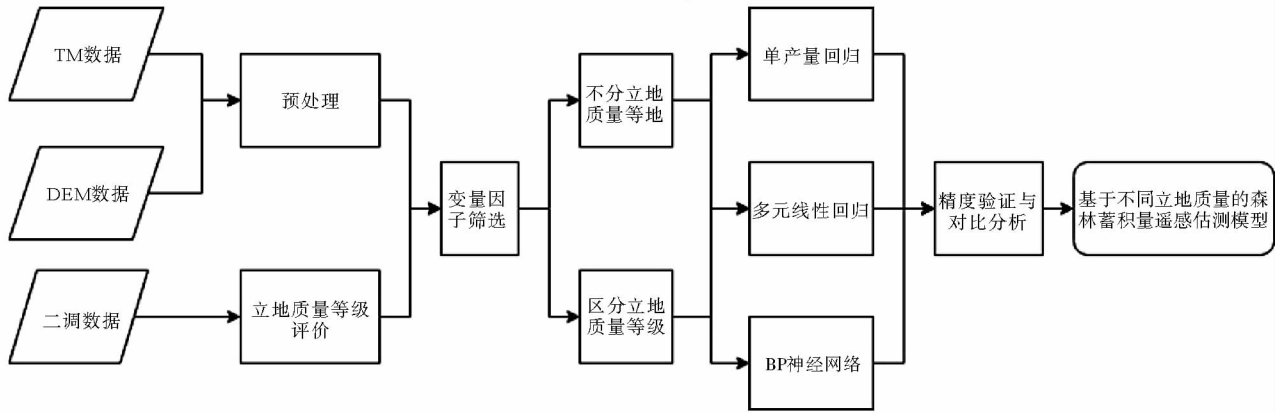


图2 技术路线图

Fig. 2 Technology roadmap

## 2 地位级划分与因子选取

### 2.1 立地质量等级评价

采用地位级法进行立地质量的评价。地位级表以林分平均高与平均年龄的关系为依据，是对同龄纯林分地区分树种编制、反映林地生产力高低的数表，是森林经营及森林调查中主要的常用数表之一<sup>[10]</sup>。按相同年龄时，林分高度的变动程度划分为若干个等级，通常为5~7级，以罗马数字I、II、III...等符号依次表示，将每一等级所对应的各个树龄时的平均高度列成表，即为地位级表<sup>[10]</sup>。编制地位级表的关键就是导向曲线的选定，前人<sup>[10-12]</sup>提出了多种计算导向曲线的函数模型与方法，但各种导向曲线模型的选择因数据而异。基于凉水自然保护区森林资源二类小班调查数据，对优势树种为红松且郁闭度0.4以上的小班，依据平均树高和平均年龄建立一条代表中等立地等级条件下的地位级导向曲线<sup>[11]</sup>(图3)。通过不同数学模型比较选出对数与逆函数结合的曲线作为红松地位级导向曲线的最优模型，最终所得方程为：

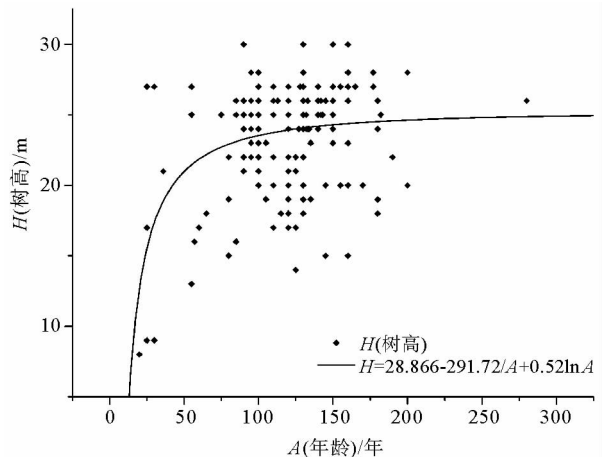


图3 最佳地位级导向曲线

Fig. 3 The best sue level guide curve

$$H=28.866-291.720/A+0.524\ln A. \tag{1}$$

其中， $H$ 代表小班的平均高， $A$ 代表小班的平均年龄。

目前的研究中，通常以导线曲线为基础，采用比例法确定地位等级的上下界线，以导向曲线的1.35倍为上界线，0.55倍为下界线<sup>[10]</sup>。用 $f(A)$ 表示导向曲线方程，则上下界模型分别为 $1.35f(A)$ 和 $0.55f(A)$ 。把杉木分为5个地位级数，各等分高度间隔为 $0.55f(A)$ 、 $0.75f(A)$ 、 $0.9f(A)$ 、 $1.05f(A)$ 、 $1.35f(A)$ ，按照以上分类办法，得到地位级表，如表1所示。

