

文章编号: 2095-2163(2019)02-0067-06

中图分类号: S771.8

文献标志码: A

基于高分遥感数据的森林郁闭度估测方法研究

杨妍婷

(浙江农林大学 信息工程学院, 浙江 临安 311300)

摘要: 森林郁闭度是研究森林生态系统和了解森林资源状况的重要参数,而传统的实地测量方法效率较低下,且仅能获取小范围的一些具有代表性的数据,不利于研究大范围或区域内郁闭度的空间分布及变化。为了估算森林郁闭度,分析其与遥感影像因子之间的相关性,本文以河北省滦平县巴克什营镇和长山峪镇为研究区域,采用分辨率较高的高分一号(GF-1)数据,结合 SRTM DEM 数据的地形因子,对该地区的森林郁闭度进行反演。本文在系统整理分析和评价国内外森林郁闭度相关研究文献的基础上,选择了红波段、近红外波段、亮度、绿度、黄度等 14 个因子作为自变量参与构建多元逐步回归(Multi-variable Stepwise Regression, MSR)、随机森林(Random Forest, RF)和 Cubist 三种模型,对该地郁闭度进行估测。实验结果表明,基于机器学习的随机森林和 Cubist 算法结果要优于传统的多元逐步回归算法,各项评价指标显示其中 Cubist 回归算法在该研究区的拟合效果最好。多元逐步回归(MSR)算法成熟简单,应用广泛,但模型不稳定,反演精度不高,不适用于大区域的郁闭度估算;随机森林(RF)处理大数据速度快,但高值低估和低值高估的情况比较严重,增大了郁闭度估测误差;Cubist 在预测连续值方面很成功,使用最近邻样本来调整规则预测结果,模型较稳定,能够得到较为准确的预测数值,不过需要花费很长时间进行计算。

关键词: 森林郁闭度; GF-1 号影像; 图像融合; 机器学习; 统计回归; 参数反演

Estimation method of forest canopy density based on high resolution remote sensing data

YANG Yanting

(College of Information Engineering, Zhejiang A&F University, Lin'an Zhengjiang 311300, China)

[Abstract] Forest canopy density is an important parameter for studying forest ecosystem and understanding forest resources. Traditional field measurement methods are inefficient and can only obtain some representative data in a small range, which is not conducive to the study of spatial distribution and change of forest canopy density in a large area or region. In order to estimate forest canopy density and analyze its correlation with remote sensing image factors, this paper takes Bakeshiying Town and Changshanyu Town in Luanping County of Hebei Province as the research area, and uses high resolution GF-1 data, combined with SRTM DEM data topographic factors, to invert forest canopy density in this area. On the basis of systematically analyzing and evaluating the related literatures of forest canopy density at home and abroad, 14 factors, such as red band, near infrared band, brightness, greenness and yellowness, are selected as independent variables to participate in the construction of three models: multi-variable Stepwise Regression (MSR), Random Forest (RF) and Cubist, and the canopy density is improved. The experimental results show that the results of random forest and Cubist algorithm based on machine learning are better than those of traditional multiple stepwise regression algorithm. The evaluation indexes show that Cubist regression algorithm has the best fitting effect in this research area. Multivariate stepwise regression (MSR) algorithm is mature and simple, and widely used, but the model is unstable, the inversion accuracy is not high, and it is not suitable for estimating canopy density in large areas; Random Forest (RF) can process large data quickly, but the situation of overestimation and underestimation is serious, which increases the estimation error of canopy density; Cubist is very successful in predicting continuous values, and uses the nearest neighbor sample to adjust the rules. The prediction results show that the model is stable and can get more accurate prediction values, but it takes a long time to calculate.

