

# 基于随机森林模型的林地叶面积指数遥感估算

姚雄<sup>1,2</sup> 余坤勇<sup>1,2</sup> 杨玉洁<sup>2</sup> 曾琪<sup>2</sup> 陈樟昊<sup>2</sup> 刘健<sup>1,2</sup>

(1. 福建农林大学3S技术与资源优化利用福建省高校重点实验室,福州350002;

2. 福建农林大学林学院,福州350002)

**摘要:** 林地叶面积指数(Leaf area index, LAI)的准确估测是精准林业的重要体现。为了快速、准确、无损监测林地LAI,利用LAI-2200型植物冠层分析仪获取福建省西部森林样地的LAI数据,结合同期Pleiades卫星影像计算12种遥感植被指数,分析了各样地实测LAI数据和相应植被指数的相关性,进而使用随机森林(RF)算法构建了林地LAI估算模型,以支持向量回归(SVR)模型和反向传播神经网络(BP)模型作为参比模型,以决定系数( $R^2$ )、均方根误差(RMSE)、平均相对误差(MAE)和相对分析误差(RPD)为指标评价并比较了模型预测精度。结果表明:全样本数据中,各植被指数与对应LAI值均呈极显著相关( $P < 0.01$ ),且相关系数都大于0.4;RF模型在3次不同样本组中的预测精度均高于同期的SVR模型和BP模型;3个样本组中RF模型的LAI估测值与实测值的 $R^2$ 分别为0.688、0.796和0.707,RPD分别为1.653、1.984和1.731,均高于同期SVR模型和BP模型,对应的RMSE分别为0.509、0.658和0.696,MAE分别为0.417、0.414和0.466,均低于同期其他2种模型。

**关键词:** 林地; 叶面积指数; 遥感反演; 随机森林模型; 支持向量回归模型; 反向传播神经网络模型

**中图分类号:** TP79; S127 **文献标识码:** A **文章编号:** 1000-1298(2017)05-0159-08

## Estimation of Forest Leaf Area Index Based on Random Forest Model and Remote Sensing Data

YAO Xiong<sup>1,2</sup> YU Kunyong<sup>1,2</sup> YANG Yujie<sup>2</sup> ZENG Qi<sup>2</sup> CHEN Zhanghao<sup>2</sup> LIU Jian<sup>1,2</sup>

(1. University Key Laboratory for Geomatics Technology and Optimize Resources Utilization in Fujian Province, Fujian Agriculture and Forestry University, Fuzhou 350002, China

2. College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

**Abstract:** Accurate estimation of forest leaf area index (LAI), which is defined as half the total area of green leaves per unit ground surface area, is the important embodiment of precision forestry. In order to monitor forest LAI faster, more accurate and non-destructively, LAI-2200 plant canopy analyzer was used to acquire LAI data from the forest plots in western Fujian. Totally 12 kinds of vegetation index based on the Pleiades satellite images in the same period were calculated and the correlation between measured LAI and the vegetation index was analyzed. The purpose was to construct LAI estimation model specifically by using random forest algorithm (RF). Additionally for each sample group, the models based on support vector regression model (SVR) and back-propagation neural network model (BP) were employed as comparison models. The estimation accuracy of the three models for each sample group was compared based on determination coefficients ( $R^2$ ), root mean square errors (RMSE), mean relative errors (MAE) and relative percent deviation (RPD). The results indicated that the vegetation indices and LAI values were significantly correlated ( $P < 0.01$ ), and the correlation coefficients were greater than 0.4 for all sample data. The forecast accuracy of RF model in three different sample groups was higher than those of the SVR and BP models in the same period.  $R^2$  of LAI estimated and measured values in the three sample groups based on RF model were 0.688, 0.796 and 0.707, respectively. RPD were 1.653, 1.984 and 1.731, respectively. These data were all higher than those of SVR model and BP model, and RF model showed a higher accuracy than the other two models (RMSE of RF model were

收稿日期: 2016-08-08 修回日期: 2016-10-21

基金项目: 国家自然科学基金项目(41401385)

作者简介: 姚雄(1990—),男,博士生,主要从事3S技术应用研究,E-mail: 424532024@qq.com

通信作者: 刘健(1963—),男,教授,博士生导师,主要从事森林经营管理与3S技术应用研究,E-mail: fljujian@126.com

0.509, 0.658 and 0.696, respectively; MAE were 0.417, 0.414 and 0.466, respectively). These results would be helpful for improving the forest LAI remote sensing estimation accuracy.

**Key words:** forest; leaf area index; remote sensing inversion; random forest model; support vector regression model; back-propagation neural network model

## 引言

叶面积指数(Leaf area index, LAI)指单位水平地面面积上所有绿叶面积总和的一半<sup>[1]</sup>,是表征森林冠层结构的重要参数,与森林植被的光合、蒸腾、降水截留、碳氮循环等生态过程密切相关,同时 LAI 也是水文、生态、气候等模型的重要输入参数<sup>[2-4]</sup>。因此,快速和准确地估测林地 LAI,对森林生态系统乃至全球碳循环的研究都具有重要意义。