结合表1，将地位级为I、II的归为立地质量差，地位级为III的归为立地质量中等，将地位级为IV、V的归为立地质量优。以此数据作为立地质量划分的结果分别用来拟合不同立地质量的森林蓄积量的估算模型。

### 2.2 影响因子定义

根据以往的研究，依据单变量或者多变量的线性组合，可以不同程度地突出某些需要的信息以及抑制无关的信息<sup>[13]</sup>。本次调查数据共有小班431个，删除优势树种不是红松和郁闭度低于0.4的小班数据，并利用标准差分析法剔除明显的异常数据。剔除异常样本后各级统计数据如表2。

表 1 红松地位级  
Tab.1 The site level of Korean Pine

小班平均年龄	各地位级的小班平均高的范围/m					小班平均年龄	各地位级的小班平均高的范围/m				
	I	II	III	IV	V		I	II	III	IV	V
30	<9	9<h<13	13<h<17	17<h<25	>25	130	<18	18<h<22	22<h<26	26<h<30	>30
50	<16	16<h<19	19<h<22	22<h<26	>26	150	<20	20<h<23	23<h<26	26<h<30	>30
70	<17	17<h<21	21<h<24	24<h<27	>27	170	<20	20<h<23	23<h<26	27<h<30	>30
90	<18	18<h<21	21<h<24	24<h<28	>28	190	<20	20<h<23	25<h<26	26<h<30	>30
110	<18	19<h<22	22<h<25	25<h<28	>28						

表 2 不同立地质量等级的建模及验证样本数  
Tab.2 Modeling and validation samples of different site qualities

立地质量类型	建模样本数/个	检验样本数/个	立地质量类型	建模样本数/个	检验样本数/个
不分立地质量	135	58	中(III)	53	23
差(I、II)	27	12	优(IV、V)	55	23

查阅相关文献后,在前人研究基础上,从 TM 影像中选取可能影响蓄积量的遥感信息参数,如 TM 影像的第一波段( $B_1$ )、第二波段( $B_2$ )、第三波段( $B_3$ )、第四波段( $B_4$ )、第五波段( $B_5$ )、第七波段( $B_7$ );多波段组合( $B_1/B_3$ 、 $B_2/B_4$ 、 $B_3/B_5$ 、 $B_3 \times B_4/B_7$ 、 $B_4 \times B_5/B_7$ 、 $(B_1 - B_7)/B_4$ 、 $(B_3 + B_4 + B_5)/(B_1 + B_2 + B_7)$ ),下文分别用  $B_{13}$ 、 $B_{24}$ 、 $B_{35}$ 、 $B_{347}$ 、 $B_{457}$ 、 $B_{174}$ 、 $B_{345127}$  表示);植被指数如归一化植被指数(normalized difference vegetation index, NDVI)、比值植被指数(ratio vegetation index, RVI)、增强型植被指数(enhanced vegetation index, EVI)、差值植被指数(difference vegetation index, DVI);主成分分析的第一分量(principal componet analysis, PCA);经过遥感影像数据反演的土壤湿度(soil moisture, SM)、地表温度(land siface temperature, LST);经过研究区数字高程数据提取的坡向(Asspect)、坡度(Slope)、海拔高度(ASL)等 23 个变量作为因子,如表 3 所示。

表 3 变量因子表  
Tab.3 Variable factor table

序号	因子	序号	因子	序号	因子	序号	因子	序号	因子	序号	因子
1	$B_1$	5	$B_5$	9	EVI	13	$B_2/B_4$	17	$(B_3 + B_4 + B_5)/(B_1 + B_2 + B_7)$	21	Aspect
2	$B_2$	6	$B_7$	10	DVI	14	$B_3 \times B_4/B_7$	18	PCA	22	Slope
3	$B_3$	7	NDVI	11	$B_1/B_3$	15	$B_4 \times B_5/B_7$	19	SM	23	ASL
4	$B_4$	8	RVI	12	$B_3/B_5$	16	$(B_1 - B_7)/B_4$	20	LST		

自变量的选择是建立模型的首要任务。分别计算表 3 选取的因子在不同立地质量下与获取的蓄积量之间的相关系数 R 和他们的方差扩大因子(variance inflation factor, VIF),所得结果如表 4 所示。

如表 4 可见,除去地形和地表温度、湿度之外,森林的蓄积量与以上各个因子的相关系数均较高,达到了 0.01 的显著性水平。无论是否划分立地质量等级,NDVI 与蓄积量的相关性均较高,其次是 TM 影像的第 3 波段、遥感图像主成分分析的第一主成分等变量,说明 NDVI 以及遥感影像的红色波段是监测森林蓄积量的敏感波段。然而,红色波段( $B_3$ )、NDVI 变量的方差扩大因子(VIF)很高,说明其自变量观察值之间存在严重的共线性。各个变量之间的相关性也比较高,会造成模型的冗余<sup>[17]</sup>。因此在建模之前必须进行因子的筛选,确定进入模型的自变量。