[Key words] forest canopy density; GF-1 image; image fusion; machine learning; statistical regression; parameter inversion

0 引言

森林郁闭度就是研究森林资源环境的重要参数之一,即等于垂直方向上林冠的投影面积和林地面积之间比值^[1-2]。过去测度此指标最常用的方法主要有:目测法^[3]、样线法^[4]、样点法^[5]、树冠投影

法^[6]、观测管法^[7]、照片法^[8]等。传统方法都需要耗费较长时间,并且只能得到小范围内典型数据,所以很难满足大区域范围测度需求。

最近数十年内,遥感技术急速发展,而且已逐步应用到林业领域。利用遥感技术计算森林郁闭度的基础数据就是遥感光学影像、实测地表样地数

作者简介: 杨妍婷(1988-),女,硕士研究生,主要研究方向:资源与环境遥感。

收稿日期: 2018-12-10

据^[9-11]。常用遥感反演方法分别是辐射传输模型法和统计模型法,前者模型结构复杂、所需要的参数多并且大都难以准确获取,进而影响到森林郁闭度的提取精度;对于后者而言,较为成熟的方法包括回归模型、以植被指数为基础的混合像元分解法。

为了实现对森林区域的基于高分辨率遥感图像的郁闭度反演,本研究采用小块区域精度较高的回归模型法进行试验,并比较不同回归模型的反演效果,再对其进行分析评估。

1 研究区概况及数据介绍

1.1 研究区概况

滦平县地属河北省承德市,植被覆盖率高。本文选取的研究区域为滦平县南部地区两镇,包括巴克什营镇和长山峪镇。全县有 58% 的森林覆盖率。其中,南部的巴克什营和长山峪镇的林地占全县植被的一大部分,经纬度范围为: $117^{\circ}3'E \sim 117^{\circ}28'E$, $40^{\circ}35'N \sim 40^{\circ}55'N$ 。研究区域的地理位置及其真彩色图像如图 1 所示。

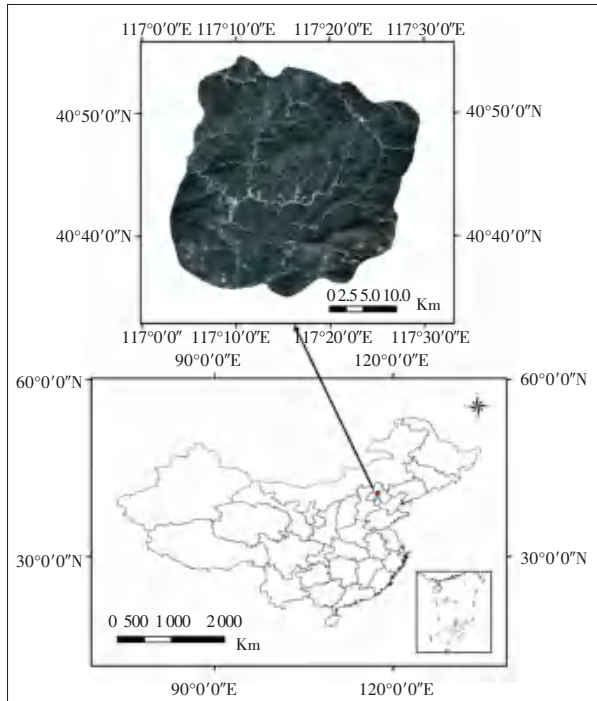


图 1 研究区域地理位置图

Fig. 1 Study of regional geographic location maps

1.2 数据介绍

GF-1 是国内首个对地进行高分辨率观测的卫星,主要优势是综合使用了分辨率较高的光学遥感、多载荷图像融合拼接等先进技术。本实验采用 GF-1 中 PMS 传感器数据。4 个可见光近红外波段和 1 个全色波段。结合实测采样日期,本研究采用 2018

年 9 月 16 日的数据,该数据云量少,质量较高。

SRTM 提供了比较完整的高分辨率 DEM。本研究运用的主要数据是 SRTM DEM 数据,90 m 空间分辨率。将 DEM 数据采用克里金法空间插值到 2 m,采用 ARCGIS 软件对其进行计算,得到坡度和坡向。然后将其与可见光近红外数据进行几何配准。

外业森林郁闭度数据于 2018 年 9 月河北省承德市滦平县南部两镇实测得到,数据表格包括经纬度、郁闭度等。本次实验采用照片法估测研究区域的郁闭度,大约每隔 1 Km 采样一个点,最终采样 308 个点的郁闭度数据。