目前,林地 LAI 的测量方法可以分为直接和间接测量两大类。直接测量法技术成熟,结果最为准确,但由于叶片样品需破坏性采集,费时费力,而且叶片样品不一定具有代表性,因此该方法仅适用于小范围区域<sup>[5]</sup>。而耦合遥感数据与地面光学仪器实测数据的林地 LAI 间接测量法具有破坏性小或无破坏性、测量迅速简便、经济高效等特点,是快速估测大范围林地 LAI 的唯一途径<sup>[6-7]</sup>。遥感估算 LAI 的方法大致可分为经验统计模型、物理模型、机器学习模型。经验统计模型通过光谱数据或表征光谱属性的光谱指数与地面实测 LAI 之间的统计关系来估算 LAI<sup>[8-9]</sup>,模型计算简单且易于实现,但模型的普适性较差。物理模型具有很强的物理理论基础,不因背景环境或植被类型的变化而变化,因此该类模型具有较好的普适性<sup>[10-11]</sup>,但模型涉及参数较多,存在解的不唯一性,而且模型运算时间过长。机器学习模型通过样本数据的不断训练以使目标最优化,它具有很强的非线性拟合能力<sup>[12-13]</sup>,是估算林地 LAI 的重要手段。

已有众多学者采用机器学习模型进行 LAI 的遥感估算:王丽爱等<sup>[14]</sup>基于中国自主研发的环境与灾害监测预报卫星 HJ-CCD 数据,采用多核支持向量回归算法建立了小麦不同生育期 LAI 的遥感估算模型;王修信等<sup>[15]</sup>使用 Landsat 数据,构建了基于最优多植被指数组合的神经网络模型反演复杂地形水源林 LAI 的空间变化;韩兆迎等<sup>[16]</sup>分别采用随机森林算法和支持向量机算法构建了苹果树冠层反射率与实测 LAI 之间关系,通过模型比对检验,指出随机森林回归模型的预测精度高于支持向量机。尽管这些学者利用机器学习模型估测 LAI 的研究都取得了相应的成果,但迄今为止,国内外关于随机森林模型在 LAI 遥感监测预报方面的应用还鲜有报道<sup>[16]</sup>,尚无

该模型在林地 LAI 遥感估测中的相关研究。为此,本研究选择福建省长汀县河田镇为研究区域,使用随机森林模型构建基于遥感植被指数的林地 LAI 遥感估算模型,并与支持向量机模型和神经网络模型进行反演精度比较,判断随机森林模型在林地 LAI 遥感估测中的适用性。

## 1 材料与方方法

### 1.1 研究区概况

福建省长汀县河田镇位于福建西部,地处武夷山脉南段(25°33'~25°48'N、116°18'~116°31'E),属中亚热带季风气候。境内气候温和,水热资源丰富,年平均气温 17.5~18.8℃,极端最高气温 39.8℃,极端最低气温 -4.9℃。全年光照充足,雨量充沛,降雨集中在夏季,年平均降雨量 1700 mm,无霜期 265 d。全镇低山高丘环绕四周,中部开阔,土壤类型以红壤为主,是南方典型的红壤丘陵区。镇域面积 296 km<sup>2</sup>,其中山地面积 213 km<sup>2</sup>,山地植被主要包括马尾松(*Pinus massoniana* Lamb.)、黄瑞木(*Adinandra millettii*)、胡枝子(*Lespedeza bicolor* Turcz.)和芒萁(*Dicranopteris dichotoma* Bernh.)。由于历史原因,研究区山地天然植被遭到严重毁坏,水土流失连年加剧,成为南方典型的红壤水土流失严重区(图 1)。

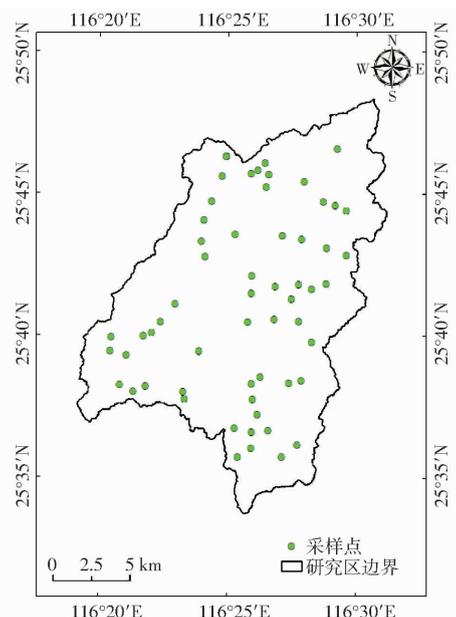


图 1 研究区地理位置图

Fig. 1 Geographic location of study area

### 1.2 地面数据采集

2015 年 1 月份在研究区开展地面数据采集工作, 以优劣兼顾为原则, 在河田镇布设 59 个 25.82 m × 25.82 m 的标准样地(包括 32 个马尾松林地、17 个杉木林地和 10 个阔叶林地), 样地面积为 667 m<sup>2</sup>, 样地间距离大于 500 m, 为了避免居民点、道路等地物对调查数据造成干扰, 各样地均位于此类地物 50 m 之外(图 1)。样地坐标通过两部麦哲伦海王星 400 手持 GPS 获取, 当两部 GPS 定位结果基本一致时记

录坐标数据。

LAI 采用 LAI-2200 型植物冠层分析仪(Li-COR Inc, Lincoln, NE, USA) 进行测量。测量时, 首先在样地外空旷处测量冠上 A 值, 再进入样地内测量 4 次冠下 B 值, 即采用 ABBBB 测量方式; 测量结束后, 仪器自动记录 LAI 值。每个样地重复测量 3 次, 取其平均值作为该样地的 LAI 值, 共获取 59 组 LAI 数据。LAI 采样数据的描述统计见表 1。

表 1 采样点 LAI 的描述统计

Tab. 1 Results of LAI for descriptive statistics

类型	样点数	最小值	最大值	平均值	标准差	偏度	峰度	变异系数/%
杉木	17	1.26	5.68	2.96	1.01	1.00	2.20	34.12
马尾松	32	0.78	3.24	1.56	0.51	1.10	2.33	32.69
阔叶树	10	2.63	6.28	4.00	1.41	0.59	-1.37	35.25
全样本	59	0.78	6.28	2.37	1.29	1.30	1.47	54.43