### 2.3 因子筛选方法

#### 2.3.1 所有子集回归法

表4 不同立地质量等级的变量因子相关信息

Tab.4 Information about the selected variable factors of different site qualities

因子	不分等级		差(I、II)		中等(III)		优(IV、V)	
	R	VIF	R	VIF	R	VIF	R	VIF
B <sub>1</sub>	0.468**	553.390	0.574**	1 167.370	0.386**	1057.400	0.458**	931.690
B <sub>2</sub>	0.557**	571.800	0.655**	1 416.340	0.584**	1191.400	0.465**	1 248.053
B <sub>3</sub>	0.644**	3215.100	0.728**	4852.300	0.736**	5 142.800	0.590**	1921.500
B <sub>4</sub>	0.429**	478.150	0.507**	1 961.150	0.328**	812.110	0.484**	326.246
B <sub>5</sub>	0.565**	242.810	0.665**	749.900	0.592**	458.670	0.481**	274.705
B <sub>7</sub>	0.612**	424.350	0.651**	1 089.630	0.572**	581.610	0.618**	572.725
NDVI	0.664**	462.190	0.703**	1934.230	0.734**	623.520	0.667**	623.370
RVI	0.519**	577.010	0.511**	1 612.720	0.620**	980.250	0.558**	677.516
EVI	0.607**	589.100	0.690**	1 495.480	0.698**	989.220	0.569**	761.488
DVI	0.588**	47.700	0.678**	84.920	0.671**	145.610	0.563**	254.738
B <sub>13</sub>	0.592**	351.690	0.585**	1 001.170	0.665**	380.250	0.632**	425.143
B <sub>35</sub>	0.611**	191.760	0.681**	783.970	0.693**	465.900	0.589**	191.992
B <sub>174</sub>	0.542**	15.380	0.631**	86.610	0.507**	14.710	0.499**	32.180
B <sub>345127</sub>	0.588**	473.350	0.610**	1 247.830	0.705**	558.740	0.593**	645.655
B <sub>347</sub>	0.486**	105.480	0.444**	105.780	0.612**	134.980	0.533**	175.938
B <sub>157</sub>	0.462**	118.280	0.409**	76.850	0.582**	180.930	0.507**	172.387
PCA	0.648**	5.560	0.700**	28.203	0.661**	5.714	0.597**	5.991
SM	-0.015	1.131	-0.227	1.965	0.038	1.223	-0.116	1.116
LST	-0.049	1.892	-0.287	4.000	0.127	1.705	0.248*	1.708
Aspect	-0.092	1.572	0.166	2.312	-0.092	1.293	-0.073	1.758
Slope	0.063	1.917	0.127	2.747	-0.171	1.482	-0.109	2.390
ASL	-0.046	1.803	0.156	2.807	-0.069	1.793	-0.191	1.980

注:1. \*\*表示其达到了0.01的显著性水平,\*表示其达到了0.05的显著性水平。2.VIF值表示自变量观察值之间复共线性程度,VIF越大代表共线性越显著。一般来说,当0<VIF<10,共线性视为不存在;当10≤VIF<100,存在一定的共线性;当VIF≥100,存在严重的共线性。

应用回归分析处理实际问题时,逐步回归是一种最常用的选择回归变量的方法。但由于这种方法不计算所有可能性,因此得到的变量子集只是局部“最优”,可能会遗漏全局最优的变量子集。因此,本研究选用所有子集回归法,即若有m个可供选择的变量,那么所有可能的回归方程就有2<sup>m</sup>个,然后利用选元准则挑出最优的方程。常见的选元准则有C<sub>p</sub>统计量、自由度调整复决定系数和赤池信息量等。本研究的目的是建立预测模型,故运用使预测值的均方根误差尽量小的C<sub>p</sub>统计量<sup>[18-19]</sup>。马洛斯(Mallows)<sup>[20]</sup>从预测的角度提出了用C<sub>p</sub>统计量来选择自变量。设n为样本数,SSE<sub>p</sub>表示p个变量模型的残差平方和,则

$$C_p = \frac{SSE_p}{\hat{\sigma}^2} - n + 2p = (n - m - 1) \frac{SSE_p}{SSE_m} - n + 2p \quad (2)$$

其中, $\hat{\sigma}^2 = \frac{1}{n - m - 1} SSE_m$ ,表示全模型中的无偏估计,此时,选择使C<sub>p</sub>最小的自变量子集,这个子集对应的回归方程就是最优回归方程。