2 数据处理

2.1 GF-1 数据预处理

本实验采用 ENVI 软件相关功能对多波段和全色波段数据进行辐射定标、大气校正和正射校正,消除传感器和地形因素等引起的几何误差。图像的预处理结果如图 2 所示。

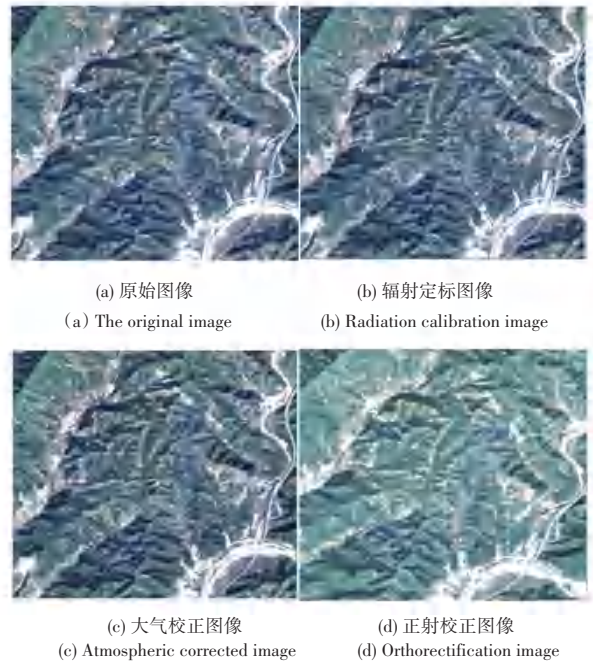


图 2 研究区多波段数据预处理结果

Fig. 2 Preprocessing results of multi-band data in the study area

2.2 图像融合

为了得到精度更高的多波段图像,考虑将 8 m 的多波段和 2 m 的全色波段进行图像融合,采用常见的图像融合算法进行试验^[12]。本研究对比了 4 种融合算法的结果,即: PC 融合、Brovey 融合、NNDiffuse Pan Sharpening 融合和 Gram-Schmidt 融合。相比之下 NNDiffuse Pan Sharpening 融合结果色彩精确,纹理清晰。不同算法的融合效果如图 3 所

示。

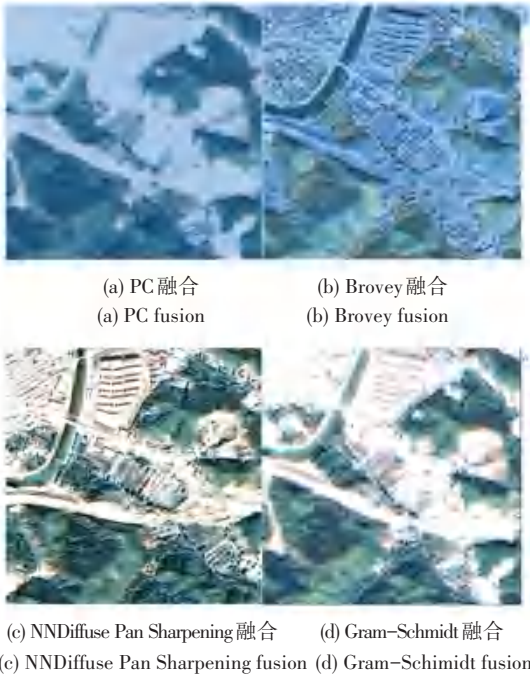


图 3 不同算法图像融合结果

Fig. 3 Image fusion results of different algorithms

灰度值均值能够体现人眼视觉观察到的图像平均亮度；标准差能够体现这一数值的离散情况。本文主要选取均值、方差和均方根误差来评价。选取植被敏感的绿波段进行定量评价，见表 1。

表 1 不同融合算法评价指标对比

Tab. 1 Comparison of evaluation indicators of different fusion algorithms

评价指标	PC	Brovey	NNDiffuse Pan Sharpening	Gram-Schmidt
均值	0.054	0.124	0.034	0.042
方差	0.148	16.180	0.132	0.163
均方根误差	2.115	20.021	1.389	2.006

结合目视和定量评价结果，本文选取 NNDiffuse Pan Sharpening 融合处理数据并进行接下来的研究。

2.3 监督分类提取林分信息

对于融合后的图像，采用滦平县南部地区巴克什营和长山峪两镇矢量边界对其进行裁剪，然后采用随机森林监督分类方法提取植被。使用 ENVI 软件对待研究区进行分类，将研究区分为水体（蓝）、植被（绿）、不透水面（红）、裸土（黄），主要提取植被区域进行森林郁闭度估算。分类结果如图 4 所示。