### 1.3 遥感数据获取及预处理

本研究使用的遥感数据在河田镇的过境时间为 2014 年 12 月 10 日, 景号分别为 0719-04222、0519-03996 的法国 Pleiades 遥感数据, 其多光谱波段空间分辨率为 2 m, 包括 4 个波段: 蓝光波段(430~550 nm)、绿光波段(490~610 nm)、红光波段(600~720 nm) 和近红外波段(750~950 nm)。

遥感数据预处理包括大气校正和几何校正, 首先使用 ENVI 5.1 软件中的 FLAASH 大气校正模块进行影像大气校正, 然后采用三阶多项式变换模型对大气校正后的图像进行几何校正, 即以河田镇 1:10 000 地形图为校正基准图, 以 Pleiades 遥感数据为匹配图, 采用匹配图对基准图采点的方法进行几何校正, 校正采用 Krasovsky 椭球体坐标和 Gauss Kruger 投影。

### 1.4 植被指数选取

遥感影像数据上的植被信息主要通过绿色植被

叶片和冠层光谱信息的差异来体现。植被在不同波段范围所体现的信息与植被的某种特征状态密切相关, 但是植被的一些信息(如生物量、植被长势等), 难以用单波段或多个单波段数据直接提取, 往往需选用多个光谱波段经简单地分析运算, 产生可以表征或指示植被长势、生物量等的数值, 用一种简单有效的形式(即植被指数)来实现对植被不同状态信息的表达, 因而植被指数是估算植被 LAI 的重要工具<sup>[17]</sup>。本研究基于 Pleiades 遥感数据, 借鉴前人研究, 在众多植被指数中选取 12 个对 LAI 敏感且能较好估算 LAI 的宽波段植被指数<sup>[9,18-19]</sup>, 各植被指数表达式见表 2。 $B_1$ 、 $B_2$ 、 $B_3$ 、 $B_4$  分别为蓝光、绿光、红光、近红外波段光谱反射率。

### 1.5 模型构建方法

在众多神经网络模型中, 反向传播(Back propagation, BP)神经网络是应用最广泛的非线性映

表 2 植被指数及其表达式

Tab. 2 Vegetation indexes and their formulas

植被指数名称	表达式
差值植被指数 DVI	$DVI = B_4 - B_3$
绿色归一化植被指数 GNDVI	$GNDVI = (B_4 - B_2) / (B_4 + B_2)$
绿色比值植被指数 GRVI	$GRVI = B_4 / B_2 - 1$
修正型土壤调整植被指数 MSAVI	$MSAVI = [2B_4 + 1 - \sqrt{(2B_4 + 1)^2 - 8(B_4 - B_3)}] / 2$
归一化植被指数 NDVI	$NDVI = (B_4 - B_3) / (B_4 + B_3)$
光谱结构不敏感植被指数 PSRI	$PSRI = (B_3 - B_1) / B_4$
再归一化植被指数 RDVI	$RDVI = \sqrt{\frac{(B_4 - B_3)^2}{B_4 + B_3}}$
比值植被指数 RVI	$RVI = B_4 / B_3$
土壤调整植被指数 SAVI	$SAVI = 1.5(B_4 - B_3) / (B_4 - B_3 + 0.5)$
冠层结构不敏感植被指数 SIPI	$SIPI = (B_4 - B_1) / (B_4 + B_1)$
三角植被指数 TVI	$TVI = 0.5 [120(B_4 - B_2) - 200(B_3 - B_2)]$
宽范围动态植被指数 WDRVI	$WDRVI = (0.2B_4 - B_3) / (0.2B_4 + B_3)$

射学习算法,它的中心思想是调整神经元节点权值使输出总误差最小<sup>[20]</sup>。本研究中,BP算法结构采用典型的三层网络标准结构,由输入层、隐含层和输出层组成,通过交叉验证优化得到各个参数。研究采用 Matlab R2014b 编程实现 BP 神经网络模型。

支持向量机(Support vector machine,SVM)是基于统计学习理论的机器学习方法,根据结构风险最小化准则,采用二次规划的方式得到问题最优解。SVM最初用于解决分类问题,即支持向量分类(Support vector classification,SVC),现已发展至可解决回归问题,即支持向量回归(Support vector regression,SVR)。目前,SVR已经被广泛应用于农作物 LAI 反演研究<sup>[14,21-22]</sup>。SVR 回归预测精度很大程度上由核函数类型和相关参数决定,因此,核函数和相关参数的确定是 SVR 建模的核心问题。该研究是利用 R 语言中的 kernlab 软件包实现 SVR 模型。

随机森林(Random forest,RF)模型是 Breiman 等提出的一种基于分类回归树的机器学习方法<sup>[23]</sup>。它通过自助法(bootstrap)抽样从原始训练集中抽取  $k$  个样本,且每个样本的样本容量均和原始训练集的大小一致;然后对每个样本分别进行决策树建模,得到  $k$  个建模结果;最后,利用所有决策树的建模结果,通过投票得到最终预测结果<sup>[24]</sup>。RF 建模的核心问题是确定森林中决策树的数量和分割节点的随机变量数。该研究是利用 R 语言中的 random Forest 软件包实现 RF 模型。

## 1.6 模型评价

对 RF 模型、BP 神经网络模型和 SVR 模型的验证集进行评价,常用的评价指标有决定系数( $R^2$ )、

均方根误差(RMSE)、平均相对误差(MAE)和相对分析误差(RPD)<sup>[16]</sup>。 $R^2$ 和 RPD 越大, RMSE 和 MAE 越小,表明模型估测精度越高。