使用SAS软件进行最优子集的选择,取C<sub>p</sub>统计量作为准则进行所有子集回归,得到森林蓄积量预测的最优子集,如表5所示。

表 5 各立地质量等级的最优子集

Tab. 5 Optimal subset of each site quality grades

	$C_p$	$R_a^2$	$R^2$	个数	变量因子名称
全部	2.452	0.553	0.572	9	$B_4, NDVI, RVI, B_{13}, B_{35}, B_{147}, B_{345127}, B_{437}$
差	1.231	0.839	0.878	9	$B_1, B_7, NDVI, EVI, B_{13}, B_{24}, B_{345127}, B_{437}, PCA$
中	2.084	0.693	0.711	4	$B_3, NDVI, B_{35}, B_{147}$
优	2.246	0.671	0.702	7	$B_3, B_1, B_5, B_{345127}, Slope, Lst, Asl$

### 2.3.2 基于平均贡献值(MIV)的因子筛选

平均贡献值(mean impact value, MIV)方法选择神经网络输入变量,用于确定输入神经元对应变量的影响大小,其符号代表相关的方向,绝对值大小代表影响的相对重要性。具体过程为:首先,在网络训练终止后,将训练样本  $P$  中每一个自变量特征在其原值基础上分别加、减 10% 构成新的样本  $P_1$  和  $P_2$ ,将  $P_1$ 、 $P_2$  分别作为仿真样本利用已经建成的网络进行仿真,得到仿真结果  $A_1$  和  $A_2$ ,求出  $A_1$  和  $A_2$  的差值,即得出该自变量对于因变量的 MIV 值;其次,根据 MIV 绝对值的大小为各自变量排序,得到各自变量对网络输出影响相对重要性的位次表;最后,根据位次采用逐步剔除法实现变量的筛选。由于 BP 神经网络的权重是随机得到的,每次训练生成的 MIV 值不同,故选取 10 次计算的均值作为变量的 MIV 值。使用 MATLAB 的神经网络工具,构建一个简单的神经网络,并实现各因子的 MIV 值计算,结果如表 6 所示。

表 6 各因子对网络的平均贡献值(MIV)

Tab. 6 Mean impact value (MIV) of each factor on the network

因子	MIV	因子	MIV	因子	MIV	因子	MIV	因子	MIV
$B_{547}$	2 845.171	$B_{147}$	604.968	$B_{24}$	334.697	$B_3$	179.725	<i>Slope</i>	4.956
$B_{13}$	2 129.974	$B_{437}$	477.314	<i>Ast</i>	303.562	<i>NDVI</i>	129.330	<i>SM</i>	1.741
<i>DVI</i>	1 576.015	$B_{35}$	434.621	<i>PCA</i>	301.688	<i>Aspect</i>	98.898	$B_4$	0.908
$B_2$	1 183.196	$B_5$	375.447	$B_7$	200.946	$B_{345127}$	81.788		
$B_1$	1 097.633	<i>RVI</i>	338.477	<i>EVI</i>	193.468	<i>Lst</i>	16.232		

## 3 蓄积量估测模型的建立

### 3.1 基于回归分析的森林蓄积量估测模型

首先进行单产量的回归,以小班的计蓄积量为因变量、小班提取的 TM 影像第一主成分遥感因子为自变量,对红松区区分立地质量等级和不区分立地质量两种办法分别建立蓄积量估算模型。根据散点图分布,选取线性函数和幂函数两种模型。各个模型的拟合结果如表 7 所示。

综合对比可见,无论是何种建模方法,模型的  $P$  值都小于 0.01,通过了 0.01 的显著性水平检验,且均通过了  $F$  检验,说明森林的蓄积量和 TM 影像的第一主成分之间存在较好的回归关系。其中,幂函数模型的决定系数均在 0.6 以上,最高为 0.707,较普通线性模型有着显著的提高。另外,基于不同立地质量的模型较不区分立地质量的模型效果略有提升,但提升并不显著。

为充分利用遥感影像提供的信息,结合最优子集法筛选出的子集,利用 SPSS 进行多元线性回归,再使用剔除法依次除去  $VIF$  值较高的变量,直至模型中所有变量的  $VIF$  都低于 10,得出不同立地质量的森林蓄积量估算模型,如表 8 所示,其中  $R_a^2$  表示自由度调整复相关系数,相比于  $R^2$ ,它能更好地刻画多元模型的拟合效果。