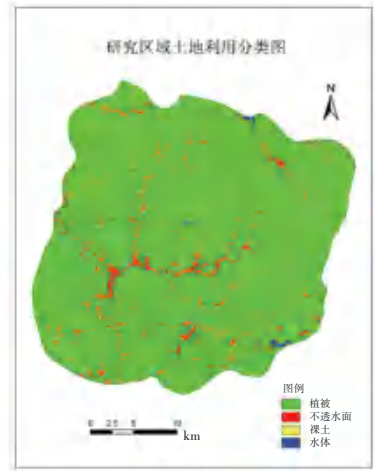


图 4 随机森林监督分类结果

Fig. 4 Random forest supervised classification results

2.4 影响因子计算

2.4.1 植被指数

由于 GF-1 数据波段有限，本研究采用常见的 4 种植被指数，即 $NDVI$ 、 DVI 、 RVI 和 $SAVI$ 进行分析。上述植被指数的计算公式如下：

$$NDVI = \frac{b_4 - b_3}{b_4 + b_3}; \tag{1}$$

$$DVI = b_4 - b_3; \tag{2}$$

$$RVI = b_4/b_3; \tag{3}$$

$$SAVI = \frac{(b_4 - b_3) * (1 + L)}{(b_4 + b_3 + L)}. \tag{4}$$

其中， b_3 和 b_4 分别是 GF-1 数据的第 3、4 波段反射率，即红波段和近红外波段， L 为调节系数，本研究取 $L = 0.5$ 。不同的植被指数计算结果如图 5 所示。