为了检验模型的稳定性,采用保留样本交叉检验方式<sup>[25]</sup>将 LAI 原始数据按 3:2 的比例随机划分为 2 部分,60% (35 个)作为建模集用以建立模型,40% (24 个)作为检验集用以评价模型,并重复进行 3 次随机分组,记为样本组 1、样本组 2、样本组 3。

## 2 结果与分析

### 2.1 LAI 与植被指数的相关性分析

表 3 为不同划分方式实测样点 LAI 值与遥感植被指数间的相关性分析结果。从中可知,除 PSRI 与 LAI 值呈显著相关( $P < 0.05$ )外,样本组 1 中的样点 LAI 值与其余 11 个植被指数均呈极显著相关( $P < 0.01$ ),其中建模集中 LAI 值和植被指数的相关系数均大于 0.6,检验组中 LAI 值和植被指数的相关系数在 0.7 之上;样本组 2 和样本组 3 中除建模集中的 PSRI 与 LAI 值无显著相关( $P > 0.05$ )外,样点 LAI 值与其余植被指数均呈极显著相关( $P < 0.01$ ),且它们的相关系数均高于 0.5;从全样本来,各植被指数与 LAI 值均呈极显著相关( $P < 0.01$ ),相关系数由大到小依次为:RVI、WDRVI、NDVI(SAVI)、MSAVI、GRVI、RDVI、TVI、GNDVI、DVI、SIPI、PSRI。所有植被指数中,除 PSRI 外,其余植被指数与各样本组及全样本中样点 LAI 值均在  $P < 0.01$  水平上显著相关。因此这些植被指数可以较好地反映研究区 LAI 值的变化情况,可以用来构建 LAI 估算模型。

表 3 不同样本组 LAI 值与植被指数间的相关性分析结果

Tab. 3 Correlation of LAI and vegetation indices for different sample groups

植被指数	样本组 1		样本组 2		样本组 3		全样本
	建模集	检验集	建模集	检验集	建模集	检验集	
DVI	0.713 **	0.790 **	0.607 **	0.899 **	0.654 **	0.851 **	0.731 **
GNDVI	0.715 **	0.826 **	0.641 **	0.852 **	0.694 **	0.797 **	0.733 **
GRVI	0.752 **	0.849 **	0.683 **	0.900 **	0.736 **	0.839 **	0.774 **
MSAVI	0.753 **	0.904 **	0.714 **	0.868 **	0.761 **	0.819 **	0.782 **
NDVI	0.770 **	0.907 **	0.729 **	0.882 **	0.774 **	0.835 **	0.795 **
PSRI	0.400 *	0.496 *	0.300	0.588 **	0.298	0.615 **	0.419 **
RDVI	0.745 **	0.858 **	0.673 **	0.900 **	0.717 **	0.855 **	0.769 **
RVI	0.802 **	0.903 **	0.765 **	0.907 **	0.801 **	0.864 **	0.824 **
SAVI	0.770 **	0.907 **	0.729 **	0.882 **	0.774 **	0.835 **	0.795 **
SIPI	0.662 **	0.763 **	0.570 **	0.825 **	0.610 **	0.789 **	0.680 **
TVI	0.723 **	0.812 **	0.629 **	0.904 **	0.671 **	0.859 **	0.743 **
WDRVI	0.789 **	0.907 **	0.748 **	0.897 **	0.789 **	0.853 **	0.812 **

注: \*、\*\*表示在  $P < 0.05$  水平(双侧)和  $P < 0.01$  水平(双侧)上显著相关。

### 2.2 林地 LAI 反演模型比较分析

分别利用不同样本组中的建模集,以植被指数为自变量,对应样点的 LAI 值为因变量,使用 BP 算法、SVR 算法和 RF 回归算法构建 3 个不同样本组的 LAI 值遥感估算模型。本文中 BP 算法结构采用三层网络标准结构,通过交叉验证优化得到输入层到隐含层传递函数为  $\text{tansig}$ 、隐含层到输出层传递函数为  $\text{logsig}$ 、训练方式为 Levenberg - Marquardt 算法、迭代次数为 3000 的误差反向传播模型。SVR 算法的建立包括核函数和相关参数的选取,本研究中,SVR 算法中核函数选用高斯核函数,惩罚因子  $C$  取

10,核函数参数  $\sigma$  取 0.1。RF 回归算法的核心是确定森林中决策树的数量和分割节点的分割变量数,根据经验及反复试验,本研究中森林决策树的数量为 500,节点的分割变量为 3。

基于不同样本组中独立于建模集的检验集,得到 3 个模型的 LAI 预测数据,通过与对应的 LAI 实测数据比较,评价不同模型的预测能力,同时对 LAI 实测数据和预测数据进行回归分析,并绘制 3 种模型 LAI 实测值和预测值的 1:1 关系线,结果如表 4 和图 2 所示。

由表 4 可以看出,3 个样本组中 RF 模型的  $R^2$

表 4 LAI 估算模型比较

Tab. 4 Comparison of LAI estimation models

评价指标	样本组 1			样本组 2			样本组 3			平均		
	RF	BP	SVR									
$R^2$	0.688	0.419	0.609	0.796	0.454	0.629	0.707	0.543	0.629	0.730	0.472	0.622
RMSE	0.509	1.057	0.592	0.658	1.634	0.820	0.696	1.241	0.781	0.621	1.311	0.731
MAE	0.417	0.888	0.483	0.414	1.152	0.611	0.466	0.905	0.604	0.432	0.982	0.566
RPD	1.653	1.338	1.435	1.984	1.136	1.515	1.731	1.308	1.465	1.789	1.261	1.472