可见,同等条件下,相较于单产量回归模型,多元线性模型的  $R^2$  有明显的提高。特别是基于不同立地地质

量等级进行建模时,模型的拟合效果显著提升, $R^2$ 分别从 0.707、0.620、0.601 提升到 0.819、0.720、0.712,这一现象表明基于不同立地质量等级分别对森林蓄积量进行建模中的有效性,且遥感因子与森林蓄积量之间存在复杂的关系。为进一步挖掘数据中存在的非线性关系,引进了处理非线性问题表现良好的 BP 神经网络技术。

表 7 单产量因子的回归拟合结果

Tab.7 Fitting result of each variable factor

立地质量	函数	模型方程	$R^2$	F 检验值	显著性
不分地位级	线性	$Y=0.522 \times PCA + 2\ 113.767$	0.498	181.242	$P < 0.01$
	幂函数	$Y=4.406 \times PCA^{0.808\ 9}$	0.619	294.436	$P < 0.01$
差 (地位级 I II)	线性	$Y=0.424 \times PCA + 2\ 401.254$	0.445	26.429	$P < 0.01$
	幂函数	$Y=3.293 \times PCA^{0.849\ 7}$	0.707	79.441	$P < 0.01$
中等 (地位级 III)	线性	$Y=0.610 \times PCA + 1\ 599.910$	0.509	75.644	$P < 0.01$
	幂函数	$Y=9.912 \times PCA^{0.711\ 1}$	0.620	83.283	$P < 0.01$
优 (地位级 IV V)	线性	$Y=0.517 \times PCA + 2\ 199.812$	0.519	78.848	$P < 0.01$
	幂函数	$Y=3.273 \times PCA^{0.843\ 6}$	0.601	115.908	$P < 0.01$

表 8 森林蓄积量的多元回归结果

Tab.8 Multiple regression result of forest stock volume

立地质量	模型方程	$R^2$	$R_a^2$	F 检验值	显著性
不分地位级	$Y=1\ 325.13+0.68 \times B_4+5.82 \times B_{13}+2.41 \times B_{35}+3.22 \times B_{147}-0.02 \times B_{137}$	0.649	0.631	32.798	$P < 0.01$
差(地位级 I II)	$Y=980.26+0.58 \times B_4+0.41 \times EVI+51.68 \times B_{24}+0.98 \times B_{437}+0.38 \times PCA$	0.819	0.783	23.311	$P < 0.01$
中等(地位级 III)	$Y=813.64+0.94 \times B_3+26.81 \times NDVI+3.24 \times B_{147}-6.85 \times B_{35}$	0.720	0.703	41.647	$P < 0.01$
优(地位级 IV V)	$Y=-25\ 291.88+2.94 \times B_4+0.68 \times B_5+9.37 \times B_{345\ 127}+176.73 \times Slope+1\ 190.43 \times Lst-7.06 \times Asl$	0.712	0.691	22.533	$P < 0.01$

### 3.2 基于 BP 神经网络的蓄积量预测模型

BP 神经网络是当前应用最广泛的神经网络模型之一,是一种按照误差逆向传播算法训练的多层前馈神经网络,具有任意复杂的模式分类能力和优良的多维函数映射能力,可以很好地处理非线性问题<sup>[21]</sup>。

利用神经网络建模的过程分为数据预处理、网络训练、调参、网络再训练、模型评价等 5 个阶段。首先,将数据进行归一化以消除量纲的影响;然后进行网络结构的设计,网络共 3 层,分别是输入层、隐含层和输出层;使用控制变量法分别对比不同传递函数、不同训练函数、不同神经元等各种网络参数下网络的表现,经多次训练、调参,得到各立地质量等级的网络结构及其在训练集上的系数。其中,模型的传递函数均选择 Purelin 函数,立地质量为差的数据因样本较少,选择 Trainlm 为训练函数,其他类别使用的训练函数为 Trainrp。模型以网络拟合的均方误差(mean squared error, MSE)为目标损失函数。各模型的网络结构均已在表 9 中说明。模型的迭代次数设置为 1 000 次,每个实验重复 10 次,取 10 次训练的 MSE 的均值及其标准差,如表 9 所示。

可见,利用神经网络技术对不同立地质量的数据进行建模时,模型的均方误差可以收敛于更小的水平,模型的拟合性能得到显著改善。为全面对比分析模型效果,下面进行模型对比验证。



表 9 不同立地质量的神经网络模型及训练参数

Tab. 9 Neural network model and training parameters of different site qualities

	传递函数	训练函数	输入层神经元数	隐含层神经元数	训练次数	MSE/( $\times 10^{-3}$ )
不分地位级	Purelin	Trainrp	22	16	893	5.682 $\pm$ 0.645
差(地位级 I、II)	Purelin	Trainlm	8	6	405	0.018 $\pm$ 0.009
中等(地位级 III)	Purelin	Trainrp	13	8	1 000	0.645 $\pm$ 0.096
优(地位级 IV、V)	Purelin	Trainrp	16	11	909	1.323 $\pm$ 0.241