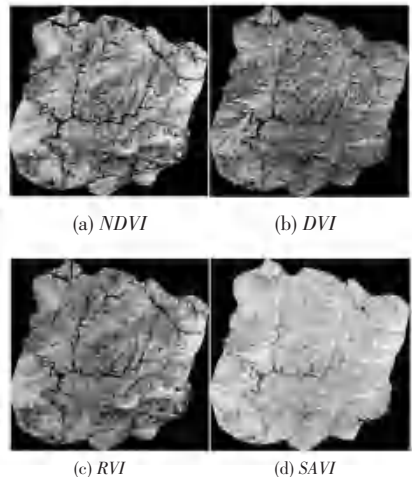


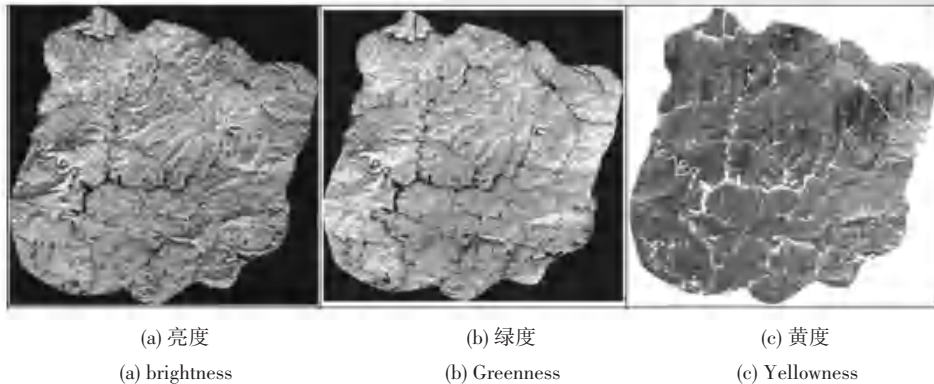
图 5 植被指数计算结果

Fig. 5 Calculation results of vegetation index

2.4.2 缨帽变换因子

缨帽变换 (K-T 变换) 是一种经验线性正交变

换,本研究采用亮度、绿度、黄度三个分量参与分析, 其计算结果如图6所示。



注:由于该地区绿色植被较多,黄度多为负值,忽略背景值为0,按亮度排序,背景为白色。

图6 缨帽变换分量计算结果

Fig. 6 Computational results of transformation components of reel cap

2.4.3 地形因子

高坡处的地面回波会与部分林分冠层回波杂糅,并且海拔也会影响植被的生长状态,诸如此类地

形因子均会影响森林郁闭度的计算^[13]。根据已有的DEM数据,将高程、坡度和坡向作为影响因子纳入回归模型,地形因子计算结果如图7所示。

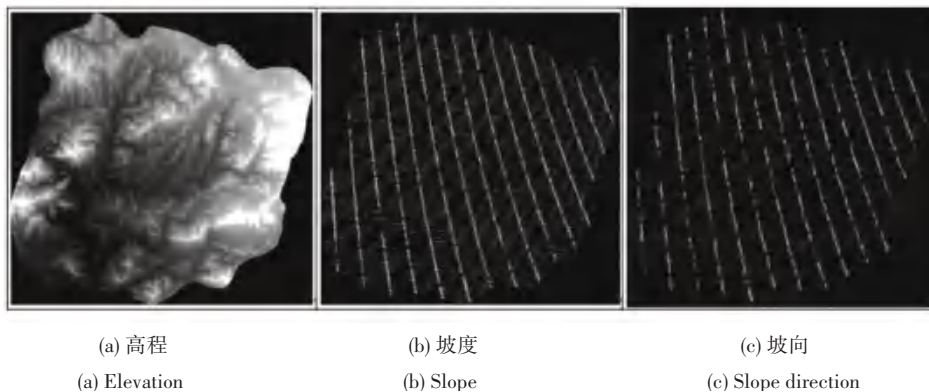


图7 地形因子计算结果

Fig. 7 Calculation results of topographic factors

3 森林郁闭度反演

根据上述计算的10个影响因子,加上原始数据的4个波段,通过相关分析从中找出存在显著相关的因子,再运用多种回归方法,分析获得郁闭度与各因子之间回归关系,并建模。

由于本研究采用的回归算法涉及传统的最小二乘回归和基于机器学习的回归,对于测试集和训练集的划分,需要统一的标准。本文采用传统的划分方法,即70%的样本用于训练集,30%的样本用于测试集。采用python的sklearn进行样本划分,最终得到215个样本用作训练集,83个样本用于验证。

3.1 自变量的选择

自变量是建立回归模型的基础。通常来讲,筛选剔除的变量主要有3种,对此分述如下。

(1)能够被其它变量代替的变量,如果都被导入模型,会导致回归反演过于复杂,却不能显著提高精度水平。

(2)变量和因变量的相关性不高,甚至完全没有相关性,此类变量对于建模分析没有任何证明影响,还会导致运算工作量的增加。

(3)整体样本因变量和自变量都没有显著相关性,部分规律较强的自变量出现的规律模型将主导建模结果,甚至导致固有规律直接影响模型。

3.2 回归建模

3.2.1 多元逐步回归

在利用SPSS进行多元逐步回归中,参数设置包括置信水平为95%,使用 f 的概率,进入为0.05,删除为0.1,对基于14种遥感因子共同引入模型进行自变量选择,最终选择了2种自变量参与模型构

建,得到最优模型。可将其解析为如下数学公式:

$$Y = 0.670 - 0.007 15 X_3 + 0.001 82 X_8. \quad (5)$$

由式(5)可知,逐步回归后留下来的自变量只 *SAVI* 和 b_3 两个,自变量数量较少。

3.2.2 随机森林回归

此方法的过程如下:

Step1 进行 bootstrap 抽样从训练原始集内选出样本集(内有样本 k 个),各样本集的样本容量都和训练原始集相同。

Step2 对样本集(k 个)构建各自的决策树模型(k 个),同时根据各树模型预测记录,得到预测值(k 组)。

Step3 对所得全部预测值求均值,得到预测最终结果^[14]。

3.2.3 Cubist 回归

此方法通常可用于连续值预测问题中。模型树代表的是某个分段线性函数,模型树内部各节点测定某些自变量,各叶节点将分析得到模型,并且所有模型均为以前期测试节点所得为自变量^[15]。此过程主要通过递归处理得到模型树,起点是全部完整训练样本集。在各层都需要选取辨别度最高的自变量为子树根,抵达此处样本都需要结合其自变量数值大小,被进一步划分为多个子集。分割节点的标准在于能够促使样本标准差下降,同时不影响平方误差。主要运用 Variance,也就是诱导来实现启发,叶节点内都必须有常数填充建立模型。

3.3 结果讨论

3.3.1 模型评价

本研究采用决定系数 R^2 、估计值的标准差 SEE 、总相对误差 TRE 和平均系统误差 MSE 进行模型评价。研究中将用到的数学公式可表示如下:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}; \quad (6)$$

$$SEE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - p}}; \quad (7)$$

$$TRE = \frac{\sum (y_i - \hat{y}_i)}{\sum \hat{y}_i} * 100\%; \quad (8)$$

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i) / \hat{y}_i * 100\%. \quad (9)$$