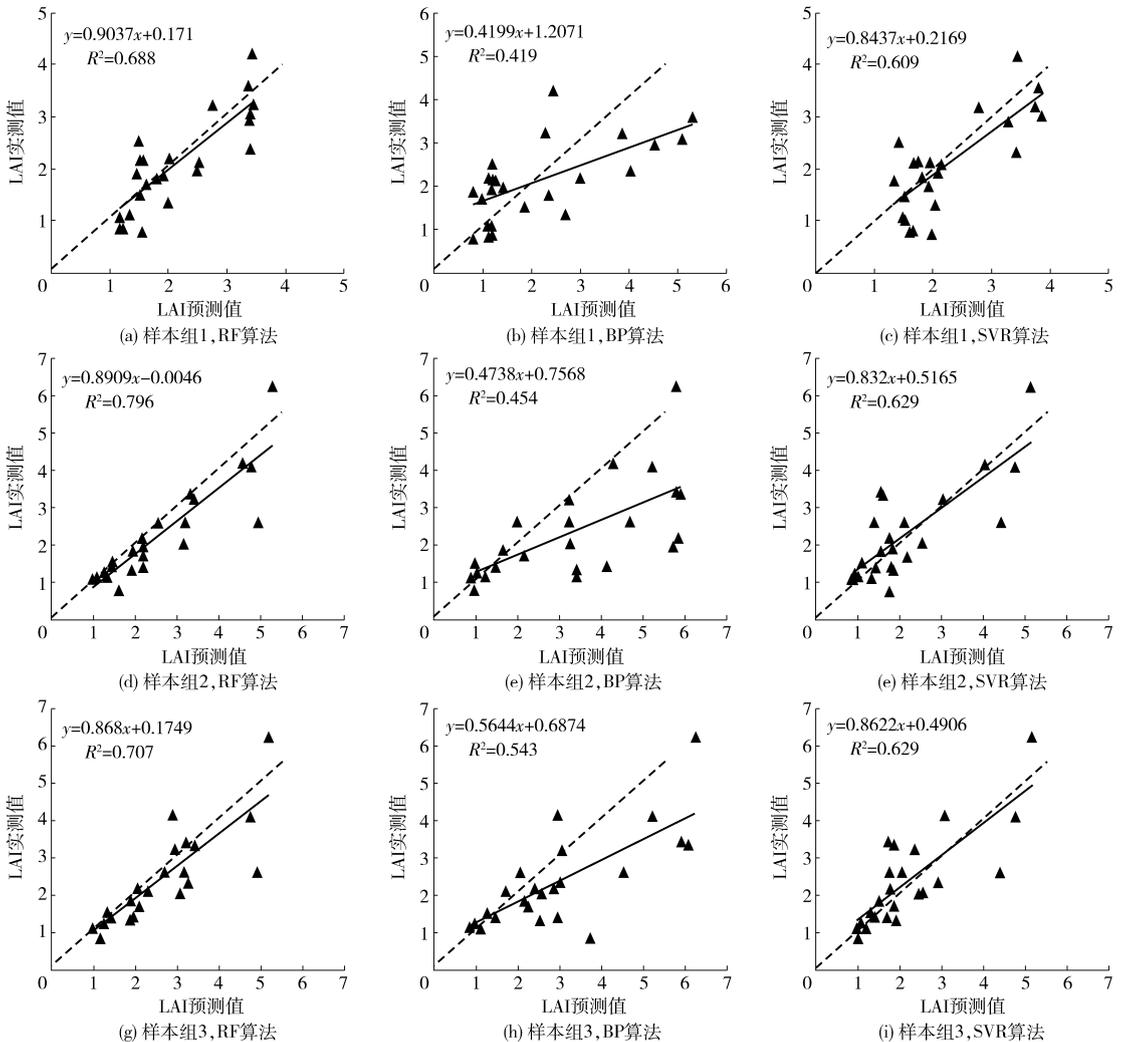


图 2 不同模型 LAI 实测值与预测值回归分析

Fig. 2 Regression analysis of measured and predicted LAI for different models

均在 0.688 以上,比 BP 模型至少提高了 16.4%,比 SVR 模型至少提高了 7.9%,相应的 RMSE 和 MAE 均最小。其中,RF 模型的 RMSE 与 SVR 模型大致相当,与 BP 模型相差较大,其值比 BP 模型依次低 0.548、0.976、0.545;RF 模型的 MAE 最大值为 0.466,低于其他 2 种模型的最小值;3 个样本组中 RPD 由大到小顺序均为 RF、SVR、BP。以上分析表明,3 个样本组中的 RF 模型预测结果与同组的 BP 模型和 SVR 模型相比,均表现出最强的预测能力。

图 2 中虚线表示 1:1 关系线,LAI 实测值与预测值线性回归线越接近 1:1 关系线,则说明模型的预测效果越好。从图 2 可以看出,各模型均不同程度

上高估了 LAI 地面实测值,不同样本组中 RF 模型和 SVR 模型的回归线与 1:1 关系线的吻合度较好,预测精度较高;BP 模型的回归线远离 1:1 关系线,预测精度较低。

综上,与 BP 模型和 SVR 模型相比,RF 模型得到的反演结果较为理想。因此,可以用 RF 模型进行林地 LAI 遥感估算。

### 2.3 林地 LAI 空间分布

通过 R 语言中的 random Forest 软件包,利用估计精度最高的样本组 2 数据对研究区林地 LAI 进行预测,采用 ArcGIS 9.3 制作基于 RF 模型的林地 LAI 的空间分布图如图 3 所示。