### 3.3 预测模型对比与验证

采用随机预留的验证数据验证模型的精度,将验证集数据应用到模型,得到森林蓄积量的估测值。将估测值与实际蓄积量之间进行线性拟合,得到的  $R^2$ 、斜率及其总体估计精度作为精度评价指标,完成模型的精度检验。其中,曲线斜率越接近 1 代表预测值和实测误差越小,总体估计精度( $P$ )的计算公式为:

$$P = \frac{\sum \text{预测值}}{\sum \text{实测值}} \times 100\% \quad (3)$$

检验结果如图 4 和表 10 所示。综合对比以上验证结果,可见:

- 1) 模型预测值与实测值之间的  $R^2$  均达 0.7 以上,总体估测精度在 78% 以上,使用遥感影像可以较好地预测森林蓄积量,其中,多自变量模型的预测效果要远好于单产量模型;
- 2) BP 神经网络模型表现要好于单产量回归和多元线性回归模型,其预测值和实测值的  $R^2$  均达到 0.9 及以上,曲线斜率也更接近于 1,总体预测精度也大大提升,达到了 95% 以上;
- 3) 基于不同立地质量等级进行建模,无论采用何种建模方式,较不区分立地质量时,预测精度、 $R^2$  都有显著的提升。特别是 BP 神经网络模型的总体预测精度在 97% 以上, $R^2$  达到了 0.94 以上,达到了很高的预测水平。由此可见基于不同立地质量建立森林蓄积量估算模型的思路是正确的,在同等条件下可以使模型性能得到显著的提升。

表 10 不同立地质量模型的精度评价结果

Tab. 10 Accuracy evaluation results of different site quality models

立地质量	样本	实测值(总)	单产量回归模型				多元线性回归模型				BP 神经网络模型			
			估测值(总)	$R^2$	总体估测精度/%	斜率	估测值(总)	$R^2$	总体估测精度/%	斜率	估测值(总)	$R^2$	总体估测精度/%	斜率
不分	61	333 442	261 685	0.713	78.48	0.607	301 731	0.762	90.49	0.761	349 014	0.899	95.33	0.856
差	11	112 097	95 069	0.996	84.81	0.677	106 672	0.926	95.16	0.824	109 810	0.996	97.96	1.001
中	25	132 603	148 369	0.778	88.11	0.687	139 233	0.861	95.00	0.809	131 714	0.941	99.33	0.839
优	24	152 524	138 324	0.773	90.69	0.783	143 586	0.822	94.14	1.015	155 651	0.940	97.95	0.890

## 4 结论与展望

本研究以 2009 年凉水自然保护区二类森林调查数据,结合 2009 年该地区的 TM 影像,以各小班面积上遥感因子的信息总量为自变量,各小班的计蓄积量为因变量,对红松建立不分立地质量和区分立地质量等级的森林蓄积量遥感估算模型。得到以下结论:

- 1) 采用地位级法对森林的立地质量进行评定,将各个小班数据分别划分为 3 种立地质量等级,然后基于不同立地质量等级的数据分别建模,蓄积量预测精度显著改善;
- 2) BP 神经网络具有很强的非线性处理能力,基于不同立地质量等级建立的模型经过反复调参之后总体预测精度可达 97% 以上,模型可以很好地跟踪实测值,较单产量回归模型和多元线性回归模型具有更强的预测能力。

模型适用于黑龙江省伊春市凉水自然保护区红松林蓄积量预测,对其他优势树种的小班是否有效还有待探讨,立地质量的划分和建模方法的选择可为之提供借鉴。建模所需样本数目有限,为了提高估测模型稳定性和预测能力,还应考虑增加样本数。