其中, y_i 和 \hat{y}_i 分别是样地 i 的郁闭度实测值和预测值; \bar{y}_i 是所有实测值的平均数; p 为模型参数个数; n 为样本数量。

3.3.2 结果分析

将剩余的验证样本数据(83 个)导入 3 种估算分析模型,估测郁闭度结果,并评价结果的精度水平。

表 2 回归模型比较

Tab. 2 Comparison of regression models

模型	R^2	SEE	$TRE/\%$	$MSE/\%$
逐步回归	0.753	0.289	3.25	3.59
随机森林回归	0.803	0.235	3.10	3.21
Cubist 回归	0.824	0.256	2.98	3.07

由表 2 可见,3 种模型对郁闭度的反演精度验证结果,综合建模精度来看,从多元逐步回归(MSR)模型、随机森林(RF)模型到 Cubist 模型精度中,Cubist 模型比较稳定,精度最高,为该研究区域郁闭度反演的最优模型,其预测能力最强,反演效果最好,模型验证精度 $R^2 = 0.824$, $SEE = 0.256$, $TRE(\%) = 2.98$, $MSE(\%) = 3.07$, 相对比较稳定。不过,不同模型进行分析统计所得指标之间对比并没有发现特别明显的差异。从 4 项指标分析,随机森林回归略逊于 Cubist 回归。

在森调尤其是遥感分析研究森林资源过程中,建模预估郁闭度是非常重要的工作步骤。因为遥感数据多而且非常复杂,所以无法明确获知可能对预测结果造成关键影响的变量集合,如果选取太多变量又必然会对建模计算带来困难,不但会导致模型不够稳定,还可能导致模型无法用于实际。故而,研究选择出最佳模型是对郁闭度进行精准预测的核心所在。

3.4 制图输出

对于以上不同模型的精度评价和结果讨论,本研究采用精度最高的 Cubist 回归模型反演的森林郁闭度结果进行制图输出,如图 8 所示。

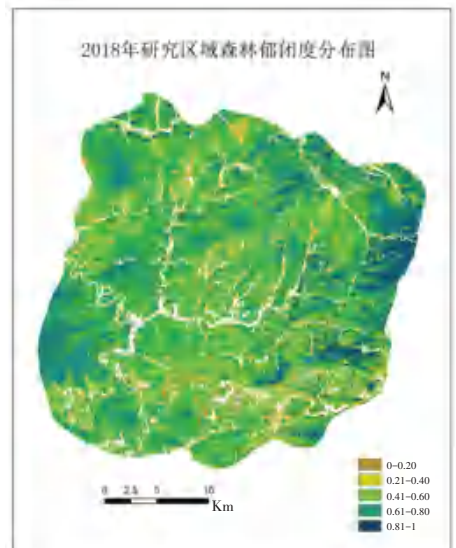


图 8 2018 年研究区域森林郁闭度分布图

Fig. 8 Distribution map of forest canopy density in the study area in 2018

从图 8 中可以看出,在不透水面和水体周围,森

林郁闭度很低;东、西部有较大高程的山区内这一指标数值较大,这一现象与实际情况相符。

4 结束语

与传统郁闭度测定方法相比,利用遥感估算模型能够更有效、更准确地获取大范围或区域的森林郁闭度,有利于降低森林资源调查成本,并提高管理工作水平。森林郁闭度的遥感估算精度与模型的好坏有关,同时也与参与建模反演的影响因子密不可分。此处将给出研究结论,具体阐述如下。

(1)以河北省滦平县南部两镇为研究区域,采用高分一号(GF-1)遥感数据融合 SRTM DEM 数据,并整合实测数据,对该地区的森林郁闭度进行反演,取得了较好的效果。

(2)采用4种常用的融合算法,包括PC融合、Brovey融合、NNDiffuse Pan Sharpening融合和Gram-Schmidt融合,对预处理之后的多波段和全色波段进行融合,从目视和定量两个角度比较不同方法的融合效果,结果表明NNDiffuse Pan Sharpening融合效果最好。

(3)采用3种统计回归模型对森林郁闭度进行估算,包括:逐步回归、随机森林回归和Cubist回归。结果表明,基于机器学习的随机森林和Cubist算法结果要优于传统的回归算法,各项评价指标总体显示Cubist回归算法在该研究区的拟合效果最好。

(4)最终的反演制图结果表明基于Cubist算法的郁闭度遥感反演整体精度较高,但也存在低值高估和高值低估的现象。

参考文献

- [1] 龚韡,吴聪连,陈奕浪. 信息技术在林业管理中的应用[J]. 南方农业,2017,11(7):35-36.
[2] 孙华. SPOT5 在森林资源调查中的应用研究[D]. 长沙:中南林

业科技大学,2006.