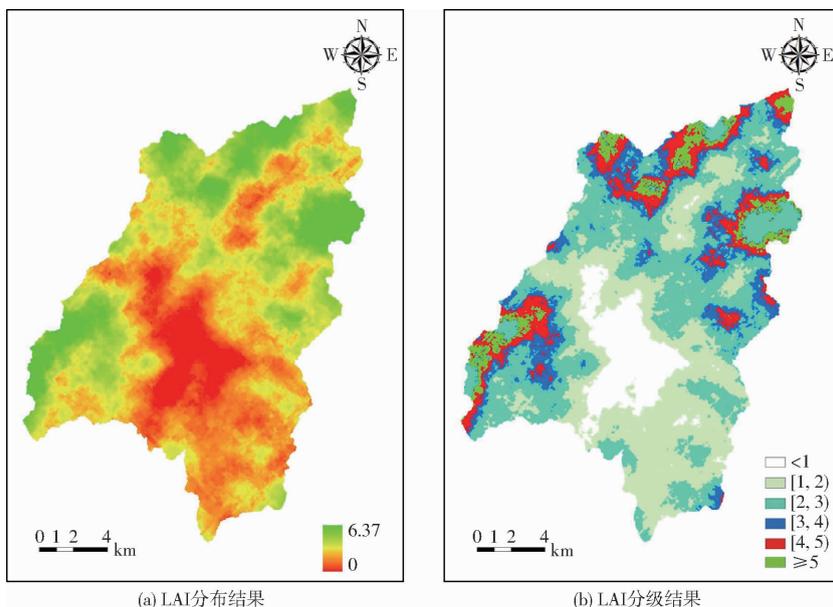


图 3 河田镇林地 LAI 空间分布图

Fig. 3 Spatial distribution of forest LAI in Hetian County

从图 3 可以看出,研究区 LAI 值范围在 0 ~ 6.37 之间,整体上呈“北高中低”的分布格局;北部地区林地 LAI 值相对较高,大部分区域的 LAI 值高于 3,主要原因在于该区是阔叶树的主要分布区,林分结构较好;研究区中部地区林地 LAI 值相对较低,原因在于该区植被稀疏,且林分多为单一马尾松纯林,林下水土流失严重。因而,基于 RF 模型预测的 LAI 值较好地还原了研究区 LAI 值的空间分布趋势。

## 3 讨论

遥感是快速获取大面积森林植被冠层光谱反射率的重要手段,但易受大气环境、植被冠层结构、土壤背景、太阳天顶角等因素的影响。有研究表明,LAI 与冠层光谱反射率之间存在较强的相关关系<sup>[24]</sup>,因而,有学者选取特征波段构建植被指数用来定量反演林地 LAI。以往遥感估算林地 LAI 的研

究中,多采用单一植被指数进行遥感反演且选取的植被指数数量较少<sup>[19,26]</sup>,由于单个植被指数只包含单波段的信息且每种植被指数均存在不同程度的饱和性,因此,基于单一植被指数得到的 LAI 遥感反演模型推广能力较差。

除自变量(植被指数)选取外,建模算法是影响林地 LAI 遥感监测的重要因素,选择一种合适的模型,有助于提高植被理化参数的遥感预测精度。本研究首次将机器学习算法中的随机森林(RF)模型引入林地 LAI 遥感估算中,针对样本组 1、样本组 2 和样本组 3,分别构建了以 12 个植被指数为自变量的 RF 模型,并将模型预测性能与支持向量回归(SVR)模型和反向传播神经网络(BP)模型性能进行了比较,结果表明 RF 模型的预测精度最高,这与王丽爱等<sup>[27]</sup>关于农作物 LAI 的遥感估算结果相一致。后期工作将进一步研究 RF 模型能否适用林地生物量、土壤养分等林地质量参数的遥感估算,以提

升其在精准林业遥感监测中的应用价值。

本研究的 RF 模型表现出比其他两种模型更好的预测效果,其原因在于 RF 模型不需要检查变量的交互作用是否显著,加上它进行了两次随机抽样,使得模型在异常值和噪声方面具有较高的容忍度,而且不容易出现过拟合现象,因此该模型具有较高的反演性能<sup>[28]</sup>;SVR 模型的性能取决于核函数及其参数的选取,目前对核函数及其参数的选取均为人经验选择,带有一定随意性和盲目性,本研究也如此,这在一定程度上制约了模型的反演性能;BP 模型的预测效果最差,原因可能在于模型在训练过程中陷入了局部最优状态或模型的学习能力过强,导致学习出的模型已表现不出训练样本所包含的规律,最终弱化了模型的泛化能力。此外,这 3 种机器学习模型均要设置参数,SVR 模型需要设置多个参数(核函数类型、核参数、惩罚因子等),BP 模型需要的参数也较多(网络层数、节点数、神经元个数、传递函数、学习率等),而 RF 模型仅需 2 个参数(森林决策树数量和节点分割变量),这大大提升了 RF 模型的应用性。另外,本研究采用的 12 个植被指数

多为红光波段和近红外波段的组合,存在严重的多重共线性,但 RF 模型对多元共线性不敏感。因此,RF 模型可以作为林地 LAI 遥感估算的优选方法。

值得注意的是,本研究中不同样本组的建模集和检验集均基于同一个年份(2015 年)的数据,模型在时间维方面的普适性,尚需要进一步验证。另外,王正兴等<sup>[29]</sup>的研究表明,植被指数存在光谱“饱和”现象,仅利用多植被指数组合作为自变量,可能会影响模型的预测精度,并减弱其泛化能力。后续工作将耦合林地遥感纹理指数和植被指数,引入地形环境因子作为校正参数,以提高林地 LAI 遥感估算的精度。