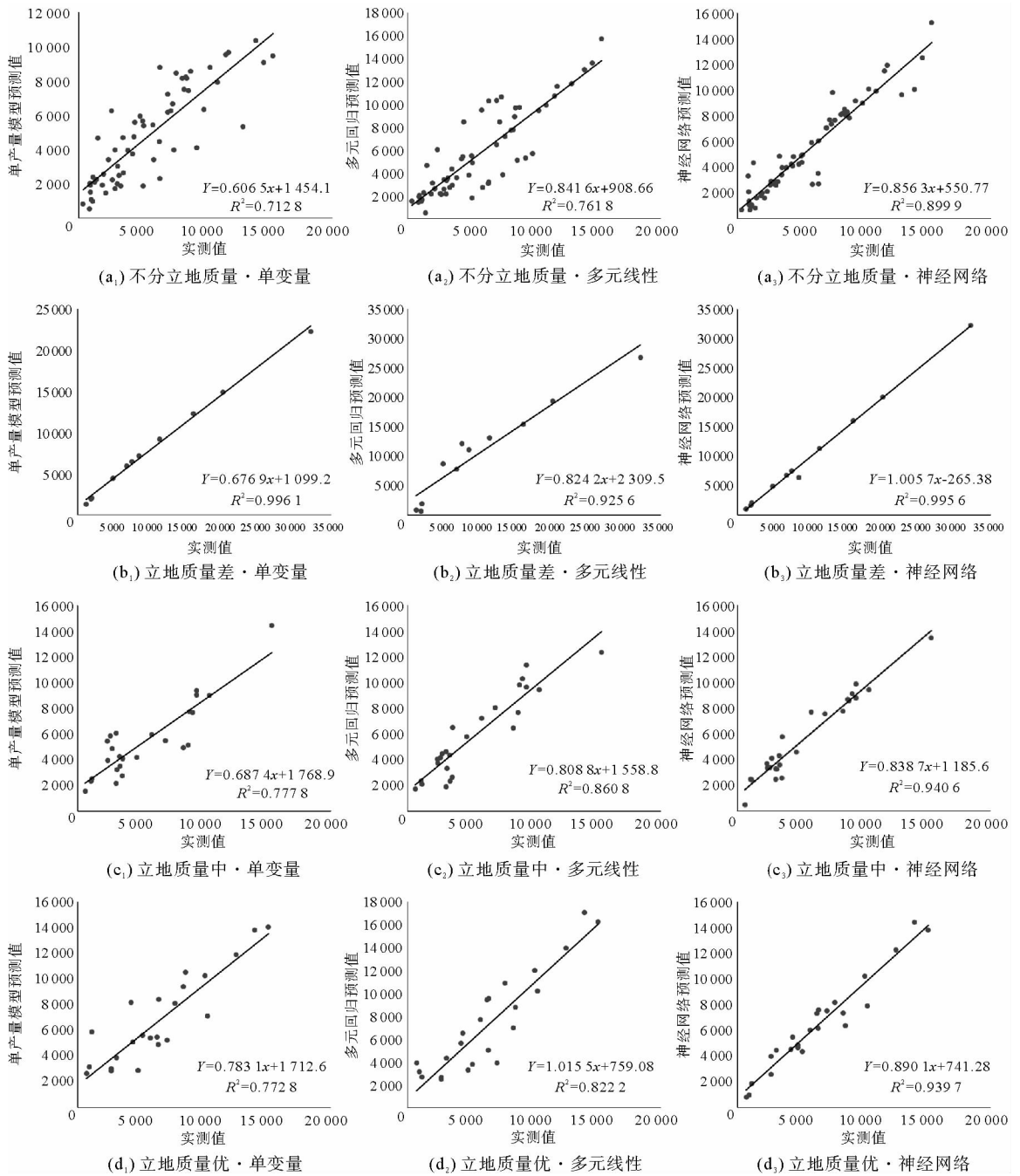


图4 各模型预测值与实测值的精度对比与验证

Fig. 4 Comparison and verification of the accuracy of predicted and measured values of each model

参考文献:

[1]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.  
 LIU Qiongge,PENG Daoli,TU Yunyan. Estimation of forest stock volume based on partial least squares regression[J]. Journal of Central South University of Forestry & Technology,2014,34(2):81-84.

[2]DIAS A T C D,MATTOS E A M,VIEIRA S A,et al. Aboveground biomass stock of native woodland on a Brazilian sandy coastal plain:Estimates based on the dominant tree species[J]. Forest Ecology & Management,2006,226(1):364-367.

[3]TROTTER C M,DYMOND J R,GOULDING C J. Estimation of timber volume in a coniferous plantation forest using Landsat TM[J]. International Journal of Remote Sensing,1997,18(10):2209-2223.

[4]WIJAYA A,KUSNADI S,GLOAGUEN R,et al. Improved strategy for estimating stem volume and forest biomass using