- [3] 赵国帅. 无人机遥感在林业中的应用与需求分析[J]. 福建林业科技,2017,44(1):136-140.
[4] 车腾腾,冯益明,吴春争. “3S”技术在精准林业中的应用[J]. 绿色科技,2010(10):158-162.
[5] 柏艺. 基于3种不同数据源的川西南常绿阔叶林郁闭度估算模型研究[D]. 雅安:四川农业大学,2016.
[6] 邹杰,诸葛祥东. 森林郁闭度及其测定方法[J]. 黑龙江科技信息,2011(35):290.
[7] 李永宁,张宾兰,秦淑英,等. 郁闭度及其测定方法研究与应用[J]. 世界林业研究,2008,21(1):40-46.
[8] 杜文峰,王凤臻,李庆. 提高郁闭度调查精度的几点建议[J]. 林业资源管理,1999(3):62-64.
[9] 王强,斯琴毕力格,肖海涛,等. 森林郁闭度测定方法探讨与展望[J]. 内蒙古林业调查设计,2014,37(1):82-83,108.
[10] 覃先林,陈尔学,李增元,等. 基于MODIS数据的森林覆盖变化监测方法研究[J]. 遥感技术与应用,2006,21(3):178-183.
[11] 吴胜义,赵强国,石昊楠. SPOT5 遥感影像在林业二类调查中的应用[J]. 福建林业科技,2011,38(3):25-31.
[12] 陈尔学. 合成孔径雷达森林生物量估测研究进展[J]. 世界林业研究,1999,9(6):18-23.
[13] NELSON R, KRABILL W, MACLEAN G. Determining forest canopy characteristics using airborne laser data [J]. Remote Sensing of Environment,1984,15(3):201-212.
[14] 樊晶,杨燕琼. 基于遥感的森林健康度分析—以东莞桉树林为例[J]. 林业与环境科学,2017,33(1):40-45.
[15] 庞勇,李增元. 基于机载激光雷达的小兴安岭温带森林组分生物量反演[J]. 植物生态学报,2012,36(10):1095-1105.
[16] 黄燕平,陈劲松. 基于SAR数据的森林生物量估测研究进展[J]. 国土资源遥感,2013,25(3):7-13.
[17] NELSON R, KRABILL W, TONELLI J. Estimating forest biomass and volume using airborne laser data[J]. Remote Sensing of Environment,1988,24(2):247-267.
[18] WULDER M A, NORMAN P, WITTE C. Special issue: Li DAR remote sensing of forest structure and terrain[J]. Can. J. For. Res.,2003,29(15):527-678.
[19] SUN G Q, RANSON K J, KIMES D S, et al. Forest vertical structure from glas: An evaluation using LVIS and SRTM data [J]. Remote Sensing of Environment,2008,112(1):107-117.
[20] HUDAK A T, EVANS J S, SMITH A M S. Li DAR utility for natural resource managers [J]. Remote Sensing,2009,1(4):934-951.

(上接第66页)

- [18] 阮松,孙开琼,覃磊,等. 基于改进的快速最大熵多阈值的成熟草莓图像分割[J]. 东北农业科学,2016,41(1):107-112.
[19] 邹小林,陈伟福,冯国灿,等. 基于谱聚类的多阈值图像分割方法[J]. 计算机科学,2012,39(3):246-248,259.
[20] PAL N R, PAL S K. A review on image segmentation techniques

[J]. Pattern Recognition,1993,26(9):1277-1294.

- [21] VTIALLIAMMAL N, GEETHALAKSHMI S N. A novel approach for plant leaf image segmentation using fuzzy clustering [J]. International Journal of Computer Application,2012,44(13):10-20.