#### 4 结束语

基于高分辨率 Pleiades 遥感数据,构建了 3 种林地 LAI 遥感估算模型,通过与地面实测的 LAI 数据进行比较,结果表明基于 RF 算法构建的林地 LAI 遥感估算模型具有较高的预测精度和泛化能力,该模型可为林地 LAI 的快速、准确、无损遥感监测提供一种新的机器学习方法。

#### 参 考 文 献

- 1 CHEN J M, BLACK T A. Defining leaf area index for non-flat leaves [J]. *Plant Cell and Environment*, 1992, 15(4): 421 - 429.
- 2 PU R, CHENG J. Mapping forest leaf area index using reflectance and textural information derived from WorldView-2 imagery in a mixed natural forest area in Florida, US [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2015, 42: 11 - 23.
- 3 DONG T, LIU J, QIAN B, et al. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2016, 49: 63 - 74.
- 4 NEINAVAZ E, SKIDMORE A K, DARVISHZADEH R, et al. Retrieval of leaf area index in different plant species using thermal hyperspectral data [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 119: 390 - 401.
- 5 任海, 彭少麟. 鼎湖山森林群落的几种叶面积指数测定方法的比较[J]. *生态学报*, 1997, 17(2): 220 - 223.
- 6 REN Hai, PENG Shaolin. Comparison of methods of estimation leaf area index in Dinghushan forests [J]. *Acta Ecologica Sinica*, 1997, 17(2): 220 - 223. (in Chinese)
- 7 STENBERG P, RAUTIAINEN M, MANNINEN T, et al. Boreal forest leaf area index from optical satellite images: model simulations and empirical analyses using data from central Finland [J]. *Boreal Environment Research*, 2008, 13: 433 - 443.
- 8 刘婧怡, 汤旭光, 常守志, 等. 森林叶面积指数遥感反演模型构建及区域估算[J]. *遥感技术与应用*, 2014, 29(1): 18 - 25.
- 9 LIU Jingyi, TANG Xuguang, CHANG Shouzhi, et al. Application of remote sensing to inverse the forest leaf area index and regional estimation [J]. *Remote Sensing Technology and Application*, 2014, 29(1): 18 - 25. (in Chinese)
- 10 朱绪超, 袁国富, 易小波, 等. 基于 Landsat 8 OLI 影像的塔里木河下游河岸林叶面积指数反演[J]. *干旱区地理*, 2014, 37(6): 1248 - 1256.
- 11 ZHU Xuchao, YUAN Guofu, YI Xiaobo, et al. Leaf area index inversion of riparian forest in the lower basin of Tarim River based on Landsat 8 OLI images [J]. *Arid Land Geography*, 2014, 37(6): 1248 - 1256. (in Chinese)
- 12 李粉玲, 王力, 刘京, 等. 基于高分一号卫星数据的冬小麦叶片 SPAD 值遥感估算[J/OL]. *农业机械学报*, 2015, 46(9): 273 - 281. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20150940&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20150940&flag=1). DOI:10.6041/j.issn.1000-1298.2015.09.040.
- 13 LI Fenling, WANG Li, LIU Jing, et al. Remote sensing estimation of SPAD value for wheat leaf based on GF-1 data [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(9): 273 - 281. (in Chinese)
- 14 DENG F, CHEN J M, PLUMMER S, et al. Algorithm for global leaf area index retrieval using satellite imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2006, 44(8): 2219 - 2229.
- 15 谷成燕, 杜华强, 周国模, 等. 基于 PROSAIL 辐射传输模型的毛竹林叶面积指数遥感反演[J]. *应用生态学报*, 2013, 24(8): 2248 - 2256.
- 16 GU Chengyan, DU Huaqiang, ZHOU Guomo, et al. Retrieval of leaf area index of Moso bamboo forest with Landsat Thematic Mapper image based on PROSAIL canopy radiative transfer model [J]. *Chinese Journal of Applied Ecology*, 2013, 24(8): 2248 - 2256. (in Chinese)
- 17 林卉, 梁亮, 张连蓬, 等. 基于支持向量机回归算法的小麦叶面积指数高光谱遥感反演[J]. *农业工程学报*, 2013,