- moderate resolution remote sensing data and GIS[J]. *Journal of Forestry Research*, 2010, 21(1): 1-12.
- [5] 刘俊, 孟雪, 温小荣, 等. 基于不同立地质量的森林蓄积量遥感估测[J]. *西北林学院学报*, 2016, 31(1): 186-191.  
LIU Jun, MENG Xue, WEN Xiaorong, et al. Remote sensing estimation of forest volume based on different site qualities[J]. *Journal of Northwest Forestry University*, 2016, 31(1): 186-191.
- [6] NELSON R, KRABILL W, MACLEAN G. Determining forest canopy characteristics using airborne laser data[J]. *Remote Sensing of Environment*, 1984, 15(3): 201-212.
- [7] GEMMELL F M. Effects of forest cover, terrain, and scale on timber volume estimation with Thematic Mapper data in a rocky mountain site[J]. *Remote Sensing of Environment*, 1995, 51(2): 291-305.
- [8] FAZAKAS Z, NILSSON M, OLSSON H. Regional forest biomass and wood volume estimation using satellite data and ancillary data[J]. *Agricultural and Forest Meteorology*, 1999, 99: 417-425.
- [9] 程武学, 倪勇, 杨存建, 等. 基于小班对象的森林蓄积量遥感估测模型合理性诊断[J]. *湖北农业科学*, 2011, 50(23): 4945-4948.  
CHENG Wuxue, NI Yong, YANG Cunjian, et al. Rationality diagnosis of remote estimation model of forest volume based on small class objects[J]. *Hubei Agricultural Sciences*, 2011, 50(23): 4945-4948.
- [10] 王臣立, 牛铮, 郭治兴, 等. 基于植被指数和神经网络的热带人工林地上蓄积量遥感估测[J]. *生态环境学报*, 2009, 18(5): 1830-1834.  
WANG Chenli, NIU Zheng, GUO Zhixing, et al. Estimating tropical forest stock volume based on vegetation indices and neural network model[J]. *Ecology and Environmental Sciences*, 2009, 18(5): 1830-1834.
- [11] 许炜敏, 陈友飞, 陈明华, 等. 基于BP神经网络的杉木林蓄积量估测研究[J]. *森林与环境学报*, 2012, 32(4): 310-315.  
XU Weimin, CHEN Youfei, CHEN Minghua, et al. Research on the estimation of Chinese fir volume based on BP neural networks[J]. *Journal of Fujian College of Forestry*, 2012, 32(4): 310-315.
- [12] 吴达胜. 基于多源数据和神经网络模型的森林资源蓄积量动态监测[D]. 杭州: 浙江大学, 2014.
- [13] 林昌庚, 周春国, 林俊钦, 等. 关于地位级表[J]. *林业资源管理*, 1997(5): 31-34.  
LIN Changgeng, ZHOU Chunguo, LIN Junqin, et al. About the status level table[J]. *Forestry Resource Management*, 1997(5): 31-34.
- [14] 周春国, 徐雁南. 地位级表编制的新方法[J]. *华东森林经理*, 1997(2): 11-13.  
ZHOU Chunguo, XU Yannan. A new method for the establishment of status table[J]. *East China Forest Management*, 1997(2): 11-13.
- [15] 黄杏模. 谈最佳立地指数导向曲线的评定[J]. *中南林业调查规划*, 1990(3): 45-47.  
HUANG Xingmo. Evaluation of guidance curve for best site index[J]. *Central South Forest Inventory and Planning*, 1990(3): 45-47.
- [16] 王艳婷, 李崇贵, 郝利军. 用岭估计估测以分类为前提的森林蓄积量[J]. *东北林业大学学报*, 2014, 42(9): 39-42.  
WANG Yanting, LI Chonggui, HAO Lijun. Forest volume estimation on the premise of classification by ridge estimate[J]. *Journal of Northeast Forestry University*, 2014, 42(9): 39-42.
- [17] 刘俊, 孟雪, 温小荣, 等. 基于不同立地质量的松树林蓄积量遥感估测[J]. *林业资源管理*, 2015(5): 55-60.  
LIU Jun, MENG Xue, WEN Xiaorong, et al. Remote sensing estimation of pine volume based on different site quality[J]. *Forestry Resource Management*, 2015(5): 55-60.
- [18] GAO J. Quantification of grassland properties; How it can benefit from geoinformatic technologies[J]. *International Journal of Remote Sensing*, 2006, 27(7): 1351-1365.
- [19] 柯宗建, 张培群, 董文杰, 等. 最优子集回归方法在季节气候预测中的应用[J]. *大气科学*, 2009, 33(5): 994-1002.  
KE Zongjian, ZHANG Peiqun, DONG Wenjie, et al. Application of optimal subset regression method in seasonal climate prediction[J]. *Chinese Journal of Atmospheric Sciences*, 2009, 33(5): 994-1002.
- [20] MALLOWS C L. Some comments on CP[J]. *Technometrics*, 1973, 15(4): 661-675.
- [21] 何芳, 王小川, 肖森予, 等. 基于MIV-BP型网络实验的房地产项目风险识别研究[J]. *运筹与管理*, 2013, 22(2): 229-234.  
HE Fang, WANG Xiaochuan, XIAO Senyu, et al. Research on risk recognition of real estate projects based on MIV-BP Neural network test[J]. *Operations Research and Management Science*, 2013, 22(2): 229-234.