- 29(11): 139 - 146.
- LIN Hui, LIANG Liang, ZHANG Lianpeng, et al. Wheat leaf area index inversion with hyperspectral remote sensing based on support vector regression algorithm [J]. Transactions of the CSAE, 2013, 29(11): 139 - 146. (in Chinese)
- 13 王丽爱, 周旭东, 朱新开, 等. 基于 HJ-CCD 数据和随机森林算法的小麦叶面积指数反演[J]. 农业工程学报, 2016, 32(3): 149 - 154.
- WANG Liai, ZHOU Xudong, ZHU Xinkai, et al. Inverting wheat leaf area index based on HJ-CCD remote sensing data and random forest algorithm [J]. Transactions of the CSAE, 2016, 32(3): 149 - 154. (in Chinese)
- 14 王丽爱, 谭昌伟, 杨昕, 等. 基于 MK - SVR 模型的小麦叶面积指数遥感反演[J/OL]. 农业机械学报, 2015, 46(5): 245 - 251. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20150535&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20150535&flag=1). DOI:10.6041/j.issn.1000-1298.2015.05.035.
- WANG Liai, TAN Changwei, YANG Xin, et al. Monitoring wheat leaf area index using MK - SVR algorithmic model and remote sensing data [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(5): 245 - 251. (in Chinese)
- 15 王修信, 孙涛, 朱启疆, 等. 林地叶面积指数遥感估算方法适用分析[J]. 生态学报, 2014, 34(16): 4612 - 4619.
- WANG Xiuxin, SUN Tao, ZHU Qijiang, et al. Assessment of different methods for estimating forest leaf area index from remote sensing data [J]. Acta Ecologica Sinica, 2014, 34(16): 4612 - 4619. (in Chinese)
- 16 韩兆迎, 朱西存, 房贤一, 等. 基于 SVM 与 RF 的苹果树冠 LAI 高光谱估测[J]. 光谱学与光谱分析, 2016, 36(3): 800 - 805.
- HAN Zhaoying, ZHU Xicun, FANG Xianyi, et al. Hyperspectral estimation of apple tree canopy LAI based on SVM and RF regression [J]. Spectroscopy and Spectral Analysis, 2016, 36(3): 800 - 805. (in Chinese)
- 17 刘洋, 刘荣高, 陈镜明, 等. 叶面积指数遥感反演研究进展与展望[J]. 地球信息科学学报, 2013, 15(5): 734 - 743.
- LIU Yang, LIU Ronggao, CHEN Jingming, et al. Current status and perspectives of leaf area index retrieval from optical remote sensing data [J]. Journal Geo-Information Science, 2013, 15(5): 734 - 743. (in Chinese)
- 18 何亚娟, 潘学标, 裴志远, 等. 基于 SPOT 遥感数据的甘蔗叶面积指数反演和产量估算[J/OL]. 农业机械学报, 2013, 44(5): 226 - 231. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20130539&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20130539&flag=1). DOI:10.6041/j.issn.1000-1298.2013.05.039.
- HE Yajuan, PAN Xuebiao, PEI Zhiyuan, et al. Estimation of LAI and yield of sugarcane based on SPOT remote sensing data [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(5): 226 - 231. (in Chinese)
- 19 刘玉琴, 沙晋明, 余涛, 等. 基于宽波段和窄波段植被指数的草地 LAI 反演对比研究[J]. 遥感技术与应用, 2014, 29(4): 587 - 593.
- LIU Yuqin, SHA Jinming, YU Tao, et al. Comparing the performance of broad-band and narrow-band vegetation indices for estimation of grass LAI [J]. Remote Sensing Technology and Application, 2014, 29(4): 587 - 593. (in Chinese)
- 20 郭孝玉, 孙玉军, 王轶夫, 等. 基于改进人工神经网络的植物叶面积测定[J/OL]. 农业机械学报, 2013, 44(2): 200 - 204. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20130237&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20130237&flag=1). DOI:10.6041/j.issn.1000-1298.2013.02.037.
- GUO Xiaoyu, SUN Yujun, WANG Yifu, et al. Improved artificial neural network for determination of plant leaf area [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(2): 200 - 204. (in Chinese)
- 21 梁栋, 管青松, 黄文江, 等. 基于支持向量机回归的冬小麦叶面积指数遥感反演[J]. 农业工程学报, 2013, 29(7): 117 - 123.
- LIANG Dong, GUAN Qingsong, HUANG Wenjiang, et al. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat [J]. Transactions of the CSAE, 2013, 29(7): 117 - 123. (in Chinese)
- 22 谢巧云, 黄文江, 蔡淑红, 等. 冬小麦叶面积指数遥感反演方法比较研究[J]. 光谱学与光谱分析, 2014, 34(5): 1352 - 1356.
- XIE Qiaoyun, HUANG Wenjiang, CAI Shuhong, et al. Comparative study on remote sensing inversion methods for estimating winter wheat leaf area index [J]. Spectroscopy and Spectral Analysis, 2014, 34(5): 1352 - 1356. (in Chinese)
- 23 BREIMAN L. Random forest[J]. Machine Learning, 2001, 45(1): 5 - 32.
- 24 李映雪, 朱艳, 戴廷波, 等. 小麦叶面积指数与冠层反射光谱的定量关系[J]. 应用生态学报, 2006, 17(8): 1443 - 1447.
- LI Yingxue, ZHU Yan, DAI Tingbo, et al. Quantitative relationships between leaf area index and canopy reflectance spectra of wheat [J]. Chinese Journal of Applied Ecology, 2006, 17(8): 1443 - 1447. (in Chinese)
- 25 史舟. 土壤地面高光谱遥感原理与方法[M]. 北京: 科学出版社, 2014.
- 26 贾玉秋, 李冰, 程永政, 等. 基于 GF-1 与 Landsat-8 多光谱遥感影像的玉米 LAI 反演比较[J]. 农业工程学报, 2015, 31(9): 173 - 179.
- JIA Yuqiu, LI Bing, CHENG Yongzheng, et al. Comparison between GF-1 imagines and Landsat-8 imagines in monitoring maize LAI [J]. Transactions of the CSAE, 2015, 31(9): 173 - 179. (in Chinese)
- 27 王丽爱, 马昌, 周旭东, 等. 基于随机森林回归算法的小麦叶片 SPAD 值遥感估算[J/OL]. 农业机械学报, 2015, 46(1): 259 - 265. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20150136&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20150136&flag=1). DOI:10.6041/j.issn.1000-1298.2015.01.036.
- WANG Liai, MA Chang, ZHOU Xudong, et al. Estimation of wheat leaf SPAD value using RF algorithmic model and remote sensing data [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 259 - 265. (in Chinese)
- 28 RODRIGUEZ-GALIANO V, MENDES M P, GARCIA-SOLDADO M J, et al. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain) [J]. Science of the Total Environment, 2014, 476: 189 - 206.
- 29 王正兴, 刘闯, HUETE A. 植被指数研究进展: 从 AVHRR-NDVI 到 MODIS-EVI[J]. 生态学报, 2003, 23(5): 979 - 987.
- WANG Zhengxing, LIU Chuang, HUETE A. From VVHRR-NDVI to MODIS-EVI advances in vegetation index research [J]. Acta Ecologica Sinica, 2003, 23(5): 979 - 987. (in Chinese)